Tensor Network Methods and Artificial Intelligence Study of the Hubbard Model

Zero- and Finite-Temperature Simulations and Optimized Encoder-only Transformer Analysis of the Hubbard Model

Changkai Zhang (张昌凯)

München 2025

Tensor Network Methods and Artificial Intelligence Study of the Hubbard Model

Zero- and Finite-Temperature Simulations and
Optimized Encoder-only Transformer
Analysis of the Hubbard Model

Changkai Zhang (张昌凯)

Dissertation
an der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Changkai Zhang
aus China

München, den 11. November 2025

Erstgutachter: Prof. Jan von Delft
Zweitgutachter: Prof. Jad Halimeh
Datum der Abgabe: 2025-11-11
Datum der Prüfung: 2025-12-17

Abstract

The two-dimensional (2D) Hubbard model, widely believed to capture the essential physics of high-T_c cuprate compounds, has attracted immense research interest ever since the discovery of superconductivity in these materials. Despite its superficial simplicity, the Hubbard model incubates abundant fascinating phenomena owing to the strong interactions between charge carriers, encompassing antiferromagnetism, the pseudogap, and unconventional superconductivity. Yet these very interactions induce strong correlations, rendering canonical perturbative approaches unreliable. With advances in modern computational hardware, numerical techniques — notably tensor network methods and Artificial Intelligence (AI) algorithms — have become indispensable to our expedition.

Tensor networks encode the amplitude information of many-body quantum states in a network of interconnected tensors. Leveraging the entanglement area law, tensor networks provide efficient, systematic, and controllable representations of quantum states that would otherwise require an exponentially large number of parameters. Past decades have witnessed the development of various tensor network ansatzes, such as Matrix Product States/Operators (MPS/MPO) for one-dimensional systems and Projected Entangled-Pair States (PEPS) for two-dimensional systems, together with diverse algorithms for optimizing these ansatzes for both zero- and finite-temperature scenarios.

Over the past years, AI technologies have revolutionized the way we analyze scientific data. The transformer architecture, in particular, has demonstrated exceptional capabilities in the domain of natural language processing. Its attention mechanism excels at capturing long-range correlations in sequential data, making it a promising tool for studying strongly correlated many-body systems. The interpretable nature of the attention mechanism further offers valuable insights into the AI's perception of the underlying physics.

In this thesis, we assemble an all-inclusive numerical toolchain for analyzing strongly correlated lattice systems. As a demonstration, we investigate the zero- and finite-temperature properties of the 2D Hubbard model on a square lattice via the infinite PEPS (iPEPS) and the eXponential Tensor Renormalization Group (XTRG), respectively. The native two-dimensional iPEPS tensor network faithfully embodies

the entanglement area law and helps settle long-standing debates regarding superconducting order in the ground state of the Hubbard model. The XTRG algorithm efficiently constructs thermal density matrices across a broad temperature range via a highly effective exponential-cooling protocol, thereby enabling detailed probes of pairing and pseudogap phenomena. Afterwards, we generate a comprehensive snapshot dataset which furnishes subsequent training of an optimized encoder-only transformer model. Our novel AI architecture features a Markovian interpretation of the attention design and affords improved parallelism. The collective technological stack opens a brand new avenue for exploring the rich physics of strongly correlated many-body systems.

Acknowledgements

I would like to express my deepest gratitude to my supervisor Prof. Jan von Delft for his continuous support, guidance, and encouragement throughout my doctoral studies. I am especially grateful for his efforts in clearing countless obstacles along the way, his thoughtful mentorship in career development, as well as the intellectual latitude and steadfast backing in the exploration of novel ideas which eventually lead to our innovative AI technologies.

My heartfelt thanks go to my parents for their enduring support and understanding during my years abroad. Their unshakeable confidence has been a constant source of motivation, reliably dispersing the distractions and anxieties from the outside world.

I am likewise grateful to my colleagues and collaborators, particularly Jheng-Wei Li, Andreas Gleis, and Markus Schieb, for their insightful discussions and inspirations. Their expertise and perspectives have eminently enriched an important portion of this research. I also benefited from valuable feedback from Phillip Corboz, Wei Li, Annabelle Bohrdt, Fabian Grudst, Zi-Yang Meng, Ming-Pu Qin, Roger Melko, Daiwei Qu, Qiaoyi Li, Ming Huang, and Oleksandra Kovalska, as well as from Andreas Weichselbaum and Seung-Sup Lee regarding the ideas and development underlying the QSpace tensor library.

Special thanks are due to a variety of AI tools, chatbots, and agents — most notably OpenAI's GPT series and Anthropic's Claude — which, since 2023, have substantially amplified my productivity in writing, coding, and analysis.

Beyond the academy, I gratefully acknowledge the early support of my friends Gongrui Cheng and Guoen Nian during the beginning of my studies in Germany. I also extend special thanks to the Chinese Students and Scholars Association in Munich (CSSA-Munich, 2022-2024) for their community-building events, to the Bilibili uploaders, and to *Honkai: Star Rail* and *Zenless Zone Zero* — all of which provided essential mental support amid the otherwise monotonous rhythms of research.

Finally, I would like to acknowledge the financial support in part from the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy EXC-2111 (Project No. 390814868), and from the Munich Quantum Valley, supported by the Bavarian state government through the Hightech Agenda Bayern Plus.

May this journey lead us starward

愿此行，终抵群星

Terms and Acronyms

AFM	AntiFerroMagnetism
AFQMC	Auxiliary-Field QMC
AI	Artificial Intelligence
CBE	Controlled Bond Expansion
CG	Clebsch-Gordan
CGC	Clebsch-Gordan Coefficient
CNN	Convolutional Neural Network
CT	Continuous-Time or Computed Tomography
CTMRG	Corner Transfer Matrix Renormalization Group
CV	Computer Vision
DFT	Density Functional Theory
DL	Deep Learning
DMFT	Dynamical Mean-Field Theory
DMRG	Density Matrix Renormalization Group
DQMC	Determinant QMC
ED	Exact Diagonalization
FFU	Fast Full Update
FU	Full Update
GRU	Gated Recurrent Unit
HOG	Histogram of Oriented Gradients
HOTRG	Higher-Order Tensor Renormalization Group
iPEPS	infinite PEPS
irrep	irreducible representation
LRA	Low-Rank Approximation
LSTM	Long Short-Term Memory
LoRA	Low-Rank Adaptation
MERA	Multi-scale Entanglement Renormalization Ansatz
METTS	Minimally-Entangled Typical Thermal States
ML	Machine Learning
MLP	Multi-Layer Perceptron
MPO	Matrix Product Operator
MPS	Matrix Product States
MRI	Magnetic Resonance Imaging
$\boldsymbol{n}$**D**	n-dimensional
$\boldsymbol{n}$**s**	n-site
NLP	Natural Language Processing
NN	Nearest-Neighbor
NNN	Next-Nearest Neighbor
NQS	Neural Quantum State

NTB	Neural Transformer Backflow
MCMC	Markov-Chain Monte Carlo
PCA	Principal Component Analysis
PEPO	Projected Entangled-Pair Operator
PEPS	Projected Entangled-Pair State
PET	Positron Emission Tomography
QMC	Quantum Monte Carlo
QNN	Quantum Neural Network
QR	QR Decomposition
QSpace	a tensor library featured in symmetry resolution
RAG	Retrieval-Augmented Generation
RBM	Restricted Boltzmann Machine
RG	Renormalization Group
RNN	Recurrent Neural Network
SC	Superconductivity
SIFT	Scale-Invariant Feature Transform
STRIPS	Stanford Research Institute Problem Solver
SU	Simple Update
SVD	Singular Value Decomposition
SVM	Support Vector Machine
tanTRG	tangent-space Tensor Renormalization Group
TDVP	Time-Dependent Variational Principle
TN	Tensor Network
TNS	Tensor Network State
TensorKit	a tensor library for Julia
VMC	Variational Monte Carlo
XOR	logical operation `XOR`
XTRG	eXponential Tensor Renormalization Group
YBCO	Yttrium Barium Copper Oxide

Contents

Chapter 1
Introduction and Motivation

Various fascinating phenomena and phases of matter have been discovered in the high-T_c cuprate compounds, including antiferromagnetism (AFM) [1–6], the pseudogap [7–9], strange metal [10–14], and high-T_c superconductivity [15–20]. Elucidating the microscopic mechanisms responsible for these behaviors remains one of the central challenges in condensed matter physics. Yet the exceedingly intricate electronic dynamics pose formidable obstacles to both theoretical analysis and numerical simulation. This motivates the development of simplified effective models that distill the essential physics of these materials.

In generic solids, itinerant electrons experience electric-field screening from the positively charged ionic cores and other electrons, which substantially reduces the effective interaction between electrons and often yields a metallic or a Fermi-liquid behavior. By contrast, experiments show that half-filled cuprates are Mott insulators, with two electronic bands separated by a large Mott gap [21–25]. This indicates an anomalously strong on-site Coulomb repulsion among itinerant electrons. The Hubbard model [26, 27] was found paradigmatic for such strongly correlated electron systems, combining a kinetic term describing electron hoppings with an on-site interaction term that penalizes double occupancy.

Despite its austere simplicity, the Hubbard model has been found notoriously challenging to solve, as strong interactions invalidate conventional perturbative methods. With the advances of high performance computational infrastructure, a suite of powerful numerical techniques has been developed to assist — including Quantum Monte Carlo (QMC) [28–31], Dynamical Mean-Field Theory (DMFT) [32–35], Density Matrix Renormalization Group (DMRG) [36–40], and a variety of zero- [41–48] and finite-temperature [49–57] Tensor Network State (TNS) algorithms. Each method, however, has its respective comfort zone; a comprehensive understanding of the Hubbard model thus requires their integrated and complementary application.

Projected Entangled-Pair States (PEPS) [41–43, 45–47] constitute a genuinely two-dimensional (2D) TNS ansatz targeting ground states of lattice many-body systems. Infinite PEPS (iPEPS) [42, 58, 59] exploits translational invariance in infinite 2D lattices, enabling direct simulations in the thermodynamic limit. Relative to

other ground-state approaches, iPEPS is free of the sign problem and faithfully preserves the native 2D geometry. Exponential Tensor Renormalization Group (XTRG) [54,55,57,60] is a finite-temperature TNS algorithm that constructs thermal density matrices efficiently across a broad temperature window via an exponential cooling protocol. Compared to alternative finite-temperature schemes, XTRG excels in its wide coverage of phase space, thereby supporting the exploration of phenomena that emerge at disparate temperature scales.

In recent years, the rapid development of artificial intelligence (AI) techniques has revolutionized the way we perceive and analyze scientific data. The transformer architecture [61], grounded in the attention mechanism [62–66], has achieved remarkable success in natural language processing (NLP) [67–72], computer vision (CV) [73–76], bioinformatics [77–79], and many other domains [80–83]. Attention enables the model to focus selectively on the most informative components of the input, capturing long-range dependencies and intricate correlation patterns [84, 85]. These properties render transformer models particularly well-suited to the study of strongly correlated electron systems such as the Hubbard model.

Our expedition starts with the zero-temperature iPEPS and finite-temperature XTRG simulations of the Hubbard model on a square lattice. We systematically chart ground-state and thermal properties across a range of doping levels and next-nearest-neighbor hopping amplitudes. The resulting TNS representations yield a comprehensive snapshot dataset via site-wise sampling, which in turn furnishes an AI-driven analysis. We design an optimized encoder-only transformer (core architecture) with an interpretable attention mechanism to classify snapshots drawn from distinct regions in phase space. The trained model proves highly effective at capturing underlying correlation structures in the snapshot data and facilitates accurate omnimetry (i.e., the simultaneous measurement of all calibrated observables) for ultracold atom experiments [86–91].

1.1 High-T_c Cuprates and the Hubbard Model

The high-T_c cuprates comprise a broad family of copper-oxide materials and exhibit a rich phase diagram [4, 92–95] as a function of charge doping and temperature, as sketched in Fig. 1.1(a). The undoped parent compounds are Mott insulators [21–25] with long-range AFM order [1–4, 96, 97]. Upon doping, AFM order is rapidly suppressed and a pseudogap phase [4, 9, 98–100] emerges below the onset temperature T^*. At lower temperatures, superconductivity (SC) [18–20, 101–106] develops below the critical temperature T_c, forming a characteristic dome in the phase diagram. The strange-metal regime [10–13], marked by non-Fermi-liquid behavior, appears at higher temperatures above the SC dome, while the overdoped regime tends to recover conventional Fermi-liquid properties [107–111]. Additional charge/spin density waves [106,112–128] and electronic nematicity [129,130] have been observed in specific doping ranges. The precise nature of these phases and their mutual interplay remain subjects of active investigation.

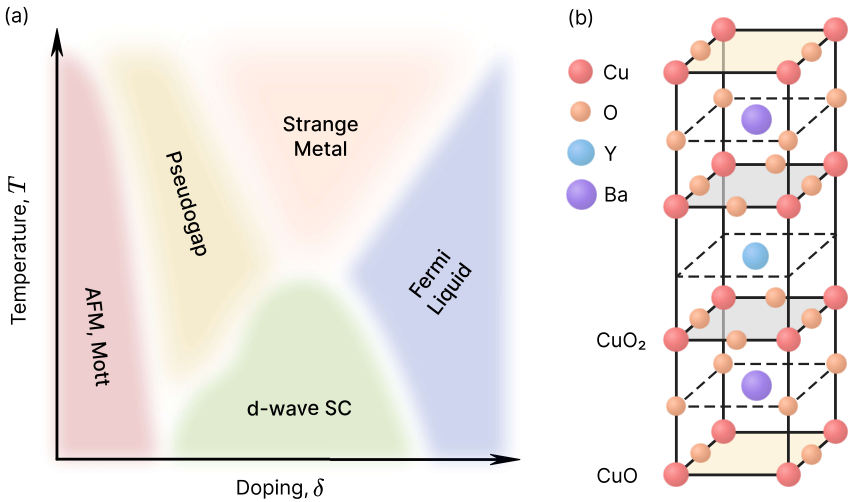


Fig. 1.1 (a) Schematic phase diagram of high-T_c cuprate compounds as a function of charge doping δ and temperature T. The phase structure of hole-doped and electron-doped cuprates differs, though the overall qualitative features remain similar. (b) Crystal structure of YBCO, a prototypical high-T_c cuprate. Superconductivity occurs in the CuO_2 planes.

Deeper insight into these emergent phenomena demands careful scrutiny of the crystal structure. Consider the prototypical high-T_c cuprate $YBa_2Cu_3O_{7-y}$ (YBCO). YBCO is a layered, perovskite-derived cuprate with an orthorhombic unit cell stacked along the vertical direction (see Fig. 1.1(b)) as CuO_2-BaO-CuO-BaO-CuO_2, with Yttrium sandwiched between the two CuO_2 planes. [131–133] Superconductivity resides within these CuO_2 layers [25, 134–137], while the quasi-1D CuO chains act as a charge reservoir: by incorporating oxygen and ordering along the chains, they dope the planes and tune the hole concentration. This intrinsically anisotropic structure suggests that high-T_c superconductivity is essentially a 2D phenomenon: the pairing condensate is confined to the CuO_2 sheets, with interlayer coupling entering weakly via apical oxygens. Consequently, the low-energy physics is well captured by 2D quantum lattice models, supplemented by a reservoir-like role of the chains that sets the carrier density and modulates scattering without themselves hosting notable portion of superconducting condensate.

At low energies, the active orbitals in a CuO_2 plane are the Cu $3d_{x^2-y^2}$ and the in-plane O $2p_{x,y}$. [138–140] A minimal three-band Emery-Hubbard model [141, 142] assigns a d orbital to each Cu site and $2p_{x,y}$ orbitals to the neighboring O sites. The Hamiltonian (in the hole picture) is

$$\begin{aligned}\mathcal{H} = \varepsilon_d \sum_{i,\sigma} n^d_{i\sigma} + \varepsilon_p \sum_{j,\alpha,\sigma} n^{p_\alpha}_{j\sigma} \\ + U_d \sum_i n^d_{i\uparrow} n^d_{i\downarrow} + U_p \sum_{j,\alpha} n^{p_\alpha}_{j\uparrow} n^{p_\alpha}_{j\downarrow} + U_{pd} \sum_{\langle i,j\rangle,\alpha,\sigma,\sigma'} n^d_{i\sigma} n^{p_\alpha}_{j\sigma'} \\ - t_{pd} \sum_{\langle i,j\rangle,\alpha,\sigma} \left(d^\dagger_{i\sigma} p_{j\alpha\sigma} + p^\dagger_{j\alpha\sigma} d_{i\sigma} \right) - t_{pp} \sum_{\langle j,j'\rangle,\alpha,\beta,\sigma} \left(p^\dagger_{j\alpha\sigma} p_{j'\beta\sigma} + \text{h.c.} \right).\end{aligned} \tag{1.1}$$

Here $d^\dagger_{i\sigma}$ creates a hole with spin σ in the Cu $d_{x^2-y^2}$ orbital at Cu site i, $p^\dagger_{j\alpha\sigma}$ creates a hole in the O $2p_\alpha$ orbital ($\alpha = x, y$) at oxygen site j, and $n^d_{i\sigma} = d^\dagger_{i\sigma} d_{i\sigma}$, $n^{p_\alpha}_{j\sigma} = p^\dagger_{j\alpha\sigma} p_{j\alpha\sigma}$. The on-site energies ε_d and ε_p define the charge-transfer gap ($\Delta = \varepsilon_p - \varepsilon_d$); U_d and U_p are the on-site Coulomb repulsions on Cu and O, respectively; U_{pd} denotes the nearest-neighbor Cu-O repulsion, while t_{pd} (Cu-O hybridization) and t_{pp} (O-O hopping) encode hole mobility within the Cu-O network.

Doped holes predominantly occupy oxygen $2p$ orbitals [138] but hybridize with the Cu $d_{x^2-y^2}$ local moments to form Zhang-Rice singlets [143–145] on CuO_4 plaquettes. Projecting onto this low-energy subspace yields an effective one-band Hubbard model on the Cu sites (the singlet band):

$$\mathcal{H} = -\sum_{i,j,\sigma} t^{(h)}_{ij} \left(h^\dagger_{i\sigma} h_{j\sigma} + \text{h.c.} \right) + U_{\text{eff}} \sum_i n^{(h)}_{i\uparrow} n^{(h)}_{i\downarrow} - \mu^{(h)} \sum_{i,\sigma} n^{(h)}_{i\sigma}, \tag{1.2}$$

where $h^\dagger_{i\sigma}$ creates a hole in the Zhang-Rice-derived band, $n^{(h)}_{i\sigma} = h^\dagger_{i\sigma} h_{i\sigma}$, and $t^{(h)}_{ij}$ collects the effective hoppings. To pass to the electron representation commonly used in numerical simulations, we perform a particle-hole transformation:

$$h^\dagger_{i\sigma} = c_{i\sigma}, \qquad h_{i\sigma} = c^\dagger_{i\sigma}, \qquad n^{(h)}_{i\sigma} = 1 - n_{i\sigma}, \tag{1.3}$$

with $c^\dagger_{i\sigma}$ creating an electron and $n_{i\sigma} = c^\dagger_{i\sigma} c_{i\sigma}$. For $i \neq j$, the kinetic term changes sign under this transformation, and the chemical potential shifts according to $n^{(h)}_{i\uparrow} n^{(h)}_{i\downarrow} = (1 - n_{i\uparrow})(1 - n_{i\downarrow})$. Up to an additive constant, the one-band Hubbard Hamiltonian becomes

$$\mathcal{H} = -\sum_{i,j,\sigma} t_{ij} \left(c^\dagger_{i\sigma} c_{j\sigma} + \text{h.c.} \right) + U \sum_i n_{i\uparrow} n_{i\downarrow} - \mu \sum_{i,\sigma} n_{i\sigma}, \tag{1.4}$$

where $t_{ij} = -t^{(h)}_{ij}$, $U = U_{\text{eff}}$, and $\mu = U_{\text{eff}} - \mu^{(h)}$. These steps make explicit how the widely used one-band Hubbard model emerges as a low-energy reduction of the three-band, charge-transfer description. While this simplification can deviate quantitatively from the actual material in certain regimes (e.g., spectroscopy or oxygen-sensitive responses), it preserves the essential ingredients — strongly correlated carriers moving on a 2D square lattice — that govern much of the cuprates' low-energy phenomenology.

1.2 Zero-temperature Methods

At zero temperature, the system resides in its ground state(s). Therefore, numerical methods usually focus on simulating or encoding the many-body wavefunction with the lowest energy for the system. Popular numerical methods to date include Quantum Monte Carlo (QMC) methods, Density Matrix Renormalization Group (DMRG), and infinite Projected Entangled-Pair States (iPEPS). Each method has its respective advantageous domains, and which method to apply thus depends on the specific system and the physical questions of interest.

Quantum Monte Carlo Simulation

Ground-state QMC approaches project a trial state in imaginary time using path-integral factorizations (Trotter-Suzuki) or Hubbard-Stratonovich fields — e.g., Determinant QMC (DQMC) [28, 146] or Auxiliary-Field QMC (AFQMC) [29] formulated in Slater-determinant space. These schemes sample the imaginary-time path integral to obtain largely unbiased estimators when the sign problem is absent, e.g., many half-filled bipartite models without frustration. However, for doped, frustrated, or generic fermionic systems, the average sign decays exponentially with system size or projection time, turning polynomial algorithms effectively exponential. [147] Workarounds (e.g., phaseless AFQMC [148], constrained-path priors [149]) restore feasibility but introduce a variational bias tied to the trial state. Thus, QMC-based methods remain challenging for generic doped Hubbard models.

Density Matrix Renormalization Group

DMRG [36, 38, 40] is a variational optimization over matrix product states (MPS), iteratively solving the ground-state eigenvalue equation under a rigorously defined canonical form. This gauge structure yields stable contractions, well-conditioned local updates via singular value decomposition (SVD), and clear truncation diagnostics (discarded weight), enabling highly accurate ground states in 1D and on quasi-1D geometries (cylinders). In 2D, however, an MPS must snake through the lattice [37, 150, 151]. For a cylinder of circumference L_y, an entanglement cut along y typically obeys an area law $S \sim \alpha L_y + \cdots$ [152–154]. An MPS with bond dimension D has an entanglement capacity across any single cut bounded by $S_{\text{MPS}} \leqslant \log D$, independent of L_y [155]. Consequently, a fixed-D MPS cannot reproduce the required scaling — one requires D that grows exponentially with width, $D \sim e^{L_y}$. Practically, this confines high-accuracy studies to limited widths, and actual simulations with different widths yield inconsistent outcomes [31, 156–159]. Also, numerical costs rise steeply with width and the required D, and long-range 2D correlations can be distorted by the mismatched geometry. Despite the challenges, DMRG is free of the sign problem, handles fermions and frustration naturally, and provides precise variational bounds. Therefore, it remains a powerful tool for the Hubbard system, es-

pecially when combined with careful finite-size scaling analysis and complementary cross-checks.

Infinite Projected Entangled-Pair States

iPEPS [41,42,44–47,160] is a 2D tensor-network ansatz defined directly in the thermodynamic limit, tiling the lattice with local tensors connected according to the native geometry. This faithfully respects the 2D area law and helps avoid finite-size effects; accuracy is controlled by the bond dimension D, with contractions performed approximately via e.g., Corner Transfer Matrix Renormalization Group (CTMRG) scheme [42,58,59,161,162]. The ground state can be attained using imaginary-time evolution [43, 163] or variational methods [164, 165], with environments approximated in a mean-field manner [43] or via CTMRG or related schemes. A key limitation is the absence of a strict canonical form for PEPS: norms and environments are only approximately conditioned, which can make optimization and contraction numerically unstable, especially for gapless states with long correlation lengths (perculiarly acute for free fermions [48, 166]). However, in many regimes of interest — e.g., the Hubbard model where a Mott gap or superconducting gap is present — these instabilities are very well controlled and iPEPS achieves competitive, often state-of-the-art, energies and order parameters. These features render iPEPS an ideal tool for exploring the 2D Hubbard model under thermodynamic limit.

Other Notable Numerical Methods

Exact Diagonalization (ED) [167, 168] fully diagonalizes the many-body Hamiltonian in a finite Hilbert space (often using Lanczos/Arnoldi), yielding numerically exact spectra, eigenstates, and correlation functions on small clusters — an indispensable benchmark despite exponential scaling limiting accessible sizes. Multiscale Entanglement Renormalization Ansatz (MERA) [169–172] encodes scale invariance and critical entanglement via disentanglers and isometries, enabling efficient representations of (1D/2D) critical ground states and extraction of scaling data. Higher-Order Tensor Renormalization Group (HOTRG) [173] performs real-space coarse-graining with higher-order SVD to approximately contract 2D/3D tensor networks and compute thermodynamics/ground-state properties with polynomial cost in bond dimension. Diagrammatic [174,175] and Green's-function [33–35,176–178] approaches perform summation and stochastic sampling of many-body Feynman diagrams and provide access to spectral and thermodynamic quantities. Neural Quantum States (NQS) [179–183] represent the many-body wavefunction with flexible neural ansätze optimized by the Variational Monte Carlo (VMC) method, combining expressive function classes with stochastic sampling to target ground states within and beyond traditional Jastrow/Slater forms.

1.3 Finite-temperature Methods

At finite temperature, one aims to access thermal density matrix $\rho \propto e^{-\beta\mathcal{H}}$ by either (i) Lanczos diagonalization with small clusters [184–186], (ii) stochastic evaluation of imaginary-time path integrals (finite-T QMC), (iii) typical-state sampling of minimally entangled pure states (METTS), or (iv) tensor-network purification or evolution of the thermal density operator itself, including finite-T PEPS, tangent-space Tensor Renormalization Group (tanTRG), and eXponential Tensor Renormalization Group (XTRG). Each method has its respective strengths and limitations, and an optimal simulation may arise from a synthetic application of multiple approaches.

Finite-T Quantum Monte Carlo

Finite-T QMC samples the imaginary-time partition function — via discrete auxiliary field factorizations (DQMC/AFQMC) [28, 187, 188] or continuous-time (CT) expansions [189, 190] — to obtain thermodynamics and correlation functions with controllable statistical errors. Again, in generic doped/frustrated fermionic models, the average sign decays exponentially with system size and inverse temperature $\beta = 1/T$ [147], so the variance explodes and the lowest reachable temperature is practically limited by the sign problem.

Minimally Entangled Typical Thermal States

METTS [49, 50, 191] represents thermal physics by sampling a Markov chain of low-entanglement pure states: collapse into a product-state basis vector, imaginary-time evolve it by $e^{-\beta\mathcal{H}/2}$ with DMRG, measure, and project to seed the next sample. METTS is technically straight-foward to implement on top of a ground-state DMRG. However, statistical noise and autocorrelation require multiple independent chains of samples to reach convergence, and the entanglement growth during imaginary-time evolution limits the lowest reachable temperature [192, 193].

Finite-T PEPS (from Purification)

Finite-T PEPS [52, 53, 194–196] encodes the thermal density matrix $\rho(\beta)$ via purification (a PEPS in an enlarged physical-ancilla space evolved in imaginary time), with environments contracted by CTMRG or boundary-MPS. As discussed in the ground-state methods, PEPS achieves a faithful, native representation of 2D geometry and area-law physics, thereby reducing finite-size artifacts. However, PEPS-based finite-T schemes typically integrate along β with many small steps; at low T this leads to long evolution times and error accumulation [197], which restricts the lowest temperature they can reach for a fixed bond dimension and environment accuracy (compounded by the lack of a strict canonical form and the cost of 2D contractions).

Tangent-space Tensor Renormalization Group (tanTRG)

TanTRG [56, 198] treats the thermal evolution equation $\partial_\beta \rho(\beta) = -\mathcal{H}\rho(\beta)$ (or related formulations) within the variational tangent space of a tensor-network ansatz (MPO/PEPO), advancing β by small steps using local projections and controlled truncations. The key advantage is favorable scaling $O(D^3)$ with bond dimension D per step and a strictly local update structure that makes tanTRG efficient and stable in 1D and on cylinders. However, since it also integrates along β with many small increments to control local errors, the number of steps grows unacceptably large at low T, leading to error accumulation and limiting the lowest reachable temperature.

Exponential Tensor Renormalization Group (XTRG)

XTRG [54, 55, 60, 199, 200] builds the thermal operator as an MPO at high temperature (small β) and then doubles β iteratively via MPO squaring and compression, i.e. $\rho(2\beta) = \rho(\beta) \cdot \rho(\beta)$. This *exponential cooling* means the number of steps grows only logarithmically with the target β, so one obtains high- to ultra-low-temperature data commonly via $\leqslant 20$ iterations with consistent truncation control and built-in diagnostics from MPO entanglement. The price is heavier contractions/compressions than tanTRG, making the per-iteration complexity higher. In practice, XTRG often outperforms linear-in-β integrators when one needs reliable data over a wide temperature window down to very low T.

Remarks on Synthetic tanTRG and XTRG

From the previous description, we find that tanTRG and XTRG have their respective comfort zones. TanTRG is more efficient at moderate to high temperatures, while XTRG excels at cooling down the system rapidly at low temperatures. To leverage the strengths of both methods, we can combine them into a synthetic approach: use tanTRG to evolve from high temperature down to an intermediate temperature T_{mid}, then switch to XTRG for even lower temperatures. This hybrid approach allows us to efficiently cover a broad temperature range while maintaining accuracy and stability in the simulations.

1.4 The Rise of Artificial Intelligence

Artificial Intelligence (AI) has developed through distinct phases with symbolic reasoning dominating early AI, and statistical learning gaining prominence later, especially with the rise of machine learning (ML) and deep learning (DL) enabled by contemporaneous advances in algorithms, data availability, and computation. [201] Over the past decades, AI evolved as a toolbox for modeling complex systems: from

hand-coded rules and logical inference (akin to deterministic solvers) to probabilistic models and optimization (statistical machinery in miniature), and finally to representation learning at scale, where high-capacity function approximators discover features directly from data.

A Historical Overview

The first decades (1950s-1980s) were dominated by symbolic AI. Early systems formalized knowledge as logical statements and searched over combinatorial spaces using algorithms such as A* [202]; planning frameworks like STRIPS [203] showed that non-trivial sequences of actions could be derived from explicit rules. Expert systems [204] — canonical examples include medical diagnosis prototypes — encoded domain knowledge as human-written if-then rules combined with inference engines. These systems demonstrated impressive performance in constrained domains but were brittle: coverage gaps and rule interactions produced failure modes that were difficult to predict or repair. In parallel, perceptrons [205] introduced learnable linear decision boundaries and catalyzed the first neural optimism, but limitations in expressivity (e.g., inability to compute XOR without hidden layers) and the absence of effective training for deep networks led to cycles of enthusiasm and retrenchment. As a result, symbolic systems maintained primacy while data-driven methods matured slowly at the margins.

From the late 1980s through the 2000s, statistical learning became the center of interest. Backpropagation [206] and multilayer perceptrons re-entered the mainstream, but depth and scale were limited by data and hardware. Probabilistic approaches [207] flourished: hidden Markov models [208] defined the state of the art in speech recognition; Bayesian networks captured structured uncertainty; and the expectation-maximization algorithm [209] provided a general recipe for learning with latent variables. Kernel methods, especially support vector machines (SVM) [210, 211] and Gaussian processes [212], delivered strong generalization via convex objectives (for SVMs) and principled regularization. In natural language processing, n-gram language models with smoothing [213], maximum entropy classifiers [214], conditional random fields [215] for sequence labeling, and topic models (e.g., latent Dirichlet allocation) [216] formed a robust toolkit. Computer vision progressed largely through engineered local features (e.g. gradient and edge statistics) such as SIFT [217] and HOG [218] combined with linear or kernel classifiers; convolutional neural networks existed — LeNet [219] for digit recognition is emblematic — but had not yet generalized to unconstrained images at scale. Representation learning ideas revived in the form of autoencoders [220] and deep belief networks [221], foreshadowing a shift from hand-crafted features to learned ones, while early neural distributional word embeddings [222–224] suggested that semantic regularities could be captured geometrically.

The inflection toward modern AI arrived when computation, data, and architectures aligned. In 2012, deep convolutional networks trained on large labeled image datasets decisively outperformed traditional pipelines [225], initiating rapid progress

in image classification, detection, and segmentation. For NLP, sequential models — recurrent neural networks [226, 227] with gated architectures — enabled end-to-end machine translation via encoder-decoder frameworks with attention mechanisms [62, 228], alleviating the difficulty of modeling long-range dependencies. The decisive conceptual shift came with the Transformer architecture [61], which replaced recurrence with stacked self-attention and position encodings. The resulting parallelizable training and direct access to global context dramatically improved global awareness, providing a general mechanism for learning non-local correlations.

Modern AI for Language, Vision, and Science

In the last decade, pretraining and self-supervision have remade NLP. Continuous word vectors matured into contextual encoders that condition on both left and right context [229, 230]; masked-token objectives yielded bidirectional representations that transfer to a wide range of downstream tasks via light fine-tuning [68, 231]. In parallel, autoregressive language models [70, 232] trained to predict the next token at scale demonstrated emergent few-shot and zero-shot capabilities: without task-specific training, they could perform translation, question answering, and code synthesis by conditioning on a handful of examples. Instruction tuning and preference optimization [233–235] transformed raw language models into aligned assistants being able to follow natural language specifications, reason over multi-step chains of thought [236], and orchestrate tools such as retrieval systems [237], calculators [238], and code interpreters [239]. Retrieval-augmented generation (RAG) [240–242] introduced an explicit memory external to model parameters, improving factuality and enabling controllable grounding in curated corpora — a particularly important feature for scientific use, where provenance matters. The net result is that text models have evolved into general language agents that can plan, decompose problems, call external resources, and produce executable artifacts.

Computer vision has experienced a parallel transformation. Residual connections [243] and normalization stabilized very deep convolutional networks, and transfer learning from large image corpora became standard practice. Generative adversarial networks [244], despite training instabilities, set new bars for photorealism and class-conditional synthesis. The recent ascendancy of diffusion [245, 246] and score-based generative models [247] has combined likelihood-based training with stable optimization to deliver high-fidelity synthesis and controllable editing. Parameter-efficient fine-tuning via low-rank adapters (LoRA) [248, 249] has become a widely-used way to adapt large vision and language models by freezing the backbone and learning low-rank updates, often approaching full fine-tuning while training orders-of-magnitude fewer weights. Vision Transformers [74] extended attention to images, providing a unified architecture across language and vision and enabling self-supervised learning with masked-patch or contrastive objectives. Cross-modal alignment models [250] — most prominently those that jointly embed images and text — opened the door to zero-shot classification, retrieval, and multimodal assistants that can interpret and generate across modalities.

These core developments have carried over directly into scientific practice. In molecular and materials science, message-passing neural networks can learn interatomic potentials that approach first-principle accuracy in many regimes while requiring far lower computational cost [251, 252]; active learning selects new configurations on the fly to cover gaps in phase space [253]. Generative models now propose molecules, catalysts, and crystal structures conditioned on desired properties [78, 254], supporting inverse design as a practical, iterative optimization process. Across imaging modalities — MRI, CT, PET, and computational microscopy — deep learned priors and unrolled optimization schemes couple physics-based forward models with data-driven regularization [255–257], often yielding higher-quality reconstructions at lower dose or faster acquisition. In high-energy physics and astrophysics, learned surrogates accelerate simulation, triggers, and event reconstruction [258, 259]; in gravitational-wave astronomy, neural models assist denoising and real-time signal characterization [260, 261]. Finally, neural operators learn mappings between function spaces, providing fast solvers for parametric partial differential equations and supporting rapid design exploration in fluid dynamics [262, 263].

Recent progress has also reframed scientific workflows around language-centric interfaces. Large language models fine-tuned for code [264] help generate and refactor analysis scripts, interface with simulation packages, and enforce unit tests; coupled with retrieval [241] over laboratory notebooks, instrument documentation, and literature, they act as copilots that reduce cognitive load and help maintain reproducibility. Tool-use agents [265] coordinate these components: they query databases, schedule parameter sweeps, invoke differentiable simulators, and summarize results for human inspection. When combined with Bayesian optimization and robotic platforms, such systems partially close the loop on experimental design, enabling semi-autonomous exploration under human oversight [266].

Why Modern AI is Powerful?

Several cross-cutting themes explain why this modern toolkit is effective. First, self-supervision [267, 268] shifts the burden from scarce labels to abundant raw data by posing pretext tasks — predict the next token, reconstruct masked content, or denoise corrupted inputs — whose solutions require learning transferable structure. Second, scaling laws [269] observed empirically relate model performance to parameter count, dataset size, and compute, providing guidance [270] on when additional resources are likely to yield returns and when data quality or objective design becomes the limiting factor. Third, alignment to human intent through instruction tuning and preference optimization [234, 235] has become increasingly important for reliable deployment; in scientific contexts, this can be complemented by retrieval for verifiability and explicit uncertainty quantification [271, 272] via ensembling, calibration, or Bayesian surrogates. Finally, rigorous evaluation [231, 273, 274] now emphasizes domain-specific probes, robustness and out-of-distribution tests, and end-to-end measures of scientific utility, ensuring that apparent improvements correspond to real, usable capability rather than benchmark overfitting.

Cutting-edge AI for Quantum Physics

ML/AI has become productive for identifying phases and phase transitions in many-body quantum systems [275], especially when order parameters are hidden or topological. In the supervised setting, classifiers trained on Monte-Carlo or tensor-network data learn decision boundaries that recover conventional phase diagrams and extrapolate to unseen parameters [276]. Unsupervised methods — principal component analysis (PCA) [277, 278] and learning-by-confusion [279] protocols — locate transition points via changes in manifold geometry without labels, while information-theoretic objectives highlight candidate collective variables that behave like emergent order parameters. For topological phases, unsupervised manifold learning can uncover signatures for winding-number sectors and identify candidate transitions [280], providing interpretable results with reduced manual feature engineering. Best practice in this literature is to couple learning with physics priors (symmetries, conservation laws, gauge constraints), quantify finite-size effects via scaling collapses, and validate against independent observables, so that *discovery* reflects underlying physics rather than dataset quirks.

Neural approaches to quantum many-body physics connect these ideas to quantum hardware and many-body simulation along two complementary tracks. First, classical neural parameterizations of quantum states — neural quantum states (NQS) such as Restricted Boltzmann Machines (RBM) [281], autoregressive flows [183, 282, 283], and fermionic networks [284] — serve as expressive variational ansätze for ground and, in some formulations, thermal states, enabling stochastic VMC training [285] directly on wavefunction amplitudes and accommodating symmetries, fermionic signs, and gauge constraints; these ansätze integrate naturally with tensor-network preconditioners [286] and can act as amortized solvers across Hamiltonian families. Second, error-mitigation techniques combine classical learning with parameterized quantum circuits. Classical or hybrid ML surrogates can approximate noise channels and thus support probabilistic error cancellation [287, 288] as well as model-based zero-noise extrapolation [289]. In parallel, parameterized quantum circuits (often termed quantum neural networks, QNNs) are used as variational ansätze and discriminators designed to be robust against specific noise patterns or to assist in data-driven readout and crosstalk calibration [290]. Generative surrogates trained on paired noisy/ideal data [291–293] can act as denoisers for expectation values, while QCNN-style architectures [294] have been proposed primarily for phase recognition and, in some exploratory work, as building blocks for shallow error-detection or mitigation schemes [295], supporting co-optimization of task fidelity and robustness. Ultimately, the success of these approaches will be judged not by benchmark scores but by the new capabilities they unlock — robust identification of phases, scalable many-body ansätze, and trustworthy quantum computations under realistic noise — at the intersection of AI and quantum physics.

Chapter 2
Models, Methods and Algorithms

This chapter introduces the Hubbard model (together with its derivatives), several highly-effective tensor network techniques, and the design and interpretation of the transformer architecture. Altogether, these methods and algorithms provide a powerful toolbox for the numerical simulation and analysis of strongly correlated electron systems across a wide range of charge doping and temperature.

This chapter is organized as follows. Section 2.1 introduces the minimal and extended Hubbard model, together with the derived t-J model and Heisenberg model. These derivation offers a basic theoretical background for the exotic phases that may emerge in the Hubbard model. Section 2.2 covers the basics of crucial tensor operations, which is essential for understanding and implementing the follow-up tensor network techniques. Section 2.3 discusses the graphical notations of tensors, the incorporation of fermionic statistics and symmetries, and presents 1D and 2D Tensor Network State (TNS) ansätze, including Matrix Product State (MPS) and Projected Entangled-Pair State (PEPS). Section 2.4 discusses several optimization algorithms used in tensor network simulations, including the Density Matrix Renormalization Group (DMRG), eXponential Tensor Renormalization Group (XTRG), and the imaginary time evolution for the ground state search of infinite PEPS (iPEPS). Finally, section 2.5 introduces the design strategy and variants of the AI architecture based on transformer, along with the attention scheme and the controversy surrounding its interpretability.

2.1 Hubbard Model and Its Derivatives

In section 1.1 we derived a single-band description of the CuO_2 planes by downfolding the three-band (Emery) model to the Zhang-Rice singlet manifold. Historically, though, the Hubbard model itself was introduced earlier by Hubbard, Gutzwiller, and Kanamori in the 1960s [26, 296–298] to describe correlation-driven (Mott) insulating behavior in narrow-band transition-metal oxides [299–301] — not specifically for cuprates. After the discovery of high-T_c superconductors, it was soon realized

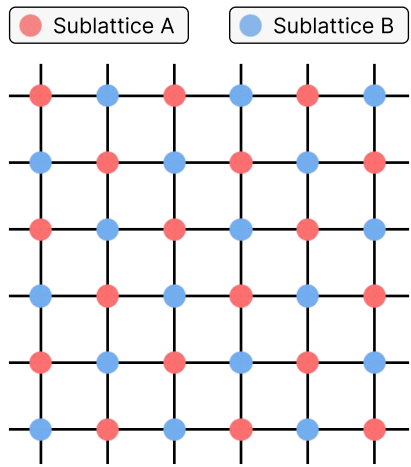


Fig. 2.1 Schematic of a 2D bipartite square lattice with two sublattices A and B. Particle-hole symmetry is preserved (up to an additive constant and a shift in chemical potential) when the hopping only occurs between different sublattices.

that the parent compounds of cuprates are also Mott insulators, and thus the same one-band framework proved useful as an effective model for the CuO_2 planes, with parameters reflecting copper-oxygen chemistry.

For our purpose, we consider the Hubbard model on a 2D square lattice. The Hubbard Hamiltonian derived from cuprates follows Eq. (1.4) with the chemical potential term. For the convenience of later discussions, we separate the full Hamiltonian into a *bare* Hamiltonian and a chemical potential term as

$$\mathcal{H} = -\sum_{i,j,\sigma} t_{ij} \left[c^\dagger_{i\sigma} c_{j\sigma} + \text{h.c.} \right] + U \sum_i n_{i\uparrow} n_{i\downarrow}, \quad \mathcal{H}_{\text{full}} = \mathcal{H} - \mu \sum_{i,\sigma} n_{i\sigma}. \quad (2.1)$$

A discrimination between $\mathcal{H}$ and $\mathcal{H}_{\text{full}}$ can be vital in numerical simulations. One can understand $\mathcal{H}$ as the Hamiltonian for a designated number of electrons (canonical ensemble), while $\mathcal{H}_{\text{full}}$ allows the electron number to fluctuate by tuning the chemical potential μ (grand canonical ensemble).

Eq. (2.1) includes hopping amplitudes t_{ij} between any two lattice sites i and j. In practice, we often truncate the hopping to a few leading terms, e.g., nearest-neighbor (NN) t, next-nearest-neighbor (NNN) t', and third-nearest-neighbor (3NN) t'' hoppings etc. The Hamiltonian of NN (minimal) Hubbard model thus reads

$$\mathcal{H}_{\text{full}} = -t \sum_{\langle i,j \rangle,\sigma} \left[c^\dagger_{i\sigma} c_{j\sigma} + \text{h.c.} \right] + U \sum_i n_{i\uparrow} n_{i\downarrow} - \mu \sum_{i,\sigma} n_{i\sigma}, \quad (2.2)$$

where $\langle i, j \rangle$ denotes NN sites. The first term describes the kinetic energy of electrons hopping between NN sites with amplitude t, the second term is the on-site Coulomb repulsion U penalizing double occupancy, and the last term controls the electron

density via the chemical potential μ. This minimal model respects a particle-hole symmetry (up to an additive constant and a shift in chemical potential) defined by

$$\mathcal{P}_{\text{ph}}: \quad c_{i\sigma} \;\mapsto\; \eta_i\, c_{i\sigma}^{\dagger}. \tag{2.3}$$

where $\eta_i = +1$ (-1) for site i on sublattice A (B) as shown in Fig. 2.1. Under this transformation, the kinetic term is invariant because hopping only occurs between different sublattices, thereby the extra fermionic sign is cancelled by $\eta_i\eta_j = -1$:

$$c_{i\sigma}^{\dagger}c_{j\sigma} \;\mapsto\; \eta_i\eta_j\, c_{i\sigma}c_{j\sigma}^{\dagger} = -c_{i\sigma}c_{j\sigma}^{\dagger} = c_{j\sigma}^{\dagger}c_{i\sigma}. \tag{2.4}$$

Also, it is straight-forward to verify that

$$n_{i\sigma} = c_{i\sigma}^{\dagger}c_{i\sigma} \;\mapsto\; 1 - n_{i\sigma}. \tag{2.5}$$

Therefore, the full Hamiltonian transforms as

$$\mathcal{H}_{\text{full}} \;\mapsto\; \mathcal{H}_{\text{full}} + (2\mu - U)\sum_{i,\sigma} n_{i\sigma} + N_s(U-\mu), \tag{2.6}$$

where N_s is the total number of lattice sites. At half-filling ($\langle n_i\rangle = 1$), the chemical potential is fixed at $\mu = U/2$ to preserve strict particle-hole symmetry. Away from half-filling, the particle-hole transformation maps a hole-doped system ($\langle n_i\rangle < 1$) to an electron-doped system ($\langle n_i\rangle > 1$) with a shifted chemical potential $\mu' = U - \mu$. Some literatures may introduce an explicit particle-hole symmetric form of the Hubbard model by rewriting the Hamiltonian as

$$\begin{aligned}\mathcal{H}_{\text{full}} = &-t\sum_{\langle i,j\rangle,\sigma}\left[c_{i\sigma}^{\dagger}c_{j\sigma} + \text{h.c.}\right] + \text{const.}\\ &+ U\sum_{i}(n_{i\uparrow} - \frac{1}{2})(n_{i\downarrow} - \frac{1}{2}) - \tilde{\mu}\sum_{i,\sigma}(n_{i\sigma} - \frac{1}{2}),\end{aligned} \tag{2.7}$$

where $\tilde{\mu} = \mu - U/2$ is the shifted chemical potential. However, we will stick to the form in Eq. (2.1) for consistency with common practice in the community.

A crucial consequence of the particle-hole symmetry is that the phase diagram of the minimal Hubbard model is symmetric about half-filling, which evidently contradicts with the experimentally observed asymmetry between hole-doped and electron-doped cuprates. To break this artificial symmetry, one can include longer-range hoppings such as the NNN t' terms. The extended Hubbard model thus becomes

$$\begin{aligned}\mathcal{H}_{\text{full}} = &-t\sum_{\langle i,j\rangle,\sigma}\left[c_{i\sigma}^{\dagger}c_{j\sigma} + \text{h.c.}\right] + U\sum_{i} n_{i\uparrow}n_{i\downarrow}\\ &- t'\sum_{\langle\langle i,j\rangle\rangle,\sigma}\left[c_{i\sigma}^{\dagger}c_{j\sigma} + \text{h.c.}\right] - \mu\sum_{i,\sigma} n_{i\sigma},\end{aligned} \tag{2.8}$$

where $\langle\langle i,j\rangle\rangle$ denotes NNN sites. The NNN hopping t' connects sites on the same sublattice, thus the particle-hole transformation in Eq. (2.3) will generate an extra minus sign in the kinetic term as

$$c^\dagger_{i\sigma}c_{j\sigma} \;\mapsto\; \eta_i\eta_j\, c_{i\sigma}c^\dagger_{j\sigma} = +c_{i\sigma}c^\dagger_{j\sigma} = -c^\dagger_{j\sigma}c_{i\sigma}. \tag{2.9}$$

Therefore, the NNN hopping term changes sign under the particle-hole transformation, breaking the symmetry between hole-doped and electron-doped systems. For cuprates, Density Functional Theory (DFT) calculations suggest that the ratio of NNN and NN hopping t'/t is typically negative ($t'/t < 0$) for hole-doped compounds and positive ($t'/t > 0$) for electron-doped ones [302, 303]. The inclusion of t' not only breaks particle-hole symmetry but also influences the stability of various competing phases, such as antiferromagnetism, superconductivity, and the charge/spin orders [48, 156, 158, 159, 304, 305].

At large U/t near half-filling, the double occupancy is energetically suppressed, and the low-energy physics of the Hubbard model can be effectively described by integrating out the high-energy degrees of freedom associated with double occupancy. Concretely, view the hopping term as a perturbation to the interaction term, and decompose the kinetic part into three contributions

$$\mathcal{T} = \mathcal{T}_0 + \mathcal{T}_1 + \mathcal{T}_{-1}, \tag{2.10}$$

where $\mathcal{T}_0, \mathcal{T}_1$, and $\mathcal{T}_{-1}$ denote the hopping processes that keep, increase, and decrease the number of doubly occupied sites by one, respectively. At large U limit, the principal low-energy correction comes from the second-order superexchange process $\mathcal{T}_{-1}\mathcal{T}_1/U$ that virtually creates and annihilates a doublon-holon pair on NN sites. The intermediate state with a doublon costs an energy U, leading to a perturbation amplitude of $1/\Delta E \approx 1/U$. Therefore, the leading order correction becomes

$$-\frac{1}{U}\mathcal{T}_{-1}\mathcal{T}_1 = \frac{4t^2}{U}\sum_{\langle i,j\rangle}\left(\mathbf{S}_i\cdot\mathbf{S}_j - \frac{1}{4}n_in_j\right), \tag{2.11}$$

where $\mathbf{S}_i = \frac{1}{2}c^\dagger_{i\alpha}\boldsymbol{\sigma}_{\alpha\beta}c_{i\beta}$ is the spin operator on site i with $\boldsymbol{\sigma}$ being the Pauli matrices, and $n_i = n_{i\uparrow} + n_{i\downarrow}$. A detailed derivation of Eq. (2.11) involves projecting the Hamiltonian onto the subspace with no double occupancy via Schrieffer-Wolff transformation [306], which we omit here for brevity. The effective Hamiltonian in the large U/t limit is known as the t-J model [307, 308]:

$$\mathcal{H} = -t\sum_{\langle i,j\rangle,\sigma}\left[\tilde{c}^\dagger_{i\sigma}\tilde{c}_{j\sigma} + \text{h.c.}\right] + J\sum_{\langle i,j\rangle}\left[\mathbf{S}_i\cdot\mathbf{S}_j - \frac{1}{4}n_in_j\right], \tag{2.12}$$

where $J = 4t^2/U$ is the superexchange coupling, and $\tilde{c}_{i\sigma} = c_{i\sigma}(1-n_{i\bar{\sigma}})$ (where $\bar{\sigma}$ denotes the opposite spin) is the electron annihilation operator projected onto the subspace with no double occupancy.

At half-filling, the kinetic terms vanish due to the no-double-occupancy constraint, and the t-J model reduces to the Heisenberg model (up to an additive constant)

$$\mathcal{H} = J \sum_{\langle i,j \rangle} \mathbf{S}_i \cdot \mathbf{S}_j, \tag{2.13}$$

which describes a quantum antiferromagnet with NN superexchange coupling $J > 0$. This provides a natural explanation for the experimentally observed AFM order and insulating behavior in undoped cuprates. Upon doping, the kinetic terms in the t-J model become active, potentially leading to high-T_c superconductivity and other exotic phases.

2.2 Basics of Tensor Analysis

The concept of *tensor* appears ubiquitously across many disciplines of physics, mathematics, and computer science, yet its precise meaning varies depending on the context. In high energy or gravitational theories, a *tensor* is essentially a *multi-linear map* that acts on vectors and covectors and yields a scalar. Formally, a **tensor** of **type** (p, q) (or **order**-$(p+q)$) defined over a vector space V is a multi-linear map that maps p vectors from the dual space V^*, and q vectors from V, to the underlying field (usually the real or complex number field). This definition is coordinate-independent: the tensor itself is an intrinsic geometric object, while the components $T^{\mu_1 \cdots \mu_p}{}_{\nu_1 \cdots \nu_q}$ arise only after a choice of basis. Operations such as index raising, lowering, or contraction are defined through the natural pairings between V and its dual and are invariant under coordinate transformations. In general relativity, such objects encode physical quantities like the metric tensor or curvature tensor, whose transformation properties guarantee covariance.

In the context of tensor network (TN) methods in many-body physics, however, the concept *tensor* acquires a more concrete and computationally accessible meaning. Here, a **tensor** is a *multi-dimensional array of (real or complex) numbers* whose entries correspond to components of a multi-linear map expanded in a basis of the many-body Fock space. Each index of the tensor carries a well-defined *dimension*[1], representing either a physical Hilbert space (such as a spin degree of freedom) or an auxiliary bond space that connects neighboring tensors in a network. Whereas the geometric notion of variance (covariant or contravariant indices) plays a central role in differential geometry, it is largely absent in the TN setting, since the underlying vector spaces are finite-dimensional Hilbert spaces equipped with canonical inner products, providing a natural isomorphism between the vector space and its dual. What matters instead is the *topological structure* — how tensors are connected together to form a network representing a physical state.

[1] This indicates that the multi-linear map acts on disparate vector (Hilbert) spaces, in contrast with the scenario in high energy physics where all (dual) vector spaces are naturally isomorphic.

In machine learning and computer science, the term **tensor** is often used in an even more relaxed sense, referring simply to a higher-order numerical array. In this language, vectors (1D arrays) are tensors of order-1, matrices (2D arrays) are tensors of order-2, and so forth. Each index may have its own range or dimension, so a tensor is characterized not only by its order but also by its **shape**, denoted as a tuple $(d_1, d_2, \ldots, d_n)$, where d_k specifies the dimension of the k-th index. Two tensors of the same order may therefore differ in shape if their index dimensions are not identical. In this computational viewpoint, tensors are concrete data structures manipulated through well-defined array operations such as reshaping, broadcasting, and contraction; and the emphasis lies on efficiency and implementation rather than geometric interpretation. Despite the differences in terminology, these views are mathematically compatible: once a basis is fixed in the abstract multi-linear definition, the tensor is completely specified by an array of its components. Conversely, any multi-dimensional array may be regarded as the coordinate representation of some multi-linear map.

Disambiguation: the *Rank* of a Tensor

Because of the coexistence of these perspectives, it is important to clarify terminology when moving between contexts. A common source of confusion lies in the use of the term *rank*. In much of the TN literature, many authors loosely refer to the number of indices of a tensor as its rank. However, in the broader linear algebraic sense, *rank* measures *linear independence*. To avoid ambiguity, in this work we adopt the convention that the **order** (also called *degree*) of a tensor denotes the number of indices it possesses, while **rank** refers to the dimension of the linear span of its components after appropriate reshaping.[2] Thus, a vector is an order-1 tensor, a matrix an order-2 tensor, and so on, regardless of its rank in the algebraic sense. When one speaks of the rank of a matrix, one refers to the number of linearly independent rows or columns, i.e. the dimension of its image as a linear map. For higher-order tensors, this concept generalizes by first *unfolding* or *matricizing* the tensor — grouping certain indices together as *rows* and the remaining indices as *columns* — and then computing the ordinary matrix rank of the resulting object. Different unfoldings generally produce different ranks, leading to the so-called *multi-linear ranks*. These ranks provide insight into the complexity and correlations encoded in the tensor, which is crucial for critical techniques like Low-Rank Approximation (LRA) for TN, and Low-Rank Adaptation (LoRA) in the context of AI architectures.

Basic Tensor Operations

As multi-linear objects, tensors admit the usual operations of addition and scalar multiplication: the sum of two tensors with identical index structure produces another

[2] However, in our publications in Chapter 3, we still follow the convention in TN community to refer to the number of indices as the rank.

tensor of the same type, and multiplying by a scalar simply rescales every component. Beyond these linear operations, tensors can be *multiplied* to form higher-order tensors through the **tensor product**. The tensor product between two tensors A and B creates a new tensor whose indices are the union of those of A and B. In component form, if $A_{i_1 i_2 \cdots i_m}$ is of order-m and $B_{j_1 j_2 \cdots j_n}$ is of order-n, their tensor product $(C = A \otimes B)$ is an order-$(m+n)$ tensor with components

$$C_{i_1 i_2 \cdots i_m j_1 j_2 \cdots j_n} = A_{i_1 i_2 \cdots i_m} B_{j_1 j_2 \cdots j_n}. \tag{2.14}$$

This operation extends the familiar *outer product* of vectors and generalizes the *Kronecker product* of matrices to arbitrary orders. It preserves multi-linearity, associativity, and distributes over tensor addition, thereby endowing the collection of tensors with an algebraic structure that mirrors the construction of composite physical systems.

A more intricate operation is the **tensor contraction**, which generalizes matrix multiplication to higher-order tensors. Contraction involves summing over one or more pairs of *shared indices* (which implies identical dimensions), reducing the total order of the contracted tensor. For example, contraction of index k between tensors A_{ijk} and B^{klm} yields a new tensor

$$C_{ij}{}^{lm} = \sum_k A_{ijk} B^{klm}. \tag{2.15}$$

Conceptually, contractions represent compositions of multi-linear maps. Thus, the contracted pair of indices should be one covariant (lower) and one contravariant (upper), but in the TN context, this distinction is often ignored due to the natural isomorphism between the spaces involved. The cost of a contraction grows polynomially with the dimensions of the contracted indices, so choosing an optimal contraction order is essential for computational efficiency. A special case of contraction is the **trace** or **partial trace**, obtained by contracting one or more pairs of indices in a single tensor.

A convenient simplification of contraction notation is the **Einstein summation convention**, which omits the summation symbol when an index appears both as a subscript and a superscript in a tensorial formula. For instance, Eq. (2.15) can be succinctly written as

$$C_{ij}{}^{lm} = A_{ijk} B^{klm}. \tag{2.16}$$

This convention streamlines expressions and highlights the underlying multi-linear structure. In many literatures, the Einstein summation convention is further relaxed to allow repeated indices in the same vertical position (both subscripts or both superscripts) to imply summation, as long as the index dimensions match. However, we insist on the original convention to respect the geometric significance of covariant and contravariant indices.

Principal Component Analysis and Low-Rank Approximation

A high-order tensor can be cast into a matrix across a chosen cut. Given an order-p tensor T with shape $(d_1, \ldots, d_p)$, select a bipartition of the index set into groups $\mathcal{I}$ and $\mathcal{J}$ and merge them into composite indices. The outcome is a matrix $M \in \mathbb{C}^{|\mathcal{I}|\times|\mathcal{J}|}$ with $|\mathcal{I}| = \prod_{i\in\mathcal{I}} d_i$ and $|\mathcal{J}| = \prod_{j\in\mathcal{J}} d_j$. This *matricization* enables the application of well-established linear algebra techniques to study the properties of the original tensor.

For tensors unfolded into a square matrix, the eigen decomposition gives the natural basis in which a linear map acts by simple rescaling. If matrix H is Hermitian, the spectral theorem furnishes an orthonormal basis U with

$$H = U\Lambda U^{\dagger}, \tag{2.17}$$

so each eigenvector spans an invariant direction and each real eigenvalue sets its dilation. In the TN context, this appears most transparently in transfer matrices: leading eigenvectors encode dominant long-range structure, while eigenvalue gaps set correlation lengths and relaxation scales. Numerically, the dense cost for a full spectrum is $O(n^3)$ for an $n\times n$ input, whereas Krylov variants (Lanczos for Hermitian, Arnoldi for non-Hermitian) target only a few extremal eigenpairs using matrix-vector products and thus may scale much better than $O(n^3)$ if only a small fraction of principal components are required.

The singular-value decomposition (SVD) is the canonical factorization for tensors unfolded to rectangular matrices. Any matrix $M \in \mathbb{C}^{m\times n}$ admits a decomposition

$$M = U\Sigma V^{\dagger}, \tag{2.18}$$

where $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ are unitary, and $\Sigma \in \mathbb{R}^{m\times n}$ is diagonal with nonnegative singular values. The columns of U and V form orthonormal bases for the domain and codomain, respectively, while the singular values quantify the strength of each mode. In computation, modern SVD routines reach this form via norm-preserving unitary bidiagonalization followed by a small-core diagonalization; for a dense $m \times n$ matrix with $m \geqslant n$ the leading cost is $O(mn^2)$.

Beyond diagnosis, the SVD yields a principled compression. The Eckart-Young-Mirsky theorem asserts that truncating to the k largest singular values produces the unique best rank-k approximation in both spectral and Frobenius norms, with error

$$\|M - M_k\|_F^2 = \sum_{i>k} s_i^2. \tag{2.19}$$

In the TN setting this *truncated SVD* furnishes a basic LRA: after matricizing across a physically meaningful cut, one retains only the dominant singular directions, thereby limiting the effective rank (and hence the bond dimension) across the very cut. The discarded weight quantifies the loss of entanglement and provides a direct measure of observable errors.

A further advantage of the SVD lies in the reduction of a high-order tensor into a series of lower-order tensors. Concretely, one may interpret $U\Sigma$ and $V^\dagger$ as two tensors joined by a new *auxiliary* index whose dimension equals the retained rank. Iterating such binary splits yields chain- or tree-structured factorizations. If used wisely, this can dramatically lower peak operation arities, memory footprint, and intermediate tensor sizes — often turning intractable follow-up SVDs into manageable ones — without sacrificing physical fidelity.

2.3 Tensor Network Techniques

Here we introduce tensor networks as a compact language and toolbox for many-body quantum systems. We begin with a graphical notation that turns multi-linear algebra into diagrammatic manipulations and facilitates transparent contraction patterns. We then incorporate fermionic statistics and global symmetries directly in the diagrams — via fermionic swap gates and symmetry-adapted block decomposition — to ensure physically consistent and computationally efficient representations. Building on this foundation, we present MPS and MPO as controlled 1D ansätze with canonical forms and principled compression; and outline PEPS as native higher-dimensional ansätze, emphasizing their structural features and expressive power for simulations directly in the thermodynamic limit.

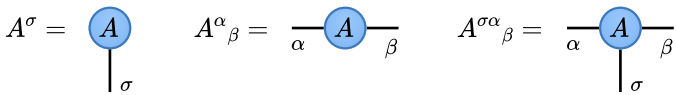
2.3.1 Graphical Notation of Tensors

Graphical notation of tensors provides a practical language for tensor networks. Algebraic expressions for even modest contractions quickly become unreadable — indices proliferate and parentheses obscure the computational flow. Diagrams concretize this structure. By showing only connectivity — which pair of indices are contracted with each other — and leaving index names implicit unless needed, a picture exposes the topology of the computation at a glance. This brings about convenience in planning efficient contraction orders and in relating algebraic manipulations to physical content (which bond encodes which correlations).

In the Penrose-style convention we adopt, a tensor is drawn as a small node with one external line (also called a leg or bond) per index; the node can be depicted as an oval, polygon, or other closed shapes derived from these; each leg may carry, if needed, a label for the corresponding index or its dimension. In this thesis, we follow the common practice of representing state-related tensors using round shapes (circles or shapes derived from circles), while reserving straight-edged polygons (squares, rectangles, etc.) for operators or environments.

Figure 2.2 illustrates the diagrammatic representation of a vector (order-1 tensor), a matrix (order-2 tensor), and a general order-3 tensor. The geometric arrangement of legs on the page has no algebraic meaning; however, for fermionic tensors, the order

of the legs matters — permuting the legs may lead to a negative sign. If symmetries are invoked, each leg may carry an arrow indicating the flow of quantum numbers. Details will be discussed in Section 2.3.2.



(a) Vector (b) Matrix (c) Tensor

Fig. 2.2 Diagrammatic expression of a vector, a matrix and a tensor of order-3. The node is depicted as a circle here, and each external line (usually called leg or bond) represents an index of the tensor.

Contraction is depicted by joining legs. Two tensors connected by a line share an index, and summation over that index is implied. Thus, if two order-3 tensors are drawn with one pair of legs connected, the result is an order-4 tensor whose remaining free legs are precisely the unconnected ones, as depicted in Fig. 2.3. This graphical calculus is faithful: every legal diagram corresponds to a well-defined algebraic contraction, and different planar embeddings of the same connectivity (together with fermionic encoding introduced in Section 2.3.2 where applicable) represent the same multi-linear map.

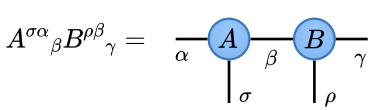


Fig. 2.3 Diagrammatic expression of a contraction operation on index β. Contraction is represented by a connected line. Indices are labelled for clarity here, but will usually be omitted in practice.

Although a tensor network diagram uniquely specifies the contraction topology, it leaves open how to carry out the contractions involved, and the choice of contraction order has a great impact on the cost without altering the outcome. The overall computational complexity depends on the sizes of intermediate tensors created along the way. Since contraction is associative but not cost-commutative, contracting two indices that produce a small intermediate tensor can be far cheaper than first forming a large one that later collapses. Therefore, planning an optimal contraction sequence is crucial for efficiency. In mature TN algorithms, optimal orders have been established manually, while in generic situations, finding the optimal contraction order can be NP-hard [309]. Therefore, one has to rely on various heuristic algorithms [310,311] to find (near-)optimal orders, often leveraging graph-theoretic insights into the network structure.

2.3.2 Fermionic Statistics and Symmetry

Fermionic tensor networks must combine a local tensor calculus with anti-commutation. If diagrams were treated as bosonic, line crossings or index permutations could flip signs. A practical fermionic formalism records statistics locally: each index carries a $\mathbb{Z}_2$ parity label; exchanges of odd objects are handled by an explicit swap gate; and genuinely odd operators are marked as such. Hence, statistics reduce to checkable, diagram-local rules independent of page layout. In parallel, we endow indices with symmetry charges or irreducible representations (irreps), and annotate them with arrows indicating the flow of quantum numbers. Symmetry-aware tensors then become block-sparse objects whose non-zero blocks satisfy charge/irrep fusion constraints; for non-Abelian groups, these constraints are mediated by Clebsch-Gordan (CG) coefficients, which we treat as explicit, small intertwiners. This bookkeeping guarantees the correct overall fermionic sign structure under planar deformations, and yields substantial speedups by contracting only symmetry-matching blocks.

Parity Conserving (Even) Tensors

Most states and evolutions in fermionic lattice systems conserve fermion parity. To implement this feature in the tensor network, we endow every index α with a parity label $p(\alpha) \in \{-1, 1\}$ in addition to its dimension. A tensor A with indices $\alpha_1, \alpha_2, \ldots, \alpha_n$ is called **even** (parity conserving) if each non-zero component satisfies the graded selection rule

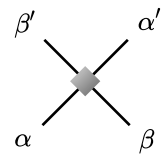
$$\prod_{i=1}^{n} p(\alpha_i) = 1. \tag{2.20}$$

Equivalently, A is block-sparse with respect to the $\mathbb{Z}_2$ charge on its indices: only blocks where the total incoming parity equals the total outgoing parity are allowed. Treating parity at the index level has two immediate benefits. Conceptually, it turns fermionic statistics into a local compatibility issue: illegal blocks are excluded rather than relying on post hoc numerical cancellations. Algorithmically, block structure reduces floating-point work and memory — only parity-allowed blocks need be stored and contracted. Contractions between even tensors preserve evenness. Here, we emphasize that $\mathbb{Z}_2$ is not merely optional symmetry bookkeeping but the minimal structure that faithfully encodes fermionic statistics.

Fermionic Swap Gates

When two fermionic indices are permuted, a minus sign arises if both indices carry odd parity. This anti-commutation is enforced by a local **swap gate** `SWAP` [44–47] inserted wherever two graded lines cross. Concretely, the swap gate can be represented as an order-4 tensor $\texttt{SWAP}^{\alpha'\beta'}_{\alpha\beta}$ defined by

$$\texttt{SWAP}^{\alpha'\beta'}_{\alpha\beta} = \delta^{\alpha'}_{\alpha}\delta^{\beta'}_{\beta}\texttt{SWAP}(\alpha,\beta) = \quad , \tag{2.21}$$

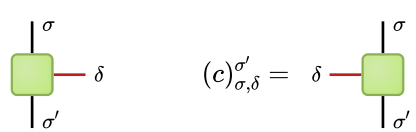


where δ is the Kronecker delta, and $\texttt{SWAP}(\alpha,\beta) = -1$ if $p(\alpha) = p(\beta) = -1$ and $+1$ otherwise. The swap gate is an involution: applying it twice gives identity. We emphasize that even though the swap gate is constructed as a tensor in the TN diagram, in actual implementation, the effect is achieved via negating odd blocks of two crossing indices — no actual contraction is performed. Diagrammatically, the swap gate is represented as a small rhombus at the crossing point of two lines, as shown in Eq. 2.21. With this construction the graphical calculus is planar and isotopy-safe: lines may be deformed freely without changing the outcome.

Parity Changing (Odd) Operators

Though the majority of the Hamiltonians in fermionic systems conserve parity, they can comprise parity-changing operators, such as single fermion creation/annihilation operators. In the graded formulation we adopt, parity-changing operators are represented by *even* tensors equipped with an additional auxiliary index of odd parity that carries a $\mathbb{Z}_2$ *parity flux*. For example, the creation and annihilation operators $c^\dagger$ and c can be translated into tensors of order-3 as

$$(c^\dagger)^{\sigma'}_{\sigma,\delta} = \qquad (c)^{\sigma'}_{\sigma,\delta} = \tag{2.22}$$



where α and α' are the physical input and output indices, respectively, and δ exemplifies the auxiliary parity flux index. Both tensors are even, satisfying

$$p(\alpha)p(\alpha')p(\delta) = 1. \tag{2.23}$$

When contracting with other tensors, the parity flux index must be connected to another parity flux index to ensure overall evenness. The auxiliary index is then routed through the network; whenever it crosses another odd index, a swap gate at the crossing contributes the appropriate -1 sign, effectively taking over the role of the Jordan-Wigner string in traditional fermionic representations. This construction ensures that the entire tensor network remains even, transforming fermionic statistics into symmetry-preserving tensors and allowing us to leverage symmetry-based implementations for tensor operations.

Symmetry in Tensor Networks

Symmetry is one of the most essential concepts in physics. In quantum many-body problems, conserved quantities (charges) generated by symmetry operators label superselection sectors; states that differ in their full set of symmetry quantum numbers cannot be coherently superposed by symmetry-preserving operations. Phase transitions are often accompanied by changes in symmetry or its realization (e.g., spontaneous symmetry breaking or topological order with emergent symmetries).

In tensor networks, enforcing global symmetries at the level of every tensor makes the entire network *block-sparse*. Each tensor index decomposes into a direct sum over irreps and an associated *multiplet* (or degeneracy) space. Contractions are permitted only when the fusion (addition) of incoming charges complies with the outgoing charge — this is the graphical *charge-flow* selection rule that prunes enormous swaths of zero elements. In practice, this yields large reductions in memory and runtime, while also improving numerical stability by ruling out symmetry-forbidden couplings. Modern libraries (e.g., QSpace [312, 313]) automate this bookkeeping and expose a clean programming interface in which algorithms can be written without explicit consideration of symmetry, yet run with full non-Abelian support under the hood.

Abelian Symmetries

For Abelian groups (e.g., $\mathbb{Z}_2$, $\mathrm{U}(1)$), all irreps are one-dimensional and commute. A basis state on an index is labeled by a charge q and a multiplet label l that distinguishes multiple states with the same charge (e.g., different orbitals carrying identical particle number). In an order-3 tensor commonly used in 1D systems, writing out the charge labels makes the block structure explicit:

$$|q''n\rangle = \sum_{ql,q'm} [A^{q}_{q'q''}]^{l}_{mn} |ql\rangle |q'm\rangle, \tag{2.24}$$

with the selection rule $q'' = q \oplus q'$ (addition mod n for $\mathbb{Z}_n$ or ordinary addition for $\mathrm{U}(1)$). Numerically, one stores only the non-zero *reduced* data blocks A for the allowed triples (q, q', q'') and omits everything else. This achieves compression by orders of magnitude in charge-conserving models (electric charge, S^z, etc.). This block-sparse structure carries over to all tensor operations, including contraction, SVD, eigen decomposition, etc., which are performed block-wise. Also, the parity structure for fermions is implemented as a special $\mathbb{Z}_2$ Abelian symmetry.

Non-Abelian Symmetries: SU(N)

For non-Abelian groups such as $\mathrm{SU}(2)$, indices carry multiplets labeled by total spin S (the irrep) and an internal magnetic component S_z within that multiplet. Crucially,

one separates *structure* from *data* using the Wigner-Eckart theorem: the *structural* part is fixed by group theory (Clebsch-Gordan objects and recoupling algebra), while the *reduced* tensors contain all variational/numerical degrees of freedom that live solely in multiplet spaces. For an order-3 tensor, this reads [314–316]

$$|q''m, q''_z\rangle = \sum_{ql q_z, qm, q'_z} [A^{q}_{q'q''}]^{l}_{mn} \cdot C^{q_z}_{q'_z q''_z} \, |ql q_z\rangle |q'm q'_z\rangle, \tag{2.25}$$

where $C^{q_z}_{q'_z q''_z} = \langle q'm, q'_z; ql, q_z | q''m, q''_z \rangle$ is the Clebsch-Gordan coefficient [317] that fuses two irreps q and q' into q''. Analogously, a symmetric operator factorizes as [314–316]

$$\langle q'l'; q'_z | \hat{O}^{\tilde{q}}_{\tilde{q}_z} | ql; q_z \rangle = (O^{[\tilde{q}]}_{qq'})^{[1]}_{ll'} \cdot C^{[\tilde{q}_z]}_{q_z q'_z} \tag{2.26}$$

with C a tabulated CGC object and $(O)^{[1]}$ the reduced data block (no q_z indices). This separation implements selection rules automatically and yields strong compression.

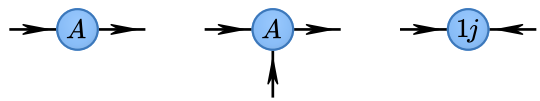
When several non-Abelian indices fuse, not only must the irreps combine to the target irrep, but some fusion channels may occur more than once (**outer multiplicity**). Implementations therefore attach a *multiplicity index* (often denoted α) in addition to the multiplet index l to uniquely label repeated channels. Careful management of fusion trees — i.e., the order in which indices are coupled — is essential for both correctness and efficiency. This fusion-tree recoupling overhead can be offloaded to **X-symbols** [316], which tabulate the Clebsch-Gordan/metric/recoupling structure into reusable connection tensors so that pairwise contractions across different fusion trees correctly resolve outer multiplicities.

The CGC/X-symbol tables are computed once per symmetry (and convention) and then cached — amortizing their cost across an entire simulation. Also, QR/SVD factorizations act *within* blocks, preserving charges and drastically shrinking the linear algebra. In practice, composite symmetries are common in models with both charge and spin degrees of freedom: e.g., $\mathrm{U}(1)_{\text{charge}} \times \mathrm{SU}(2)_{\text{spin}}$ or $\mathbb{Z}_2 \times \mathrm{SU}(2)_{\text{spin}}$. Libraries like QSpace (v4.0, now open-source) [312, 313] and TensorKit.jl [318] support various Abelian and non-Abelian symmetries and handle outer multiplicities automatically.

Graphical Features of Symmetric Tensors

In symmetric tensor diagrams, each index carries an *arrow* that encodes charge flow and duality: ingoing arrows represent ket-type (primal) indices, outgoing arrows bra-type (dual) indices. At every vertex, arrows enforce the local selection rule implied by Eqs. (2.24), (2.25), and (2.26): for Abelian groups the signed sum of charges on outgoing indices should equal that on incoming indices; for non-Abelian groups, fusion or splitting are only allowed if the corresponding fusion channel exists.

Figure 2.4 illustrates the basic diagrammatic primitives. [312] In particular, an *arrow-reversal* $1j$ symbol converts an irrep to its dual, inserting a fixed structural factor without affecting the reduced data, so that contractions only occur between



(a) Matrix (b) Tensor (c) $1j$ Symbol

Fig. 2.4 Diagrammatic expression of a matrix, a tensor of order-3 with non-Abelian symmetry, and a $1j$ symbol for arrow inversion. Each index carries an arrow indicating the flow of conserved charges.

oppositely oriented indices in accordance with the selection rules. In this calculus, a standalone vector (tensor of order-1) cannot exist as a non-trivial symmetric object: local charge conservation forces any single index carrying a non-trivial irrep to vanish. Thus, the only allowed one-index object is a trivial singlet. Consequently, symmetric tensors of interest always appear with at least two indices.

Contractions are arrow-sensitive: only indices with opposite arrow directions may be contracted directly. If two would-be contracted indices share the same orientation, one first applies the arrow-reversal $1j$ symbol to one index before performing the contraction. However, with a consistent global convention, the scenarios of same-orientation contractions can be extremely rare. A well-designed TN algorithm in most cases only needs to contract indices with opposite arrows. Exceptions arise in very specific situations such as the square root of a diagonal singular-value matrix. Therefore, one should avoid invoking $1j$ symbols unless absolutely necessary.

2.3.3 Matrix Product States / Operators

Matrix Product States (MPS) — and their operator counterparts — offer a way to study many-body quantum matter without drowning in exponential complexity. They replace a monolithic wavefunction or operator with a necklace of small, local tensors whose links carry entanglement, so structure and scale emerge from simple building blocks. The bond dimension becomes a tunable knob: increase to resolve more intricate correlations, narrow for speed and memory. In this picture, Hamiltonians, density matrices, and observables fit naturally as Matrix Product Operators (MPO), enabling us to compute expectations and evolve states while staying within the same compact framework. MPS and MPO recast many-body quantum states/operators into a compact, local tensor language — economical when correlations are simple and gracefully expandable when they grow — making large quantum systems tractable and systematically controllable.

Consider a 1D lattice of L sites, each with a local Hilbert space of dimension d. A generic many-body wavefunction $|\psi\rangle$ can be expanded in the Fock basis $\{|\sigma_1\rangle|\sigma_2\rangle\cdots|\sigma_L\rangle\}$ as

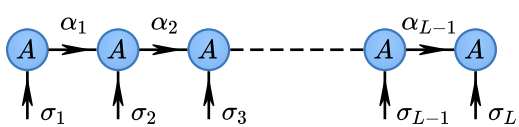
$$|\psi\rangle = \sum_{\sigma_1,\sigma_2,\cdots,\sigma_L} \Psi^{\sigma_1\sigma_2\cdots\sigma_L} |\sigma_1\rangle|\sigma_2\rangle\cdots|\sigma_L\rangle. \tag{2.27}$$

The expansion coefficients $\Psi^{\sigma_1\sigma_2\cdots\sigma_L}$ form a tensor of order L. Directly storing and manipulating this tensor is infeasible for large L due to the exponential growth of its size (d^L components). Tensor network methods provide structured factorizations of this tensor. The MPS ansatz approximates this high-order tensor as a chain of site-local tensors contracted along *auxiliary* (or *virtual*) indices:

$$\Psi^{\sigma_1\sigma_2\cdots\sigma_L} = A^{\sigma_1}_{\alpha_1} A^{\sigma_2\alpha_1}_{\alpha_2} A^{\sigma_3\alpha_2}_{\alpha_3} \cdots A^{\sigma_{L-1}\alpha_{L-2}}_{\alpha_{L-1}} A^{\sigma_L\alpha_{L-1}} \tag{2.28}$$

where each $A^{\sigma_i\alpha_{i-1}}_{\alpha_i}$ is a local tensor at site i with one physical index σ_i and two virtual indices α_{i-1} and α_i. The dimensions of the virtual indices, called the *bond dimensions* D_i, control the amount of entanglement that can be captured between different parts of the system. By adjusting these bond dimensions, MPS can efficiently represent states with limited entanglement, such as ground states of gapped 1D systems, while still allowing systematic improvement by increasing D_i.

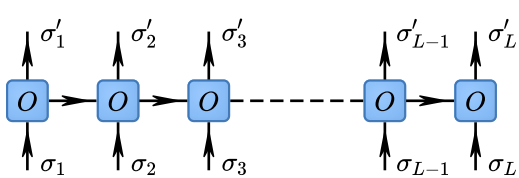
Diagrammatically, one can translate the chain of tensor contractions in Eq. (2.28) into a tensor network as



(2.29)

where we have assumed that symmetry has been incorporated, and each index carries an arrow indicating the flow of conserved charges.

An MPO can be constructed similarly to represent operators acting on the many-body Hilbert space. A many-body operator O factorizes analogously as



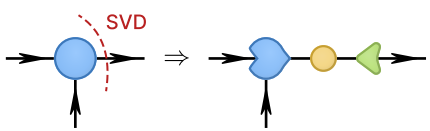
(2.30)

Hamiltonians, density matrices, and observables can often be written compactly as MPOs, enabling efficient evaluation of expectation values and evolutions within the same TN language. We note that in occasional cases (e.g., in density matrix operations), it is possible to transform an MPO into an MPS by grouping the input and output physical indices together (Choi transformation) [56]. However, this manipulation must be carried out with extra caution to avoid unconsciously invoking line crossings that introduce fermionic signs.

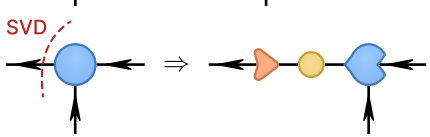
Compression and Canonical Forms

The genuine MPS encoding of the many-body wavefunction of Eq. (2.28) can be exact, with the bond dimensions $D_i = d^{\min(i,L-i)}$ growing exponentially with system size, rendering it impractical for large L. However, physical states of interest, such as ground states of local Hamiltonians, often exhibit limited entanglement that can be efficiently captured via moderate bond dimensions. To achieve this, one employs appropriate *compression* to truncate the bond dimensions while retaining the most significant correlations.

A prevalent approach for compression is the LRA via a sequence of site-wise SVDs. By reshaping the MPS tensors into matrices across a chosen bond, performing SVD, and truncating to the leading singular values, one can systematically reduce the bond dimension while controlling the approximation error. There are in general two ways of performing such an SVD:

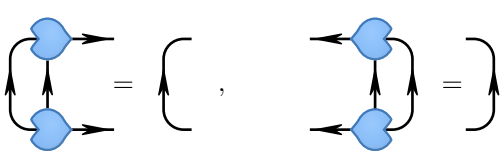


(2.31)



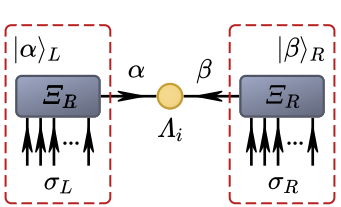
(2.32)

where we have introduced a special notation: a circle with a protruding wedge on one side and a notch in the other side represents a *normalized* MPS tensor (an *isometry*) satisfying the left- or right-normalization conditions:



(2.33)

The normalization conditions ensure that the MPS is in a canonical form, which simplifies the computation of observables and improves numerical stability. By sweeping through the chain and applying these SVDs iteratively, one can bring the MPS into a bond-canonical form with a central diagonal matrix that captures the entanglement across a specific bond, while the tensors to the left Ξ_L and right Ξ_R are left- and right-normalized, respectively:



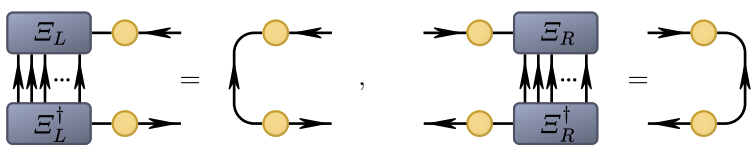
(2.34)

This bond-canonical form reveals the entanglement structure of the state explicitly. The many-body wavefunction under this form reads

$$|\psi\rangle = \sum_{\alpha} \Lambda_i^{\alpha\alpha} |\alpha\rangle_L |\alpha\rangle_R. \tag{2.35}$$

Therefore, the singular values $\Lambda_i^{\alpha\alpha}$ correspond to the Schmidt coefficients across the bond i, which quantify the entanglement between the left and right partitions of the system.

Another crucial consequence of the canonical form is the reduction of the environment to identity due to the orthonormality of the left- and right-normalized isometries:



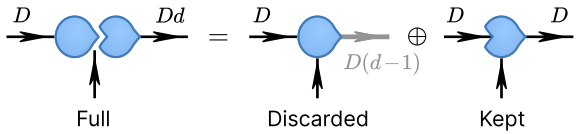
(2.36)

This property greatly simplifies the evaluation of local observables and correlation functions. In addition, Eq. (2.36) implies that the singular values $\Lambda_i^{\alpha\alpha}$ carry full information about the environmental weights for state $|\alpha\rangle_L$ and $|\alpha\rangle_R$ in a mean-field sense. Moreover, the right-hand side of Eq. (2.36) coincides with the messenger matrix in the language of belief propagation [319–322] for 1D networks. Consequently, this feature becomes critical in generalizing mean-field environments to higher-dimensional tensor networks.

Kept and Discarded Subspaces

In 1D TN formulations, the limited bond dimension defines a restricted subset of the full many-body Hilbert space. This naturally leads to a geometric decomposition of the variational space at each bond into the kept subspace, spanned by the orthonormal basis retained in the isometry, and its orthogonal complement, the discarded subspace [323]. The kept space represents the manifold itself, capturing the variational degrees of freedom already encoded in the current tensor network, whereas the discarded space collects directions along which the state can be systematically expanded or corrected. This decomposition provides a rigorous language for defining tangent-space projectors, which underpin post-MPS methods such as the time-dependent variational principle (TDVP) [324] and the construction of low-lying excitations [323]. More recently, the same idea has been extended to the Controlled Bond Expansion (CBE) [40] framework, where the discarded subspace serves as a reservoir for bond-enlarging directions beyond the original variational manifold, enabling adaptive accuracy improvements.

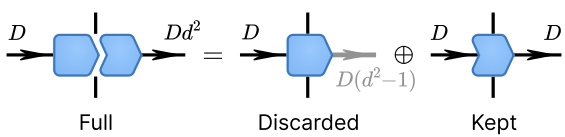
Diagrammatically, we write the decomposition of a full isometry into the kept and discarded parts as



(2.37)

where the left-hand side represents the full isometry which combines the D-dimensional space (marked by the left index) and the d-dimensional local Hilbert space into the dD-dimensional space on the right. The right-hand side decomposes this full isometry into the discarded part (left, with a protruding wedge and no notch) and the kept part (right, with both a protruding wedge and a notch). This decomposition will be frequently utilized in an efficient DMRG optimization.

Analogously, one can define the kept and discarded isometries for the MPO representation of operators as



(2.38)

where we similarly use a square with a protruding wedge and a straight edge on the other side to represent the discarded isometry, and a square with both a protruding wedge and a notch to represent the kept isometry. This decomposition plays a crucial role in an efficient exponential cooling of the thermal density matrix via the XTRG algorithm.

Traversal of 2D Lattices via Snake-Like Mapping

A common strategy to extend 1D tensor network architectures, such as the MPS or MPO, to describe 2D systems is to map the lattice sites onto a 1D chain in a prescribed order. Among various mapping schemes, the so-called *snake-like traversal* (or *serpentine ordering*) [37, 150, 151] provides a conceptually simple and computationally efficient approach. In this mapping, the sites of a rectangular $(L_x \times L_y)$ lattice are sequentially ordered along a single path that winds through the lattice in alternating directions — typically from top to bottom in one row, and from bottom to top in the next — forming a continuous, 1D sequence of sites. This procedure effectively linearizes the 2D geometry, allowing the use of standard 1D tensor-network representations and algorithms.

Mathematically, one may define the mapping function $i = f(x, y)$, which associates each site at coordinate (x, y) with a position i along the 1D chain. For the snake-like path, this is explicitly given by

$$f(x,y) = \begin{cases} (x-1)L_y + y, & \text{for } x \text{ odd}, \\ xL_y - y + 1, & \text{for } x \text{ even}, \end{cases} \tag{2.39}$$

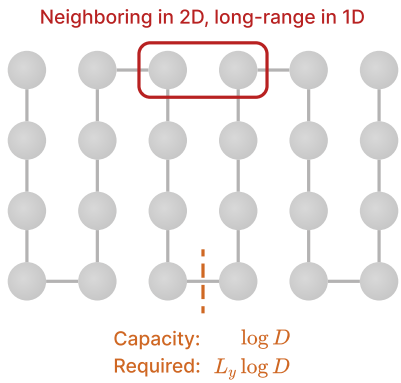


Fig. 2.5 Snake-like traversal of a 6×4 lattice mapped onto a 1D chain. Some neighboring sites in 2D become distant in the 1D ordering. Horizontal bonds can encode much less entanglement than required.

Despite its simplicity, the snake-like ordering introduces an artificial *non-locality* in the representation of physical correlations. In the original 2D geometry, neighboring sites along the horizontal direction may become distant in the 1D ordering, causing long-range couplings in the corresponding MPO or Hamiltonian representation, as shown in Fig. 2.5. This artificial increase in entanglement typically leads to a growth in the required bond dimension to faithfully represent the state, as the effective entanglement entropy between bipartitions of the 1D chain scales with the boundary length (L_y) in the 2D lattice. Consequently, while the snake mapping enables the reuse of powerful 1D algorithms such as the DMRG, its efficiency deteriorates for wider systems due to this entanglement bottleneck.

2.3.4 Projected Entangled-Pair States

Central limitations of using 1D tensor networks for 2D quantum systems include the artificial long-range structure and a lack of sufficient capacity for entanglement. PEPS circumvent this by *natively* encoding the 2D geometry: tensors are placed on the sites of the lattice and connected along all the 2D nearest-neighbor bonds. As a consequence, PEPS adhere to the 2D area law by construction. For a bipartition with boundary ∂R, the number of virtual-bond cuts scales with the boundary length $|\partial R|$, and the entanglement entropy satisfies (up to an additive constant) [153, 154]

$$S(R) \leqslant |\partial R| \log D \tag{2.40}$$

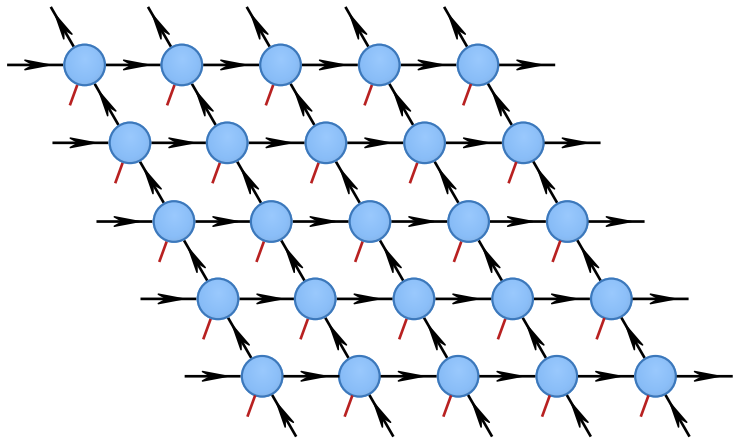


Fig. 2.6 Diagrammatic representation of a 5×5 supercell for an infinite PEPS (iPEPS) on a square lattice. Each tensor is connected to its nearest neighbors via virtual bonds (black lines), forming a 2D network that captures the entanglement structure of the system. This diagram is only schematic for fermionic systems, as many physical indices (red lines) are enclosed within the loops.

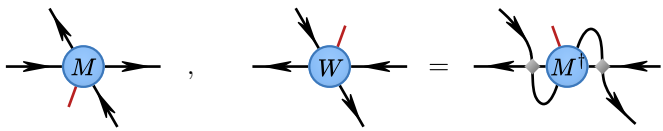
for bond dimension D. This length scaling matches the entanglement structure expected for gapped local Hamiltonians in 2D and is the key motivation for adopting PEPS as a ground-state ansatz for 2D quantum systems.

A further advantage of PEPS in two dimensions is that they naturally admit a thermodynamic-limit formulation by exploiting lattice translational symmetry. In an infinite PEPS (iPEPS), one specifies a finite supercell (unit cell) of tensors which is periodically tiled over the infinite lattice. This establishes the translational symmetry at the tensor level (or a controlled breaking thereof when larger supercells are chosen). This construction eliminates explicit finite-size and boundary effects *ab initio*, giving direct access to bulk observables and spontaneous symmetry breaking in the thermodynamic limit. Henceforth, we adopt this infinite setting exclusively and work with iPEPS throughout.

Figure 2.6 exemplifies an iPEPS on a square lattice with a 5×5 supercell. Each tensor has one physical index (red line) representing the local Hilbert space at that site, and four *virtual* (also called *auxiliary*) indices (black lines) connecting to its nearest neighbors. We emphasize, however, that this tensor network diagram is only partially legal for fermionic systems: many physical indices are enclosed within the loops and do not extend beyond the boundary of the entire network. This brings ambiguity in the ordering of the physical indices and neglects all the fermionic swap gates that would have incurred if extended outwards. Therefore, special techniques must be employed for a correct and efficient contraction of the fermionic iPEPS.

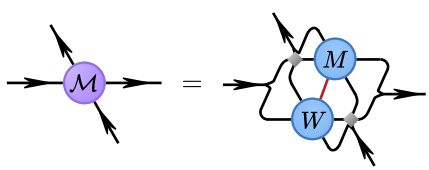
To prepare for the subsequent discussion on the contraction of iPEPS (especially fermionic iPEPS), we introduce the concept of *double-layer* tensors, which involves

contracting the physical indices of the PEPS tensors and their conjugates locally. Concretely, the order-5 PEPS tensor M and its conjugate W can be written as



(2.41)

Notice that the conjugate W does not equal the direct Hermitian conjugate $M^\dagger$ due to the extra fermionic swap gates. The double-layer *bulk* tensor $\mathcal{M}$ is then formed by contracting the physical indices of M and W:



(2.42)

Also notice the two fermionic swap gates incurred here. This double-layer tensor $\mathcal{M}$ has four virtual indices, each with a squared bond dimension D^2. In the case of evaluating expectation values of local operators, one can further define *impurity* tensors $\mathcal{M}_O$ by sandwiching the local operator O into the physical index contraction where relevant. Multi-point correlators can also be constructed analogously by inserting multiple impurity tensors at the corresponding locations. For parity-changing operators, the auxiliary parity flux index should be carefully routed through the network, maintaining the correct fermionic signs via swap gates.

After constructing all the double-layer bulk tensors, the contracted tensor network for evaluating the norm $\langle\psi|\psi\rangle$ or expectation values $\langle\psi|O|\psi\rangle$ reduces to a 2D network of these bulk tensors, as shown in Fig. 2.7. The remaining challenge is to efficiently contract this infinite 2D tensor network to obtain accurate approximations of the desired quantities.

Contraction of iPEPS via CTMRG

A prominent strategy relies on the Corner Transfer Matrix Renormalization Group (CTMRG) [42, 58, 59, 161, 162], which approximates the infinite environment surrounding a finite patch of the lattice using a set of corner and edge tensors, as marked by colored shadows in Fig. 2.7. The corner tensors C (green) capture the contributions from the four corners of the environment, while the edge tensors T (orange) represent the contributions along the edges.

These tensors are iteratively updated via directional coarse-graining steps that absorb layers of bulk tensors into the environment, followed by truncation to control the bond dimensions. This process continues until convergence is achieved, yielding an accurate approximation of the infinite environment. Detailed algorithms can be found in Refs. [315, 325].

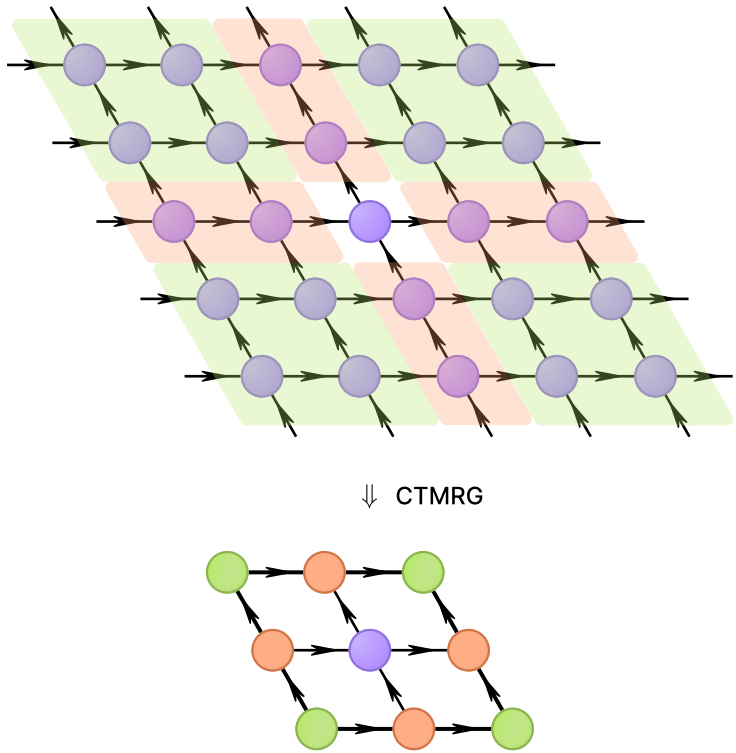


Fig. 2.7 Schematic illustration of the CTMRG scheme for contracting an iPEPS tensor network. The environment is approximated by corner tensors C (green) and edge tensors T (orange), which are iteratively updated via directional coarse-graining steps.

The environmental bond dimension χ controls the accuracy of the CTMRG approximation: larger χ allows for a more faithful representation of the environment at the cost of increased computational resources. Heuristically, we expect $\chi \geqslant D^2$ to fully capture the correlations encoded in the double-layer bulk tensors. However, in practice, smaller χ values can often yield satisfactory observations, possibly attributable to the redundant information encoded in the double-layer tensors. Systematic convergence checks with respect to both D and χ can be performed to ensure the reliability of the results.

CTMRG can struggle with critical systems or those exhibiting long-range correlations, where the environment becomes highly entangled. In these scenarios, the CTMRG iterations may fail to converge properly, with a typical signal of ill-conditioned or divergent matrix inversions when computing the RG projectors. Notable examples include the tight-binding model — an analytically-solvable model that superficially seems well-suited for benchmarking. However, this simple model

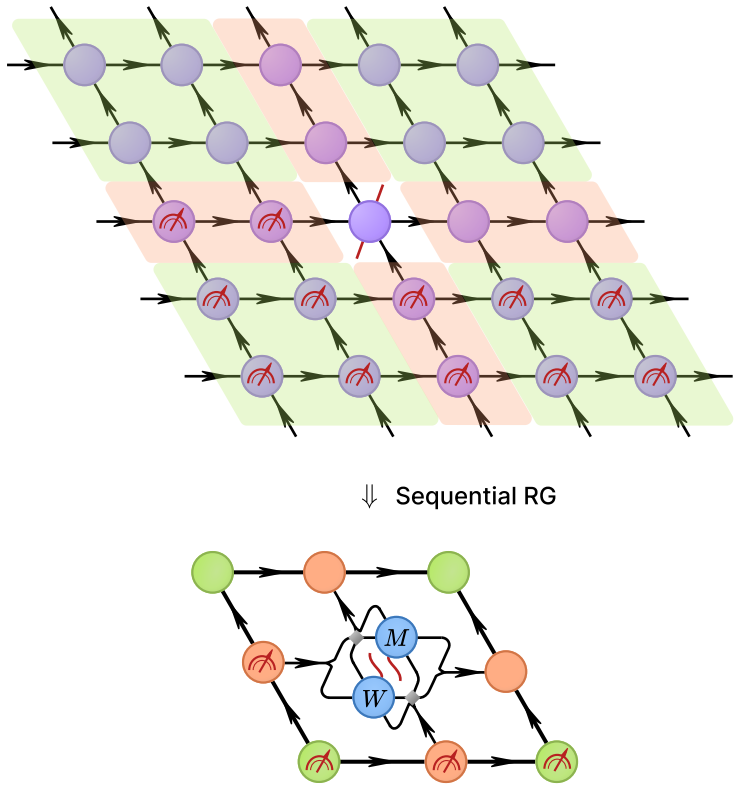


Fig. 2.8 A sequential sampling scheme for iPEPS with CTMRG environment. The measurements and projections (marked by the instrument symbol) are performed row-by-row, and environmental tensors are updated accordingly.

possesses strong criticality with logarithmic violations of the area law [166]. Therefore, one should include a pairing term to open a gap in the system for appropriate benchmarks [46, 48].

Sampling of iPEPS with CTMRG Environment

Once the CTMRG environment has converged, one can sample the iPEPS wavefunction to obtain a set of most probable configurations in the many-body Fock space. This is achieved via a sequential sampling scheme, where one traverses the lattice row-by-row, followed by a projection onto a specific local state according to the measurement outcome. The probability of a measurement outcome is given by the diagonal elements of the local reduced density matrix constructed from the iPEPS tensor and the surrounding CTMRG environment, as illustrated in Fig. 2.8. After each pro-

jection, the environment tensors are updated to reflect the new boundary conditions imposed by the sampled configuration. This process continues until all sites are sampled, yielding a complete configuration of the system.

The coarse-graining renormalization with projected bulk tensors can be performed by either a direct SVD of the enlarged edge tensors or recycling the CTMRG projectors. Pragmatically, both methods work comparably well, since the standard deviation of sampling a small cluster is usually much higher than the tiny inaccuracies introduced by different coarse-graining schemes. The sampling procedure can be repeated multiple times to generate a statistically significant ensemble of configurations, which forms a valuable snapshot dataset for a forthcoming AI analysis or the pretraining [182, 183, 326] of neural quantum states.

2.4 Optimization Algorithms

This section introduces optimization algorithms for leading TNS ansätze, emphasizing the artistry involved in balancing accuracy, stability, and computational cost that guides practical strategies in large-scale simulations. For 1D tensor networks, we revisit DMRG in its one-site (1s) and two-site (2s) variants, highlighting how the global optimization is converted into effective local eigenvalue problems and truncations. We then discuss a hybrid update scheme [40] that interpolates between 1s and 2s sweeps, designed to retain the robustness and the exploration of the enlarged variational space from 2s DMRG while operating at a cost profile comparable to 1s updates. The same idea will be ported to the finite-temperature setting with XTRG, where the square of the thermal density matrix is computed and compressed at each step; there, too, the hybrid optimization scheme yields near 2s accuracy with essentially 1s complexity, which is especially beneficial when temperature decreases rapidly or entanglement grows quickly at intermediate scales.

For 2D iPEPS tensor networks, we focus on ground-state search via imaginary-time evolution using the simple update (SU) [43], which approximates environmental effects through local bond weights and thus scales favorably with bond dimension. To contextualize SU we compare it conceptually with belief propagation schemes [319–322] that propagate messages on the lattice, and with environment-aware alternatives — the fast full update (FFU) [163, 327] and the full update (FU) [46, 328] — that explicitly incorporate an approximate environment from contraction methods like CTMRG. Moreover, variational optimization of iPEPS via automatic differentiation [165, 329–331] has gained traction as a powerful alternative to imaginary-time evolution. We discuss when each approach is preferable for the study of the Hubbard model at its challenging regimes. Throughout, we emphasize unifying principles — each algorithm and ansatz has its own strengths and systematic flaws, and the choice of method should be guided by the specific physical demands, system parameters, and available computational resources.

2.4.1 Ground State DMRG

Historically, DMRG emerged from real-space renormalization ideas and was first formulated in terms of iteratively growing *blocks* and optimizing with respect to reduced density matrices. In contemporary language, however, DMRG is most naturally understood as a variational optimization over the MPS manifold of designated maximum bond dimension. This modern perspective differs in emphasis from the original proposal: rather than coarse-graining degrees of freedom in real space, the algorithm performs a sequence of constrained variational updates on local MPS tensors, guided by the entanglement structure revealed by the Schmidt decomposition.

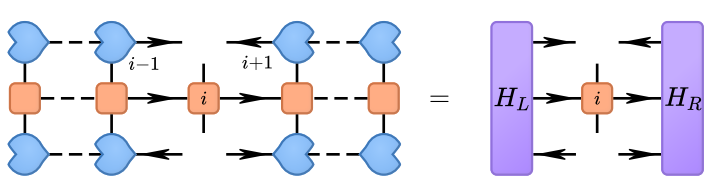
Suppose the target system is described by a Hamiltonian $\mathcal{H}$ and we restrict the ansatz to MPS at maximum bond dimension D. The ground-state search becomes the minimization of the Rayleigh quotient

$$E(A) = \frac{\langle\psi(A)|\mathcal{H}|\psi(A)\rangle}{\langle\psi(A)|\psi(A)\rangle} \tag{2.43}$$

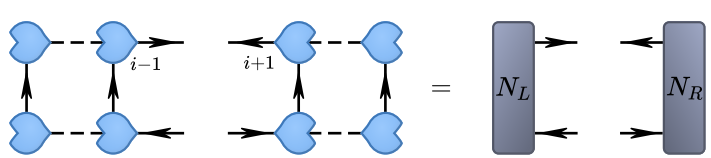
with respect to the set of tensors $A = \{A_i\}$. We do not assume normalization of the state $|\psi(A)\rangle$ represented by the MPS. DMRG tackles this high-dimensional problem by fixing all tensors except one (1s DMRG) or a nearest-neighbor pair (2s DMRG), and minimizing with respect to the remaining variational degrees of freedom while keeping the others as an *environment*. For 1s DMRG, this leads to a local optimization for the tensor at site i:

$$\begin{aligned} \min_{A_i} \frac{\langle\psi(A)|\mathcal{H}|\psi(A)\rangle}{\langle\psi(A)|\psi(A)\rangle} \quad &\Rightarrow \quad \partial_{A_i^*}\langle\psi(A)|\mathcal{H}|\psi(A)\rangle = E\,\partial_{A_i^*}\langle\psi(A)|\psi(A)\rangle \\ &\Rightarrow \quad \mathcal{H}_{1\mathrm{s}}^{(i)} A_i = E\,\mathcal{N}_{1\mathrm{s}}^{(i)} A_i \end{aligned} \tag{2.44}$$

where $\mathcal{H}_{1\mathrm{s}}^{(i)}$ and $\mathcal{N}_{1\mathrm{s}}^{(i)}$ are the effective Hamiltonian and normalization matrices for site i, respectively, obtained by contracting the full Hamiltonian and identity operators with the fixed MPS tensors forming the environment:

$\mathcal{H}_{1\mathrm{s}}^{(i)} =$ 

, (2.45)

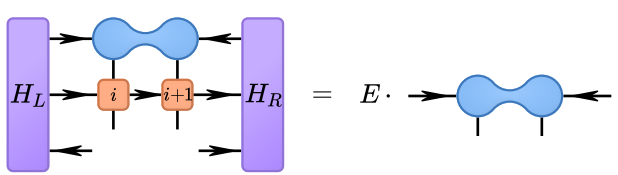
$\mathcal{N}_{1\mathrm{s}}^{(i)} =$ 

. (2.46)

With appropriate gauges, one can make all MPS tensors to the left (right) of site i left- (right-) normalized, so that $\mathcal{N}_{1s}^{(i)}$ reduces to identity, and Eq. (2.44) becomes a standard eigenvalue problem. The optimal tensor A_i is then given by the ground-state eigenvector of $\mathcal{H}_{1s}^{(i)}$. This can be solved efficiently using iterative eigensolvers such as the Lanczos or Davidson algorithms, which only require the action of $\mathcal{H}_{1s}^{(i)}$ on trial vectors without explicitly constructing the full matrix. After updating A_i, one proceeds to the next site, updates the environment, and repeats the process, sweeping back and forth through the chain until convergence.

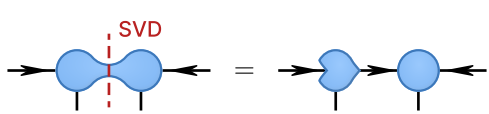
A critical limitation of 1s DMRG is its confinement to the original symmetry sectors defined by the initial MPS. One can observe from Eq. (2.45) that the effective Hamiltonian $\mathcal{H}_{1s}^{(i)}$ preserves the quantum numbers during the optimization of A_i. Consequently, if the initial MPS is restricted to a particular symmetry construction, 1s DMRG cannot explore states outside this submanifold, potentially missing lower-energy configurations that differ in quantum numbers. This limitation can be particularly severe in systems with spontaneous symmetry breaking, where the ground state may acquire a symmetry property disparate from that of the initialization.

In order to overcome this limitation, one can employ 2s DMRG, which allows for a more extensive exploration of the Hilbert space. For 2s DMRG, the optimization is performed over a pair of neighboring tensors A_i and A_{i+1}, leading to the 2s eigenvalue problem:



(2.47)

where the shape of two connected circles represents the combined tensor A_iA_{i+1}. After optimizing this combined tensor, one performs an SVD to split it back into two tensors, truncating the bond dimension as necessary:



(2.48)

The singular-value matrix can be absorbed into either A_i or A_{i+1} to maintain the MPS structure, depending on the direction of the sweep. This 2s update allows an exploration of an enlarged variational space, enabling recombination of quantum numbers across the bond. As a result, 2s DMRG can escape symmetry-constrained subspaces and find lower-energy states that differ in global quantum numbers from the initial MPS.

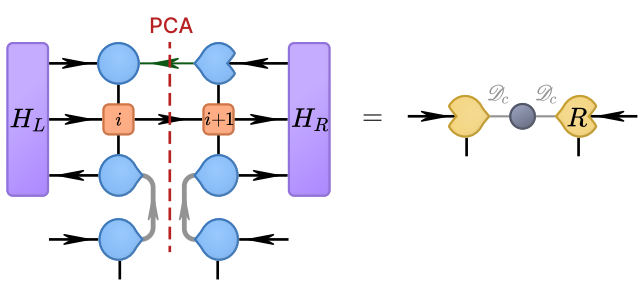
However, the benefit of 2s DMRG comes with a price: the computational cost increases significantly due to the larger effective Hamiltonian matrix, which now

scales with the square of the physical bond dimension. This makes 2s DMRG more expensive and less scalable for systems with large local Hilbert spaces.

Hybrid 1s$^+$ DMRG

A careful examination reveals that the price of 2s DMRG is largely unnecessary. The combined tensor $A_i A_{i+1}$ generically has a rank of (at most) Dd, which is subject to a truncation back to D after the SVD, i.e. a total of (up to) $D(d-1)$ components will simply be discarded. Therefore, one only needs to include a handful of the principal components from the would-be discarded subspace to capture most of the benefits of 2s DMRG. This observation motivates the hybrid 1s$^+$ DMRG, which enters as a part of the Controlled Bond Expansion (CBE) technique [40], to efficiently approximate the 2s update by augmenting the 1s optimization with a small number of additional components from the discarded subspace. By doing so, one can retain much of the robustness and exploration capabilities of 2s DMRG while operating at a computational cost comparable to 1s updates.

The key idea of hybrid 1s$^+$ DMRG is to perform a principal component analysis (PCA) on the two-site tensor at the left-hand side of Eq. (2.47), to extract the most significant components in the discarded subspace that contribute to lowering the energy, namely



(2.49)

where the left-hand side applies the discarded-space projectors [323] to the full two-site tensor. The right-hand side retains only the top $\mathscr{D}_c$ components from the discarded subspace, forming a complement isometry R. Empirically, choosing $\mathscr{D}_c$ to be a small fraction of D (e.g., 10%) or $\sqrt{D}$ suffices to capture most of the benefits of 2s DMRG.

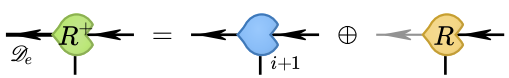
The PCA can be carried out in multiple ways. The guiding principle is to avoid a full construction of the two-site discarded-space tensor, as this would inevitably incur the full 2s DMRG cost. Ref. [40] proposes an orthodox approach, the *shrewd selection*, which involves an SVD on the left half, obtaining the remainders, followed by another SVD on the combination of the remainder and the right half.

An alternative and more efficient method is to perform a reduction of the bond dimension marked green in Eq. (2.49), via, e.g., a truncation of the Schmidt spectrum on that bond. This allows an effective approximation of the two-site discarded-space tensor without an explicit construction. One can then perform an SVD on this reduced

tensor to extract the top $\mathscr{D}_c$ components, forming the complement isometry R. This method, although less accurate than the shrewd selection, works sufficiently well in most practical scenarios, while significantly reducing the computational overhead.

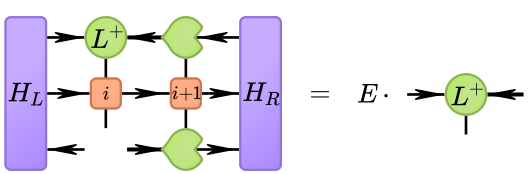
A plausible explanation for the effectiveness of this naive PCA approach is that, the variational optimization still lies in the iterative local updates. The PCA of the two-site discarded-space tensor merely needs to provide a reasonable expansion of the variational space, rather than an exact LRA of the discarded subspace. As long as the selected components capture the dominant directions that escape the symmetry constraints, the subsequent 1s optimization can effectively utilize them to lower the energy.

With the complement isometry R obtained, one can proceed to the subsequent steps for 1s$^+$ DMRG optimization. We next take the left-to-right sweep as an example; the right-to-left sweep follows analogously. The isometry at site $i+1$ is first enlarged by absorbing the complement isometry:



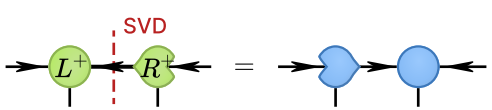
(2.50)

The resulting R^+ tensor now acquires an enlarged bond dimension of $\mathscr{D}_e = D + \mathscr{D}_c$ on the left side. The local eigenvalue problem for the MPS tensor at site i now becomes



(2.51)

where the obtained MPS tensor L^+ also has an enlarged bond dimension $\mathscr{D}_e$ on the right side. After solving this eigenvalue problem, one contracts L^+ and R^+, performs an SVD to decompose the updated tensor back into the standard MPS form, and truncates the bond dimension back to D:



(2.52)

The 2s DMRG has a computational cost scaling as $\mathcal{O}(D^3d^2w)$ (for MPO bond dimension w), while the hybrid 1s$^+$ DMRG scales as $\mathcal{O}(D^2\mathscr{D}_e dw)$. Since $\mathscr{D}_e = D+\mathscr{D}_c$ with $\mathscr{D}_c \ll D$, the hybrid scheme effectively retains the favorable cost profile of 1s DMRG while capturing much of the exploration capacity of 2s DMRG. Note that there is still a subleading cost from the SVD of the two-site tensor with cost $\mathcal{O}(D^3d^3)$, and the 1s$^+$ DMRG also introduces extra overhead in constructing the complement isometry R. Therefore, the actual speed-up compared to 2s DMRG depends on the specific parameters and implementations. Nevertheless, the hybrid 1s$^+$

DMRG provides a compelling compromise between accuracy and efficiency, making it a valuable tool for large-scale simulations where full 2s DMRG may be prohibitively expensive.

2.4.2 Thermal XTRG

Finite-temperature TN simulations have long relied on two paradigms: purification-based imaginary-time evolution and sampling of typical states. In purification, one embeds the thermal state into a pure state on an enlarged Hilbert space and evolves in small imaginary-time steps, i.e., $\rho(\beta) = \mathrm{tr}_{\mathrm{anc}}|\Psi(\beta)\rangle\langle\Psi(\beta)|$ with $|\Psi(\beta+\delta\beta)\rangle \approx e^{-\frac{\delta\beta}{2}H}|\Psi(\beta)\rangle$. While robust, this incurs an ancilla overhead and requires $\mathcal{O}(\beta/\delta\beta)$ Trotter steps, with accumulated Trotter and truncation errors and bond-dimension growth that becomes challenging at low T. The METTS approach, by contrast, avoids ancillas and estimates thermal expectation values by a Markov chain of product-state collapses and short imaginary-time evolutions. METTS can be very effective at intermediate temperatures, yet its stochastic variance and autocorrelations demand many samples for high accuracy and, again, the cost grows essentially linearly with β. Against this backdrop, XTRG reframes finite-T calculations as deterministic operator-space compression that reduces both the number of evolution steps and the sources of systematic error.

The central idea of XTRG is to represent the thermal density matrix $\rho(\beta) = e^{-\beta H}$ as an MPO and to evolve it by repeatedly squaring the MPO. Starting from a high-temperature thermal state $\rho(\epsilon)$ at small inverse temperature ϵ, one expands the density matrix into a power series

$$\rho(\epsilon) = e^{-\epsilon H} \approx \sum_{k=0}^{N_c} \frac{(-\epsilon)^k}{k!} H^k \tag{2.53}$$

truncated at a large enough order N_c. Usually, for sufficiently small ϵ (e.g. $\epsilon = 2^{-11}$), a low order N_c (e.g. $N_c \leqslant 8$) suffices. The construction of the initial density matrix involves pairwise summation and multiplication of MPOs, relevant techniques of which will be discussed subsequently. Once $\rho(\epsilon)$ is constructed, one takes the square repeatedly to reach the desired inverse temperature:

$$\rho(2\beta) = \rho(\beta) \cdot \rho(\beta). \tag{2.54}$$

This procedure gives $\rho(2^n\epsilon)$ after n iterations, allowing an exponential reduction in temperatures. In reality, the squaring procedure may significantly scale the norm of the density matrix (or equivalently, the partition function $Z(\beta) = \mathrm{tr}\,\rho(\beta)$), leading to exotically large tensor elements. This issue is particularly pronounced for XTRG iterations $n \geqslant 10$. To mitigate this, one typically normalizes the density matrix at each step, and keeps track of the partition function at each temperature. The free

energy and other thermodynamic quantities can then be computed with an aid of the accumulated normalization factors.

The summation and product of MPOs at each step inevitably increase the bond dimension, necessitating a truncation back to a manageable size. Multiple strategies exist for this truncation, including SVD-based LRA compression and variational optimization. The SVD method performs a local decomposition and truncation of the MPO tensor, while the variational approach minimizes the distance between the exact and truncated MPOs.

MPO-MPO Summation and Product

The naive summation of two MPOs $\Omega^{(1)}$ and $\Omega^{(2)}$ with bond dimensions ω_1 and ω_2 proceeds via a site-wise direct sum of their local tensors

$$\Omega_i = \Omega_i^{(1)} \oplus \Omega_i^{(2)} = \begin{bmatrix} \Omega_i^{(1)} & 0 \\ 0 & \Omega_i^{(2)} \end{bmatrix}. \tag{2.55}$$

This construction yields an MPO Ω with total bond dimension $\omega = \omega_1 + \omega_2$. While elementary and robust, such a procedure induces rapid bond-dimension inflation when aggregating many terms, as occurs in the initial assembly of $\rho(\epsilon)$ in Eq. (2.53). A standard countermeasure is to perform an SVD-based truncation after each summation step.

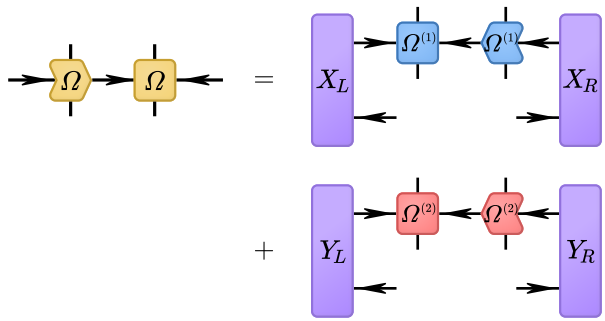
However, the cost of repeated SVD truncations can scale unfavorably with the growing bond dimension, particularly when a large number of summands are involved. A more scalable alternative is a variational optimization scheme [54], which seeks the truncated MPO by minimizing the Frobenius norm of the deviation between the exact sum and its compressed representation

$$\begin{aligned} &\min_{\Omega} \|\Omega - (\Omega^{(1)} + \Omega^{(2)})\|_F^2 \\ =&\min_{\Omega} \left[\mathrm{tr}(\Omega^\dagger \Omega) - \mathrm{tr}((\Omega^{(1)} + \Omega^{(2)})^\dagger \Omega) - \mathrm{tr}(\Omega^\dagger (\Omega^{(1)} + \Omega^{(2)})) + \mathrm{const}\right]. \end{aligned} \tag{2.56}$$

The optimization can be performed using local updates via taking derivatives with respect to each pair of neighboring MPO tensors, leading to a set of linear equations that can be solved iteratively

$$\frac{\partial \Omega^\dagger \Omega}{\partial (\Omega_i \Omega_{i+1})^*} = \frac{\partial \Omega^\dagger (\Omega^{(1)} + \Omega^{(2)})}{\partial (\Omega_i \Omega_{i+1})^*}. \tag{2.57}$$

With appropriate gauging of the MPO, the left-hand side reduces directly to the updated two-site tensor. Diagrammatically, this update can be represented as (take left-to-right sweep as an example; right-to-left sweep is analogous)



(2.58)

The product of two MPOs $\Omega^{(1)}$ and $\Omega^{(2)}$ can be constructed by contracting their physical indices site-by-site, resulting in a new MPO $\Omega = \Omega^{(1)} \circ \Omega^{(2)}$ with bond dimension $\omega = \omega_1 \cdot \omega_2$. Analogously, the bond dimension can grow exponentially with repeated products, as in the squaring step of XTRG in Eq. (2.54). To control this growth, one can again employ SVD-based truncation or variational optimization similar to the summation case.

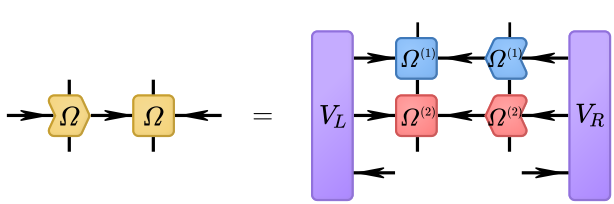
The variational optimization for the product MPO [54] adheres to the same guiding principle as Eq. (2.56), namely, minimizing the Frobenius norm of the discrepancy between the exact product and its truncated MPO approximation

$$\begin{aligned} &\min_{\Omega} \|\Omega - (\Omega^{(1)} \circ \Omega^{(2)})\|_F^2 \\ =&\min_{\Omega} \left[\mathrm{tr}(\Omega^\dagger \Omega) - \mathrm{tr}((\Omega^{(1)} \circ \Omega^{(2)})^\dagger \Omega) - \mathrm{tr}(\Omega^\dagger (\Omega^{(1)} \circ \Omega^{(2)})) + \mathrm{const}\right]. \end{aligned} \tag{2.59}$$

By invoking the same projection framework, one arrives at local update rules that mirror those for MPO summation, yielding a set of linear equations for each pair of neighboring MPO tensors

$$\frac{\partial \Omega^\dagger \Omega}{\partial (\Omega_i \Omega_{i+1})^*} = \frac{\partial \Omega^\dagger (\Omega^{(1)} \circ \Omega^{(2)})}{\partial (\Omega_i \Omega_{i+1})^*}. \tag{2.60}$$

Diagrammatically, the update for the product MPO can be represented as (again take left-to-right sweep as an example)



(2.61)

In the foregoing, we have outlined the 2s formulation of variational optimization for both MPO summation and product. In close analogy with DMRG, one may also devise a 1s$^+$ variant that reinjects a limited set of components from the discarded

subspace, thereby enriching the accessible variational manifold while preserving a favorable computational profile. This hybrid strategy is the central subject of Section 3.2, which pushes the lowest temperature accessible to a brand new level.

2.4.3 Ground State iPEPS

In the landscape of iPEPS optimization, two broad paradigms are commonly employed: imaginary-time evolution [43, 46, 47] and direct variational minimization of the energy functional [53, 164, 329–332] — often via gradient-based schemes [165] constrained to the iPEPS manifold. While fully variational routes provide self-adaptive control over optimization procedures, imaginary-time projection offers a robust and computationally efficient prior that is particularly effective under a sufficiently accurate environment approximation. The central idea is to regard ground-state preparation as spectral filtering: acting with $e^{-\beta H}$ on any trial state with non-zero ground-state overlap progressively damps excited components at a rate set by their energy gaps. This *damping* procedure usually provides a stable and efficient path to the ground state preceding a more refined variational optimization.

Formally, if an arbitrary initial state can be expanded in the Hamiltonian eigenbasis as $|\psi_0\rangle = \sum_n c_n |E_n\rangle$ with $\mathcal{H}|E_n\rangle = E_n|E_n\rangle$, then imaginary-time evolution yields

$$e^{-\beta H}|\psi_0\rangle = e^{-\beta E_0}\left[c_0|E_0\rangle + \sum_{n>0} c_n\, e^{-\beta(E_n-E_0)}|E_n\rangle\right], \tag{2.62}$$

so the normalized state $|\phi(\beta)\rangle = e^{-\beta H}|\psi_0\rangle / \|e^{-\beta H}|\psi_0\rangle\|$ converges to the ground state as $\beta \to \infty$ whenever $c_0 \neq 0$. At the asymptotic level, the energy monotonically decreases with increasing β, and the convergence rate is characterized by the first excitation gap $\Delta = E_1 - E_0$. In the presence of symmetry, the projection converges to the preset symmetric subspace; enforcing symmetries at the tensor level both targets the desired sector and improves numerical stability.

The full Hamiltonian is a high-order tensor object, and the imaginary-time propagator $e^{-\beta H}$ for large β cannot be applied to iPEPS tensors in a single step without incurring prohibitive cost and loss of locality. A standard remedy is to invoke a Trotter–Suzuki factorization [333] that replaces the global evolution by a sequence of strictly local gates acting on neighboring sites. For a Hamiltonian decomposed into two-site terms $\mathcal{H} = \sum_{\langle i,j\rangle} h_{ij}$, a first-order Trotter expansion yields

$$e^{-\tau\mathcal{H}} \approx \prod_{\langle i,j\rangle} e^{-\tau h_{ij}} + \mathcal{O}(\tau^2), \tag{2.63}$$

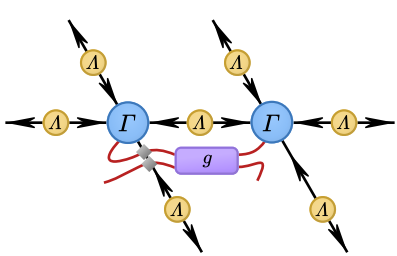
where τ needs to be sufficiently small to constrain the Trotter errors. In order to reach a large β, one then applies the local gates $e^{-\tau h_{ij}}$ sequentially in many small imaginary-time increments, updating the iPEPS tensors after each gate application.

This iterative cooling is continued until convergence, typically diagnosed by negligible variation in the Schmidt spectra across all updated bonds.

The local update necessitates an efficient mechanism for absorbing the two-site gate into the iPEPS tensors and subsequently truncating the enlarged bond dimension to its working value. In principle, an optimal truncation is variationally defined by the full environment surrounding the updated tensors; this environment can be approximated via CTMRG as discussed in Section 2.3.4. This consideration naturally leads to the full update (FU) scheme [46, 328], wherein, after each imaginary-time step, an environment is (re)constructed and used to determine the truncation that best preserves the post-gate state. Although FU markedly improves fidelity by aligning the truncation with the global network geometry, recomputing the environment at every step renders the procedure computationally demanding. The fast full update (FFU) [163, 327] ameliorates this cost by incrementally updating the environment tensors after each gate application — rather than rebuilding them from scratch — thereby retaining much of FU's accuracy at a reduced overhead, albeit with an overall scaling that remains substantial.

The simple update (SU) [43] provides a pragmatic surrogate for full-environment methods by replacing the global contraction with local bond weights that encode the corresponding Schmidt spectra. This approximation drastically lowers the computational burden while retaining the ability to capture the most salient short- to intermediate-range correlations along each bond. In practice, SU preserves a generic Lambda–Gamma representation [334,335] that records the local entanglement structure on every virtual link, thereby enabling efficient and numerically stable tensor updates without explicitly constructing the full environment.

As an illustrative example, consider applying a two-site gate g on neighboring sites i and j connected by a bond surrounded by the Schmidt weights Λ as the environment. The update proceeds as

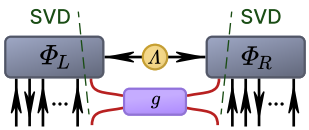


(2.64)

Other neighboring or next neighboring updates can be performed analogously. Detailed procedures for the SVD truncation and absorption of the Schmidt weights can be found in Ref. [325] and in Section 3.1.

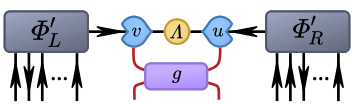
Within iPEPS (and, more generally, higher-dimensional tensor networks), gate updates frequently act on tensors of very high order; for example, Eq. (2.64) involves an order-8 tensor prior to truncation. A monolithic SVD at such arity is typically computationally prohibitive. A pragmatic remedy is to factor the object through a controlled sequence of lower-order decompositions — most commonly SVDs or

QR factorizations — thereby expressing the original high-order tensor as a compact network of low-order constituents (preferably order-3)



(2.65)

In particular, this construction decomposes the high-order tensors Φ_L and Φ_R into order-3 constituents v and u. The gate application and the subsequent truncation can then be executed locally on v and u,



(2.66)

which markedly reduces computational overhead while maintaining the intended variational structure. After these updates, the smaller tensors are recombined to restore the original connectivity. This sequential-decomposition strategy is particularly effective at taming the complexity of high-order tensor manipulations during iPEPS optimization. As will be demonstrated in Section 3.1, the technique is crucial for enabling efficient imaginary-time evolution in simulations of the next-nearest-neighbor Hubbard model.

2.5 Transformer and Attention Mechanism

Transformers have rapidly become popular for scientific machine learning. For physicists, this flexibility is attractive: fields on grids, sets of particles, graphs of interacting subsystems, or symbolic expressions can all be serialized into tokens and processed by the same computational primitive — attention — whose capacity to integrate non-local information mirrors many long-range correlations in physical systems. Also, scientific use demands more than predictive power: we must articulate design choices that respect the native regularities of the real-world scientific data, select transformer variants that align with task structure, and scrutinize what — if anything — attention weights reveal about learned mechanisms. In this section, we assume basic familiarity with transformer architectures. A beginner's guide can be found in the supplemental material in Section 3.3.

The first part of this section outlines a principled workflow for designing AI architectures for physics. We frame model construction as three coupled mappings: (i) encoding physical objects into token sequences with (relative) positional information, (ii) processing with modular blocks that can be augmented to enforce awareness of intrinsic patterns or correlations, and (iii) decoding back to physically meaningful

quantities with calibrated uncertainties. Throughout, the emphasis is on architectural levers — tokenization schemes, positional structure, conditioning, and objective design — that balance data efficiency, stability, expressivity, and fidelity.

The second part discusses transformer variants by the tasks common in physics. Encoder-only models, which aggregate information into contextualized representations, are naturally suited to discrimination and regression (phase classification, property estimation, anomaly detection). Decoder-only models, trained with autoregressive objectives, target generative synthesis and sequential prediction (trajectory rollout, quantum information processing), where one conditions on past tokens to produce plausible futures. Encoder-decoder models separate information acquisition from controlled generation, enabling transduction problems (translating information from one domain to another) and context-sensitive simulation (conditioning on given physical parameters while generating sampled sequences of quantum states).

The final part addresses the interpretability debate surrounding attention. Attention matrices are accessible and sometimes align with salient physical couplings, which makes them tempting explanatory artifacts. Yet they are not uniquely causal: multiple internal pathways can yield the same output, and learned reparameterizations may redistribute attribution away from weights alone. We recommend a cautious approach — using attention as one corroborating signal among many. Together, these perspectives set the stage for a disciplined, physics-aware interpretation of the attention mechanism.

2.5.1 Design of AI Architectures

Designing learning systems for physics requires a coherent incorporation of the physical data structure and the inductive biases of modern architectures. Experimental readouts, lattice configurations, and geometric constraints need to be properly mapped into a machine-interpretable representation, transformed across depth, and decoded into observables or information of interest. The central motivation of this section is to develop a unified architectural perspective that guides the design — how discrete and continuous inputs are tokenized and embedded; how to select intermediate processing blocks (fully connected, convolutional, recurrent, attention-based, etc.); and how output heads are aligned with discriminative prediction or generative tasks. Framing design in these terms clarifies where to inject physical priors, how to balance computational cost with fidelity to correlation structure, and how to promote interpretability and reproducibility across disparate physics tasks.

Architectural Overview: From Inputs to Outputs

Most contemporary models can be organized as a modular pipeline with three stages. First, an input interface **encodes** raw data — discrete symbols (e.g., local states on lattice sites, particle species, measurement outcomes) or continuous quantities (coor-

dinates, fields, spectra) — into numerical vectors suitable for subsequent processing. When the input is naturally a collection of items (sites of a lattice, time steps, detector channels), it is convenient to represent each item as a **token**; the collection of tokens forms a sequence or set that constitutes the model input. In this stage, one may also attach *positional information* (e.g., lattice coordinates or time indices) to preserve geometry or ordering.

Second, the model applies a stack of processing blocks. A **block** (also called **layer** or **module**) denotes a standardized unit that combines a principal transformation (e.g., a fully connected, convolutional, recurrent, or attention operation) with stabilization mechanisms such as residual connections and normalization. The stack acts on a matrix of size $L_{\text{seq}} \times d_{\text{model}}$, where L_{seq} is the sequence length (number of tokens) and d_{model} is the model size (number of features per token).

Third, one or more **output heads** transform the final internal representation into the desired target: class labels (e.g., phase identification), continuous quantities (energies, order parameters), or full probability distributions over configurations (for generative modeling or sampling). This modular decomposition separates concerns — how physics enters, how it is processed, and how predictions are read out — facilitating a streamlined design paradigm that turns otherwise complicated black-box processing into a block-building exercise.

Internal Representations: Feature Vectors

Within the model, each token is represented by a feature vector $\boldsymbol{h} \in \mathbb{R}^{d_{\text{model}}}$. Although features in neural representations are typically *distributed* (i.e., a specific physical quantity may be encoded across many dimensions), it is helpful to view $\mathbb{R}^{d_{\text{model}}}$ as a feature space spanning directions that correlate with physically relevant attributes. For example, suppose a token corresponds to a spin-up quantum state located at a specific lattice site. The features may align with quantities such as particle number, total spin S, or the spin-z component S_z, as well as contextual information (sublattice identity, boundary proximity, or membership in a symmetry sector). The width d_{model} controls the capacity to encode such structure: larger d_{model} affords finer representational granularity at the cost of increased computational complexity and memory. When system-level information (temperature, global magnetization, conserved charges) is important, it can be injected (e.g., added or appended) uniformly to all tokens or implemented as *cross attention* that interacts with local site tokens in the processing stack.

Input Interface: Encoding, Embedding, and Continuous Projections

The input interface maps raw physical data into token sequences with associated feature vectors. The design depends on the data modality: discrete symbols (e.g., spin states, particle occupations) are typically embedded via learned lookup tables, while

continuous quantities (e.g., coordinates, fields) may be projected through learned linear layers. Two standard mechanisms translate discrete inputs into vectors:

- **Encoding** refers to a fixed, non-trainable mapping. Examples include one-hot vectors for $\{\uparrow, \downarrow\}$, binary occupation encodings $\{0, 1\}^{L_{\text{seq}}}$, or hand-crafted coordinates aligned with symmetry irrep labels. Encodings faithfully preserve semantics and can enforce known constraints, but they may be high-dimensional and do not adapt to data statistics.
- **Embedding** denotes a learned lookup: each discrete symbol is assigned an associated vector that is optimized during training. Embeddings are compact and data-adaptive, often yielding better statistical efficiency, though they require sufficient data and regularization to avoid overfitting spurious correlations.

For continuous inputs (positions, momenta, fields, spectra), the canonical front end is a **linear projection** — a fully connected layer mapping $\mathbb{R}^p \to \mathbb{R}^{d_{\text{model}}}$ that transforms p physical properties into d_{model} features. This may be preceded by standardization and can be augmented with engineered bases (e.g., sinusoidal/Fourier features) to facilitate the representation of fine spatial or temporal variation. When geometry or ordering matters, **positional encodings** — absolute or relative — are added so subsequent blocks can exploit lattice structure or temporal causality.

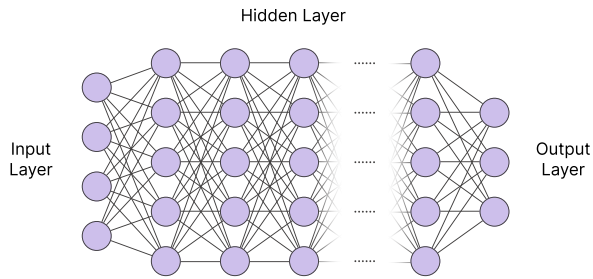
Output Heads: Discrimination, Regression, and Generation

The choice of output head is dictated by the scientific objective. **Discriminative** heads produce conditional predictions given the input representation. Typical instances include (i) **classification** (e.g., phase labels, defect types) [276, 336] via a softmax function trained with cross-entropy, and (ii) **regression** (energies, magnetizations, susceptibilities) [337–339] via linear or shallow non-linear heads trained with mean-squared or absolute error.

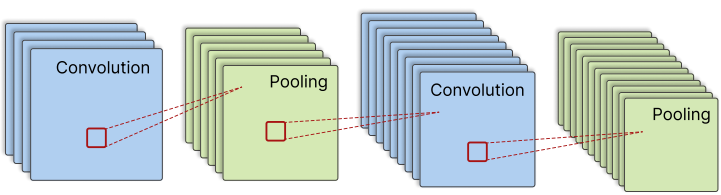
Generative heads model distributions over configurations or signals. **Autoregressive** framework emits one token at a time conditioned on predecessors, enabling sampling of state configurations or sequences [182, 183] consistent with learned physics. **Diffusion/score-based** heads learn to denoise from injected noise, which may be effective e.g. for astronomical error mitigation [340, 341]. In physical workflows, generative heads are valuable for proposing equilibrium configurations, accelerating sampling, or reconstructing signals under noise.

Processing Blocks: MLPs, Convolutions, Recurrence, and Attention

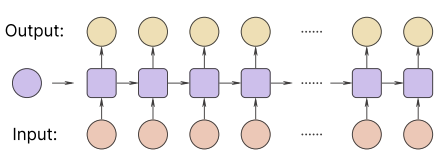
The processing stack captures correlations and inductive biases. **Fully connected multi-layer perceptron (MLP)** blocks are universal approximators and work well when inputs are already compressed into informative descriptors (a small set of scalars per system). However, they function as a complete black box which lacks explainability and thus diagnostic tools, and scale poorly with system size due to the



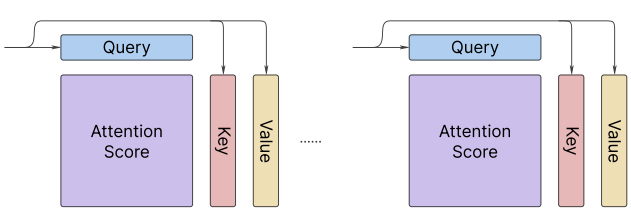
(a) Multi-layer Perceptron (MLP)



(b) Convolutional Neural Network (CNN)



(c) Recurrent Neural Network (RNN)



(d) Attention Mechanism

Fig. 2.9 Architectural processors commonly used in physics-aware AI models. (a) Fully connected MLP block interconnects all input features, making them flexible but computationally expensive. (b) CNN involves repeated convolutional filtering and pooling, capturing local patterns and translational equivariance. (c) RNN maintains internal states across sequence steps, preserving order-dependent context. (d) The attention mechanism computes attention scores between all token pairs, enabling global information integration.

quadratic growth of parameters with input dimension. Therefore, contemporary architectures usually employ MLPs only as *feed-forward* sublayers within more structured blocks (e.g., transformers).

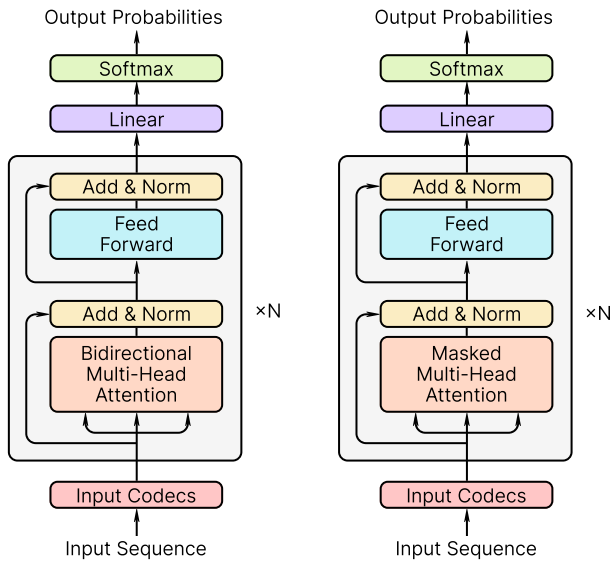
Convolutional blocks encode locality and translational equivariance through weight sharing and receptive fields, making them natural for lattice and image-like data (spin textures, density maps, microscopy). Multiscale behavior can be captured by strides, pooling, or dilated convolutions. However, strictly local receptive fields can impede modeling of very long-range dependencies unless many layers or dilation mechanisms are used.

Recurrent blocks e.g. Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTMs), Gated Recurrent Units (GRUs), process sequences with explicit state, preserving order-dependent context. RNNs are most effective when the data possess an inherent causal order (e.g., time series or sequential experimental protocols). Their main limitations are reduced parallelism and difficulties with very long dependencies, which gating alleviates but does not eliminate.

Attention (as in transformer architectures) enables *global awareness*: every token can condition on every other token within a layer, allowing the model to learn long-range correlations and non-local couplings that frequently arise in many-body systems and critical phenomena. Attention is highly parallelizable over tokens and adapts naturally to sequences, sets, or graphs once appropriate positional or structural encodings are provided. The principal cost is quadratic complexity (note the difference from quadratic growth of parameters) in sequence length for dense attention, which motivates sparse or multi-expert variants for large or extensive systems. In exchange, attention often provides superior fidelity when correlations are global or long-range.

2.5.2 Variants of Transformers

Transformers come in three canonical wiring patterns — *encoder-only*, *decoder-only*, and *encoder-decoder*. All three reuse the same two building blocks (multi-head attention and position-wise feed-forward), but they differ in how attention is masked and how the blocks are stacked. These choices shape the kinds of tasks each variant serves best. Encoder-only models [68] process an observed sequence with *bidirectional* self-attention (each token attends to both its left and right context) and are most effective when the goal is discrimination: classification, regression, tagging, or retrieval. Decoder-only models [70] use *causal* (left-to-right) self-attention and directly implement autoregressive likelihoods, making them the natural fit for *generative* tasks where outputs are produced token by token. Encoder-decoder models [61] couple a bidirectional encoder to a causal decoder via cross-attention and are thus tailored to *transduction* or *context-sensitive generation*, where the output sequence must be generated conditioned on a separate input (e.g., translation or summarization).



(a) Encoder-only Transformer (b) Decoder-only Transformer

Fig. 2.10 Two variants of transformer architectures. Encoder-only transformers stack multiple layers of bidirectional self-attention, while decoder-only transformers stack layers of causal (left-to-right) self-attention.

Encoder-Only Transformers (for Discrimination)

An encoder-only transformer contains a stack of layers that maps an input token sequence $\boldsymbol{\Sigma}_i$ (see also Section 3.3) to contextual representations $\boldsymbol{g}_i$ (the output of the gray shaded box in Figure 2.10). Each layer applies multi-head self-attention without a causal mask, so token i can attend to $\{1, 2, \ldots, L_{\text{seq}}\} \setminus \{i\}$. This *bidirectional* context is crucial: it allows the representation of a token to incorporate information from both its left and right companions, which is precisely what many discrimination tasks require (e.g., deciding a sentence's sentiment or identifying phases of matter).

After processing through a stack of N attention and feed-forward layers, the final representations $\boldsymbol{g}_i$ are pooled (e.g., via averaging or a special classification token `CLS`) and passed to an output head for classification or regression. For instance, the logit of a K-category classification can be constructed as

$$y_c = \text{pool}(\{\boldsymbol{g}_i\})\boldsymbol{W}_c + b_c, \quad c = 1, \ldots, K, \tag{2.67}$$

where $\boldsymbol{W}_c \in \mathbb{R}^{d_{\text{model}}}$ and b_c are learned parameters. Probability p_c of category c is then given by a softmax over logits y_c. The model can be trained end-to-end with cross-entropy loss on labeled data. Regression tasks follow a similar pattern, replacing with e.g. mean-squared error loss.

In addition to the standard training procedure, encoder-only transformers can benefit from an optional pretraining phase based on **masked language model (MLM)** [68]. During this phase, a portion of the input tokens is randomly masked, and the model is trained to predict the masked tokens based on their context. Let $M \subset \{1, 2, \ldots, L_{\text{seq}}\}$ index positions to mask, and let $\vec{\sigma} \setminus \sigma_{\{M\}}$ be the corrupted sequence (e.g., masked or noised tokens at indices in M). The model then produces contextual representations $\boldsymbol{g}_i$ for $i \in M$, which are passed to a generative head

$$z_{i,\sigma} = \boldsymbol{g}_i \cdot \boldsymbol{W}_\sigma + b_\sigma, \quad p_{i,\sigma} = \text{softmax}_i(z_{i,\sigma}), \quad i \in M, \tag{2.68}$$

where $\boldsymbol{W}_\sigma \in \mathbb{R}^{d_{\text{model}}}$ and b_σ are learned parameters. A common efficiency trick is *weight tying* [342], i.e. $\boldsymbol{W}_\sigma = \boldsymbol{e}(\sigma)$, where $\boldsymbol{e}(\sigma)$ is the embedding vector for token σ. The model is trained to maximize the likelihood of the original tokens at masked positions

$$\mathbb{E}_{\vec{\sigma}, M} \left[-\sum_{i \in M} \log p(\sigma_i \mid \vec{\sigma} \setminus \sigma_{\{M\}}) \right], \tag{2.69}$$

where $\mathbb{E}_{\vec{\sigma}, M}$ denotes expectation over a mini-batch of sequences $\vec{\sigma}$ and random mask sets M. After pretraining, the model can be fine-tuned on downstream discrimination tasks with supervised loss. This two-stage training encourages the model to learn richer representations by forcing it to leverage bidirectional context, which often improves data efficiency and robustness, especially when labeled data are scarce.

Decoder-Only Transformers (for Autoregressive Generation)

A decoder-only transformer differs from the encoder-only variant by employing *causal* self-attention: each token i can only attend to its left context $\{1, 2, \ldots, i-1\}$. This induces the left-to-right factorization, which is essential for autoregressive generation

$$p(\vec{\sigma}) = \prod_{i=1}^{L_{\text{seq}}} p(\sigma_i \mid \sigma_{<i}), \tag{2.70}$$

where $\sigma_{<i} = \{\sigma_1, \sigma_2, \ldots, \sigma_{i-1}\}$. The generative head is similar to Eq. (2.68), but now applied to all positions. Training uses next-token prediction with teacher forcing: the model is fed the gold prefix $\sigma_{<i}$ and optimized by the cross-entropy loss

$$\mathbb{E}_{\vec{\sigma}} \left[-\sum_{i=1}^{L_{\text{seq}}} \log p(\sigma_i \mid \sigma_{<i}) \right]. \tag{2.71}$$

At inference time, generation proceeds autoregressively: starting from a prompt (possibly empty), the model samples one token at a time, appending each sampled token to the input for the next step. Decoding can be deterministic (greedy, beam

search with length normalization) [343] or stochastic (temperature scaling [344], top-k/nucleus sampling [345, 346]), trading off fidelity and diversity.

The causal mask enforces an information flow analogous to time-ordered conditioning: each state aggregates only past context, making the model particularly well-suited to tasks where outputs must be emitted in sequence. Practical implementations cache key-value pairs from previous layers so that each new step only attends to stored states, enabling efficient long-form decoding.

Conditional tasks can be handled by concatenating the conditioning information into the prefix [70, 347, 348]. If $\vec{\gamma}$ denotes a context string (instructions, a problem statement, or physical parameters), the model generates sequence $\vec{\sigma}$ from $p(\vec{\sigma}|\vec{\gamma})$ using the same autoregressive mechanism. This *prefix-LM* view works well when the conditioning signal is short and of fixed-length; for richer or complex multi-modal inputs, the encoder-decoder pattern below provides a more explicit and scalable conditioning pathway.

Encoder-Decoder Transformers (for Context-Sensitive Generation)

An encoder-decoder transformer combines a bidirectional encoder and a causal decoder connected via cross-attention, and separates the *understanding* of a source sequence from *generation* of a target sequence. The encoder processes an input sequence $\vec{\gamma}$ into contextual representations $\boldsymbol{u}_j$, while the decoder generates the output sequence $\vec{\sigma}$ autoregressively, attending to both its own history and the encoder contexts, yielding the conditional factorization

$$p(\vec{\sigma} \mid \vec{\gamma}) = \prod_{i=1}^{L_{\text{seq}}} p(\sigma_i \mid \sigma_{<i}, \vec{\gamma}). \tag{2.72}$$

Specifically, at each decoder layer, the self-attention block is followed by a cross-attention block where each decoder token attends to all encoder outputs $\boldsymbol{u}_j$, i.e. take queries from the decoder and keys/values from the encoder. This structure allows the decoder to flexibly incorporate information from the source sequence at every generation step.

Decoupling source encoding from target generation confers several advantages for *transduction* (machine translation, abstractive summarization, speech/text transcription, etc.) and, more broadly, *context-sensitive generation*. The encoder can build globally consistent, bidirectional features over the entire input — often much longer or structurally different from the output — while the decoder focuses on fluent emission conditioned on those features. This is especially helpful when there are strong length mismatches, reordering, or many-to-one alignments between source and target. In multi-modal settings, the encoder serves as a dedicated front-end, providing great scalability while retaining the same decoding machinery. Practically, the encoder/decoder separation also enables flexibility in partial fine-tuning: one can freeze the encoder and adapt only the decoder to new generation tasks, or vice versa. At

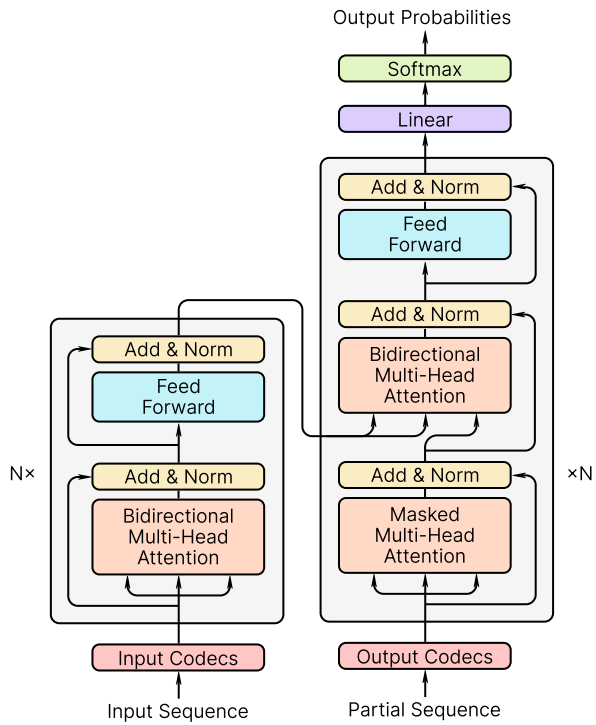


Fig. 2.11 Encoder-decoder transformer architecture. The encoder processes the input sequence with bidirectional self-attention, while the decoder generates the output sequence autoregressively with causal self-attention and cross-attention to the encoder outputs.

inference time, this separation allows pre-encoding of static contexts, speeding up generation when the same source is applied to multiple outputs.

2.5.3 Interpretation of Attention Scheme

The attention mechanism is often lauded for its *interpretability* or *explainability*. In transformer architectures, an attention head at the ℓ-th layer with *model temperature* $\mathfrak{T}$ computes a row-stochastic matrix

$$A_{ij}^{(\ell)} = \text{softmax}_j \left(\boldsymbol{\mathcal{Q}}_i^{(\ell)} \cdot \boldsymbol{\mathcal{K}}_j^{(\ell)} / \mathfrak{T} \right), \tag{2.73}$$

where $\boldsymbol{\mathcal{Q}}_i^{(\ell)}$ and $\boldsymbol{\mathcal{K}}_j^{(\ell)}$ are the query and key vectors for tokens i and j at layer ℓ. Read as an alignment, $A^{(\ell)}$ often highlights linguistically [349] or visually [350] meaning-

ful relations (e.g., dependency links in text, object-part structure in images), and in scientific settings it frequently correlates with physically significant neighborhoods or interactions [84, 351]. In general, attention externalizes where the model concentrates its limited computational bandwidth. It is therefore a *useful* explanatory signal directly tied to the model's native computation, inexpensive to obtain at inference time, and naturally comparable across layers and heads.

The main caveat is structural rather than philosophical. Because attention operates inside a deep neural network, the numerical weight on a particular edge does not by itself equal causal influence on the final prediction. Residual pathways, value content, and later non-linearities (e.g., layer-norm, feed-forward) can compensate for changes to $A^{(\ell)}$ [352]. As a result, attention should be interpreted with the same care physicists apply to intermediate quantities that depend on gauge or representation: attention is informative — often highly so — provided we check that its salient patterns are stable under benign perturbations [353] and consistent with simple causal probes (e.g., masking the highlighted tokens or heads degrades performance) [354]. Under these routine checks, attention maps become reliable, compact summaries of the model's internal routing, effective for communicating which correlations are being prioritized.

Early stress tests [355] clarified why such checks are needed. Across standard NLP tasks, it was shown that raw attention weights can be only weakly correlated with gradient/perturbation measures [352, 355] of importance and that one can often construct substantially different attention patterns that leave the output nearly unchanged [355, 356]. These findings do not negate the value of attention; rather, they delimit when raw weights *alone* are insufficient and motivate joint interpretation in conjunction with sensitivity analyses or simple interventions to confirm that the highlighted routes matter for the prediction.

Subsequent work [357] reframed the question from "is attention the explanation?" to "under what conditions is it a *useful* explanation?" Practical diagnostics emerged: compare against a uniform-weights baseline [357]; assess stability across random seeds [358]; probe with encoders whose parameters are frozen to separate the contribution of attention from the rest of the network [357]; and test adversarially perturbed attention to ensure that plausible-looking maps still carry functional weight [356]. When attention passes such checks, the maps can be both human-readable and decision-relevant.

Later syntheses systematized these ideas, distinguishing *plausibility* (human-readable alignment with domain structure) from *faithfulness* (measured counterfactual dependence) [359], and recommending enhanced methodology via e.g. supervised effective attention [360,361]. The emerging consensus for practitioners is pragmatic: treat attention as an explanatory lens that is credible when triangulated with simple audits and unreliable when those audits fail.

This viewpoint balances value and caution. In many practical systems, the heads that persist across training seeds and whose masking measurably reduces task performance also display coherent, domain-aligned patterns in their attention maps [362]. Conversely, diffuse maps, special-token *sinks* [363], or maps that vary erratically across runs are best treated as descriptive diagnostics rather than explanations.

Framed this way, attention provides an accessible explanatory lens while remaining compatible with stricter notions of faithfulness.

Synthesis Across Multiple Attention Layers

For multi-layer models, single-layer maps can be misleading because they only show one step of routing. A widely used remedy is **attention rollout** [364–366], which aggregates routing across depth. If we write a residual block schematically as

$$\boldsymbol{\Sigma}_i^{(\ell+1)} = \alpha^{(\ell)} \sum_j A_{ij}^{(\ell)} \boldsymbol{\mathcal{V}}_j^{(\ell)} + (1-\alpha^{(\ell)}) \boldsymbol{\Sigma}_i^{(\ell)}, \tag{2.74}$$

with $\boldsymbol{\Sigma}_i^{(\ell)}$ the block input and $\alpha^{(\ell)} \in [0,1]$ a shorthand mixing coefficient (ignoring position-wise projection/LayerNorm/MLP effects), then an *effective* end-to-end routing operator is approximated by

$$\mathcal{R} = \prod_{\ell=L}^{1} \left[(1-\alpha^{(\ell)}) \mathbb{1} + \alpha^{(\ell)} A^{(\ell)} \right], \tag{2.75}$$

multiplying from last to first layer. Visualizing rows of $\mathcal{R}$ yields a global attention map that captures multi-hop pathways and typically aligns more closely with perturbation-based importance than any single $A^{(\ell)}$. For multi-head attention, $A^{(\ell)}$ denotes the head-pooled map (e.g., mean or value-weighted aggregation), which preserves the notion of a single routing operator per layer while retaining head-level diagnostics if needed. In practice, rollout is simple to compute, stable across small perturbations, and offers an intuitive *transport* picture: information travels through a sequence of stochastic *propagations* plus identity *bypasses*, so end-to-end influence corresponds to the composition of propagation channels defined by each layer.

A complementary construct is **attention flow** [364,367], which treats the stack as a directed graph and measures how much of an output unit's mass can be *sent back* to inputs under the network's routing constraints. Concretely, each layer contributes edges with capacities given by the attention weights; residual connections add identity edges. One then computes, for a fixed output query, the distribution over inputs obtained by pushing a unit of flow backward through this capacitated graph. Intuitively, attention flow treats attention weights as *maximum allowable traffic* along each edge and asks how much of the output's influence can be traced back to each input when respecting these limits.

Attention flow is defined with respect to a specific output and respects the multi-layer nature of the computation; unlike generic gradients, it remains in the same *units/dimensions* as attention and is easier to compare with rollout. Empirically, flow and rollout tend to agree on dominant routes while differing in regions where residual propagation causes strong path competition — usually the regimes which may warrant deeper inspection.

Other tools are often used as cross-checks rather than replacements. Gradient-augmented variants (e.g., attention-gradient integration [368] or relevance propaga-

tion through attention [369]) take importance measures into account and sharpen maps near decision boundaries; head-importance and pruning [354, 362] quantify how much performance depends on particular heads; and perturbation protocols (deletion/insertion curves [370], remove-and-retrain controls [371]) provide scalar faithfulness scores. In combination with rollout or flow, these methods integrate attention information with genuinely influential routes, enhancing the reliability of the interpretation of attention mechanisms.

Chapter 3
Findings on the Hubbard Model

This chapter presents results on the 2D Hubbard model obtained through a coherent numerical program that interrogates the model across the entire temperature axis and subjects the generated many-body data to AI-based analysis. The three constituent components — ground-state iPEPS, finite-temperature XTRG, and an optimized encoder-only transformer — collectively establish an integrated pipeline for quantum many-body research.

First, symmetry-preserved iPEPS provide controlled access to the landscape of zero-temperature competing orders, with the global symmetry serving as a *control knob* that can admit or suppress static antiferromagnetism so as to expose the uniform superconducting phase and its competitors. This capability helps clarify the long-standing debate surrounding the existence and strength of superconductivity in the Hubbard model and offers a glimpse into a potential pairing mechanism driven by spin frustration.

Second, an enhanced $1s^+$ XTRG scheme pushes thermal simulations deep into the ultracold regime, thereby establishing a continuous bridge between finite-T and $T \to 0$ information. This enables the construction of a partial temperature-doping phase diagram of the Hubbard model. Notably, we corroborate the superconducting features identified in the aforementioned ground-state studies and scrutinize the pseudogap regime characterized by suppressed spin susceptibilities.

Third, we develop an attention-based AI tailored to the analysis of quantum many-body data. Acting as an automated analyzer, this AI model ingests finite-T snapshots generated from XTRG and extracts salient correlation patterns, enabling a universal *omnimeter* that infers calibrated quantities (e.g., temperature, charge doping, etc.) from the perceived correlations in the experimental snapshots. Together, these elements constitute a comprehensive toolbox for the Hubbard model and are readily extensible to other quantum lattice systems.

List of Publications

The findings presented in this chapter are based on the following publications:

- *Frustration-Induced Superconductivity in the t-t' Hubbard Model*
  Changkai Zhang, Jheng-Wei Li, Dimitra Nikolaidou, and Jan von Delft
  Section 3.1 / arXiv: 2307.14835 Phys. Rev. Lett. **134**, 116502 (2025)

- *Finite-Temperature Study of the Hubbard Model via Enhanced Exponential Tensor Renormalization Group*
  Changkai Zhang and Jan von Delft
  Section 3.2 / arXiv: 2510.25022 intended for Phys. Rev. B

- *Interpretable Artificial Intelligence (AI) Analysis of Strongly Correlated Electrons*
  Changkai Zhang and Jan von Delft
  Section 3.3 / arXiv: 2510.26864 intended for Phys. Rev. X

3.1 Ground State Properties and Superconductivity

The long-standing puzzle of whether the repulsive 2D Hubbard model can sustain a uniform d-wave superconducting phase is complicated by the prevalence of stripe order — unidirectional modulations of charge and spin — that frequently dominates in state-of-the-art numerics. This work highlights the role of *magnetic frustration*. By allowing an NNN hopping t' and by increasing charge doping, one frustrates AFM correlations that scaffold stripe textures. We implement a symmetry-resolved iPEPS protocol that runs, in parallel and under otherwise matched conditions, (i) a $\mathrm{U}(1)$ symmetric ansatz that admits static AFM order and thus stripes, and (ii) an $\mathrm{SU}(2)$ symmetric ansatz that forbids static spin order and hence exposes the best competing uniform state. The head-to-head comparison yields a conceptually sharp conclusion: as frustration is increased, stripe order loses its energetic advantage and a uniform state with robust d-wave singlet pairing becomes preferred. In other words, frustration tips the balance toward a uniform superconducting phase by penalizing magnetism-based spatial ordering.

A notable methodological advance is the design of a *controlled comparison* rather than an isolated variational search. Because both $\mathrm{U}(1)$ and $\mathrm{SU}(2)$ simulations use the same algorithmic pipeline, differences in energy and correlations can be attributed to the presence or absence of static AFM order only. This strategy furnishes a useful paradigm for disentangling intertwined orders in correlated-electron models. On the experimental side, it suggests concrete targets for ultracold atom simulators of the Hubbard model: by engineering NNN tunneling and charge density, one can realize frustrated backgrounds in which singlet pairing should be maximized.

The study helps reconcile a substantial body of work reporting strong stripe tendencies in the underdoped Hubbard model and closely related t-J models. Landmark QMC, DMRG and PEPS calculations have found robust stripes at and near $1/8$ doping, often outcompeting uniform SC under minimal settings. By identifying frustration as a lever that selectively destabilizes spin orders, the present results rationalize why increased doping or long-range hopping leads to enhanced SC. The explicit demonstration that positive t'/t favors a uniform SC state complements cylinder DMRG studies which are heavily influenced by finite-size effects.

The results also sharpen the open problem regarding the electron-hole asymmetry. A positive t'/t — loosely associated with electron-doped cuprates according to DFT downfolding — emerges here as notably pairing-friendly, whereas empirically many hole-doped compounds achieve higher critical temperatures; bridging this gap likely requires augmenting minimal models with additional ingredients such as density-assisted hopping. On the other hand, this mismatch likely suggests that the underlying mechanism for unconventional superconductivity in cuprates may have little relevance to the band structure in the material.

Frustration-Induced Superconductivity in the t-t' Hubbard Model

Changkai Zhang (张昌凯),[1] Jheng-Wei Li,[2] Dimitra Nikolaidou,[1] and Jan von Delft[1]
[1]Arnold Sommerfeld Center for Theoretical Physics, Ludwig-Maximilians-Universität München, 80333 Munich, Germany
[2]Université Grenoble Alpes, CEA, Grenoble INP, IRIG, Pheliqs, F-38000 Grenoble, France

(Received 22 August 2023; revised 5 December 2024; accepted 24 February 2025; published 18 March 2025)

The two-dimensional Hubbard model is widely believed to capture key ingredients of high-T_c superconductivity in cuprate materials. However, compelling evidence remains elusive. In particular, various magnetic orders may emerge as strong competitors of superconducting orders. Here, we study the ground state properties of the doped two-dimensional t-t' Hubbard model on a square lattice via the infinite projected entangled-pair state method with U(1) or SU(2) spin symmetry. The former is compatible with antiferromagnetic orders, while the latter forbids them. Therefore, we obtain by comparison a detailed understanding of the magnetic impact on superconductivity. Moreover, an additional t' term accommodates the particle-hole asymmetry, which facilitates studies on the discrepancies between electron- and hole-doped systems. We demonstrate that (i) a positive t'/t significantly amplifies the strength of superconducting orders; (ii) at sufficiently large doping levels, the t-t' Hubbard model favors a uniform state with superconducting orders instead of stripe states with charge and spin modulations; and (iii) the enhancement of magnetic frustration, by increasing either the strength of next-nearest neighbor interactions or the charge doping, impairs stripe orders and helps stabilize superconductivity.

DOI: 10.1103/PhysRevLett.134.116502

Introduction—Despite continuous efforts during the past few decades, the physics of high-T_c superconductivity in cuprate materials [1] remains unclear [2,3]. The two-dimensional (2D) Hubbard model [4] on a square lattice is believed to capture the essential low-energy features of cuprates. Various numerical methods [5–10] have been used to tackle this issue. Nevertheless, previous computational attempts generate numerous candidate ground states [11,12] very close in energies with abundant combinations of charge and spin orders. Experiments [13–20] also confirm simultaneous charge and spin modulated states coexisting or competing with superconductivity. This triggers our curiosity on the interplay between the antiferromagnetic (AFM) background and the high-T_c superconductivity in cuprates.

Typical candidates encompass a uniform state [21–38] and various stripe states [11,33,39–52]. The former features a uniform charge density and is commonly associated with d-wave superconductivity, while the latter often exhibit charge-density and spin-density waves with diverse periods, with only part of them displaying coexisting superconductivity. For the nearest neighbor (NN) minimal Hubbard model, a series of advanced numerical methods reached a consensus [11] that the ground state at 1/8 hole doping is a filled (one hole per unit cell of the charge order) period 8 stripe state devoid of superconducting orders. The half-filled period 4 stripe state [13,20,53] favored more in, e.g., LaSrCuO materials emerges primarily with negative next-nearest neighbor (NNN) hopping, as demonstrated in numerous computational simulations [48,51,54–59]. This motivates our investigations beyond the minimal Hubbard model.

Concurrently, multiple recent studies [60–62] focusing on the extended t-J model have uncovered substantially more robust superconducting orders in electron-doped settings as opposed to hole-doped configurations, a finding that contradicts experimental observations. Explorations of the extended Hubbard model using the density matrix renormalization group have yielded inconsistent outcomes [63,64], further underscoring the significance of research beyond the minimal Hubbard model.

In this Letter, we use the infinite projected entangled-pair state (iPEPS) [9,10] ansatz and simple update algorithm [65] to study the ground state properties of the t-t' Hubbard model. Our iPEPS ansatz is less susceptible to finite-size effects than the density matrix renormalization group on cylinders. Leveraging our cutting-edge QSpace tensor library [66,67], we are capable of conducting simulations with U(1) or SU(2) spin symmetry, where the former admits local magnetic moments and the latter forbids them. This allows us to scrutinize the impact of magnetic orders on pairing properties. Our simulations demonstrate that (i) a positive t'/t significantly amplifies the strength of

superconducting orders; (ii) at sufficiently large doping, the t-t' Hubbard model favors an SU(2) uniform state with d-wave pairing orders instead of a U(1) stripe state in [51]; and (iii) the enhancement of magnetic frustration, by increasing either the strength of NNN interactions or the charge doping, impairs stripe orders and helps stabilize superconductivity.

Model—The 2D t-t' Hubbard model on a square lattice is defined via the following Hamiltonian:

$$\mathcal{H} = -\sum_{i,j,\sigma} t_{ij}[c^{\dagger}_{i\sigma}c_{j\sigma} + \text{H.c.}] + U\sum_{i} n_{i\uparrow}n_{i\downarrow}. \qquad (1)$$

Here, $t_{ij} = t$ or t' for NN or NNN, respectively, and zero otherwise; U measures the on-site Coulomb repulsion. Throughout this Letter, we use $U/t = 10$, as established to be realistic for cuprate materials [68,69], and set $t = 1$ for convenience.

Method—In our computations, we apply the fermionic iPEPS [70–77] ansatz, a tensor network method targeting 2D lattice models, to simulate the t-t' Hubbard model in the thermodynamic limit. The ansatz exploits translational symmetry by assuming that the infinite tensor network consists of periodically repeated supercells of tensors. Each supercell comprises several rank-5 tensors with one physical index carrying states in the local Hilbert space, and four auxiliary indices connecting neighboring sites. The accuracy of the simulation can be controlled by the bond dimensions of the auxiliary indices. Different supercell sizes yield stripe states with different periods in charge or spin orders. Previous research [51,60,63] on the Hubbard model or the t-J model has identified stripe states with period 4 charge orders as a representative stripe state. Therefore, we hereby focus on the period 4 stripe state. Further discussions and details regarding stripes with longer periods can be found in Supplemental Material [78].

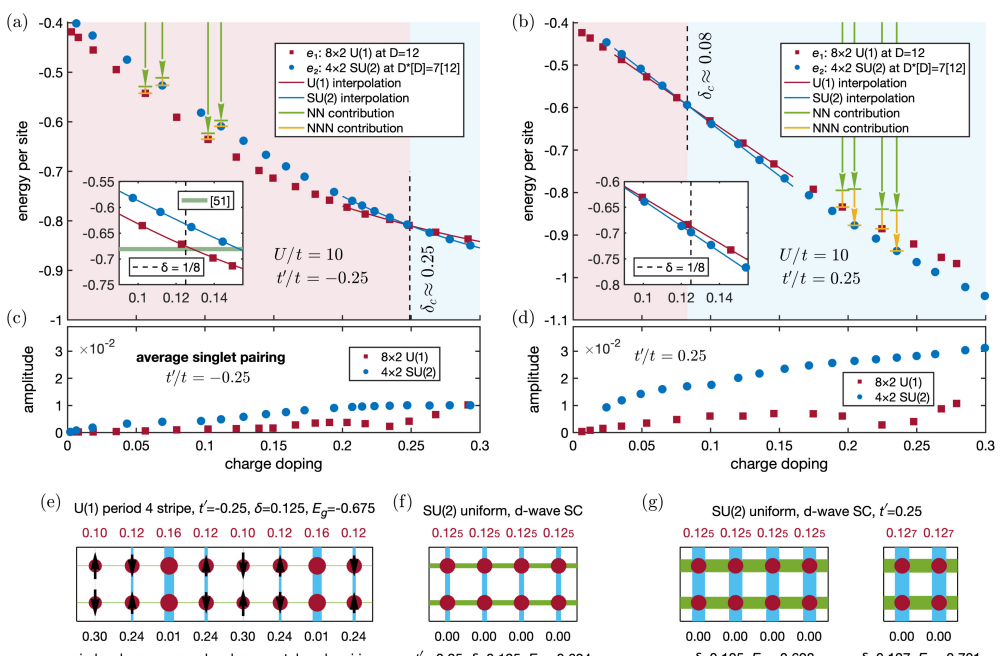


FIG. 1. The ground state energy per site (a), (b) and singlet pairing (c), (d) vs doping δ of the t-t' Hubbard model at $U/t = 10$ and (a), (c) $t'/t = -0.25$ or (b), (d) $t'/t = 0.25$, computed via U(1) iPEPS (red squares) on an 8×2 supercell at bond dimension $D = 12$ and SU(2) iPEPS (blue circles) on a 4×2 supercell keeping $D^* = 7$ multiplets (bond dimension $D = 12$). Green and yellow arrows, respectively, indicate the NN (including on site) and NNN contributions to the energy for several typical data points. Inset: enlargement of the region near $1/8$ doping. (e)–(g) Details of the U(1) and SU(2) symmetric ground states on 8×2, 4×2, and 2×2 supercells. Radii of red circles and lengths of black arrows are proportional to the charge density (top rows) and the local moments (bottom rows), respectively. Bond widths indicate NN singlet pairing amplitudes and two different colors indicate opposite signs. For (f), (g), we used $D^*[D] = 8[13]$ for reasons explained in Supplemental Material [78].

The optimization is performed via imaginary time evolution [79] in which projector $\exp\{-\tau(\mathcal{H}-\mu N)\}$ (τ is a small number, $\mathcal{H}$ the Hamiltonian, μ the chemical potential, and N the charge density) is repeatedly applied to some random initial state until the ground state energy converges. Models with NNN interactions are computationally very expensive. Therefore, we choose the simple update scheme [65,70,71] for a balance between accuracy and computational complexity. Observables are extracted by contracting the tensor network using the corner transfer matrix method [70,72,80–83]. The QSpace tensor library [66,67,84] is used to implement either U(1) or SU(2) spin symmetry.

The U(1) iPEPS simulations are conducted on an 8×2 supercell at bond dimension $D=12$. This is required for capturing the period 4 charge orders, as the corresponding spin order periods are typically twice as long as the charge periods. The SU(2) iPEPS simulations are performed on a 4×2 or 2×2 supercell by keeping $D^*=7$ symmetry multiplets (corresponding to a bond dimension $D=12$) [66]. Spin orders are suppressed upon enforcing SU(2) symmetry, making a 4×2 supercell adequate to detect any potential period 4 orders, while the 2×2 supercell is employed to ascertain the uniformity of the ground state. Charge doping is adjusted by tuning the chemical potential.

Energetics—Figures 1(a) and 1(b) show the ground state energy per site of the t-t' Hubbard model as a function of doping under $U/t=10$ and $t'/t=\mp 0.25$, computed via the U(1) and SU(2) iPEPS and denoted as e_1 (red) and e_2 (blue), respectively. Figures 1(c) and 1(d) show the corresponding singlet pairing amplitudes. Figures 1(e) and 1(f) display, respectively, the detailed characteristics of the U(1) and SU(2) ground states with a negative t'/t at the predominantly studied 1/8 doping. Figure 1(g) presents SU(2) ground states with a positive t'/t, showcasing numerically significant d-wave singlet pairing orders.

Utilizing an 8×2 supercell, our U(1) iPEPS generates a nonsuperconducting stripe state with a period 4 charge-density wave and a period 8 antiferromagnetically ordered spin-density wave. These attributes, along with the ground state energy acquired, are generally consistent with the findings in [51]. By contrast, when we enforce the SU(2) symmetry and suppress spin orders, we find a uniform state without any charge orders, at odds with finite-size studies [56,60,63]. Moreover, strong d-wave pairing emerges for positive t'/t, which implies superconductivity. The SU(2) iPEPS on 4×2 and 2×2 supercells produces physically identical states, confirming the uniformity of the ground state.

Near zero doping, we find $e_2>e_1$. This is consistent with the well-established fact that the Heisenberg model on a square lattice has an AFM ground state that breaks SU(2) symmetry. However, as the doping increases, e_2 decreases faster than e_1. They intersect at $\delta_c\approx 0.25$ for $t'/t=-0.25$ and $\delta_c\approx 0.08$ for $t'/t=0.25$ (first order transition), as depicted in Figs. 1(a) and 1(b), in agreement with prior observations [51] that a negative or positive t'/t favors stripe or uniform states, respectively. Intuitively, a positive t'/t promotes diagonal hopping of the doped charges, which in turn disrupts the AFM background in the vicinity of the domain wall within the stripe states, rendering the presence of domain walls less desirable [45].

The lower energy of the SU(2) relative to the U(1) ground state at large doping can be understood as the result of magnetic frustration induced by the NNN hoppings. The U(1) stripe state still accommodates AFM orders and thus suffers strongly from magnetic frustrations with NNN hopping. By contrast, the SU(2) uniform state is less frustrated since it hosts no local spin orders. Indeed, the NNN terms contribute much less to lowering the energy e_1 of the stripe state than to the energy e_2 of the uniform state, as indicated via the yellow arrows in Figs. 1(a) and 2(b).

This issue is further elaborated in Figs. 2(a) and 2(b), showing the contribution of NN (including on site) and NNN terms to the total energy per site as a function of doping, respectively. Throughout the entire doping range in

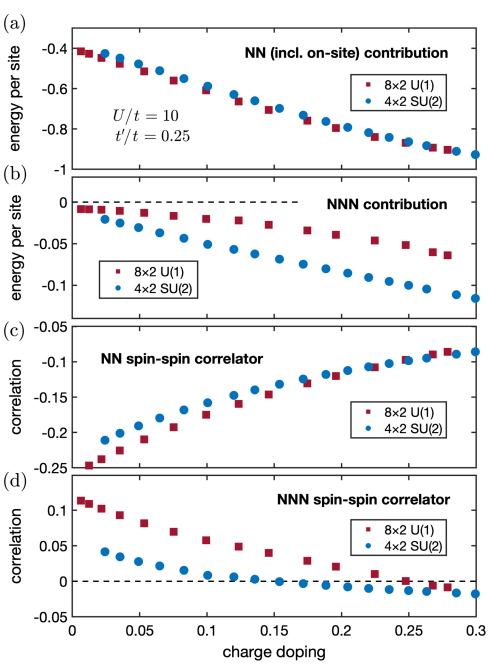


FIG. 2. The contribution of (a) the NN (including on site) and (b) the NNN terms to the total energy per site in the U(1) and SU(2) ground states, respectively, as a function of doping. (c) The NN and (d) NNN spin-spin correlators in the U(1) and SU(2) ground states, respectively.

our study, the NN contribution is marginally lower in the U(1) states than in the SU(2) states. Conversely, the NNN contribution is substantially lower in the SU(2) than the U(1) cases, ultimately leading to a lower overall energy for the SU(2) states at large doping levels. As a comparison, Figs. 2(c) and 2(d) show the NN and NNN spin-spin correlators, respectively. The NN correlations stay negative for both the U(1) and SU(2) states, reflecting the overall AFM background. The NNN correlations, however, turn negative considerably sooner for the SU(2) states than for the U(1) states, echoing the findings in ultracold atom experiments that doped charges drive the NNN spin-spin correlation negative [85–89]. This indicates that the SU(2) state better reconciles the magnetic frustration, thereby achieving a lower NNN energy. Such behaviors exemplify how the enhancement of magnetic frustration through NNN hopping inhibits the formation of stripes and promotes the emergence of superconductivity.

Pairing order—The superconducting order can be characterized by the singlet pairing amplitude $\Delta_{\mathbf{r},\mathbf{s}} = \langle c_{\mathbf{r}\uparrow}c_{\mathbf{s}\downarrow} - c_{\mathbf{r}\downarrow}c_{\mathbf{s}\uparrow}\rangle$. Specifically, we focus on the NN singlet pairing. As illustrated in Figs. 1(e)–1(g), we observe finite singlet pairing orders for both U(1) and SU(2) ground states. However, the pairing amplitude (averaged over the supercell) of the SU(2) states can be substantially larger than that in the U(1) states throughout the entire doping range for positive t'/t, as presented in Fig. 1(d). This can be attributed to the fact that the SU(2) iPEPS is, by construction, a spin-singlet state. Indeed, the latter can be interpreted as a generalized version of the resonating valence bond state [90]. Therefore, the existence of d-wave pairing order is reminiscent of Anderson's original resonating valence bond proposal [91,92].

Moreover, we discover that the singlet pairing for positive t'/t can be considerably larger than that for negative t'/t. Intuitively, this could be perceived as pair formation being enhanced (reduced) by the constructive (destructive) interference between NN and NNN hopping at positive (negative) t'/t [93]. This is in line with prior findings in the extended t-J model [60–62,94] and Hubbard model [63] using density matrix renormalization group. Electronic structure analysis [68,69,95,96] suggests that positive (negative) t'/t corresponds to electron (hole) doped cuprates. Consequently, the numerics so far yield outcomes that are opposite to the experimental observations, where hole-doped cuprates exhibit stronger superconductivity. This emphasizes the necessity for further investigations regarding the appropriate parameter settings in the effective models [61,97,98].

Long-range order—Figure 3(a) displays the long-range spin-spin $S_{\mathbf{rs}} = \langle \mathbf{S}_\mathbf{r} \cdot \mathbf{S}_\mathbf{s}\rangle - \langle \mathbf{S}_\mathbf{r}\rangle \cdot \langle \mathbf{S}_\mathbf{s}\rangle$ and pair-pair $P_{\mathbf{rs}} = \langle \Delta^\mathbf{y}_\mathbf{r}\Delta^\mathbf{y}_\mathbf{s}\rangle - \langle \Delta^\mathbf{y}_\mathbf{r}\rangle\langle \Delta^\mathbf{y}_\mathbf{s}\rangle$ (where $\Delta^{\boldsymbol{\alpha}}_\mathbf{r} = \Delta_{\mathbf{r},\mathbf{r}+\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha} = \mathbf{x},\mathbf{y}$ is the horizontal or vertical unit vector) correlators for two specific ground states with U(1) or SU(2) symmetry for $t'/t > 0$. Figure 3(b) shows the corresponding correlation lengths. Our data indicate that all these correlators decay exponentially, and the correlation lengths never exceed two

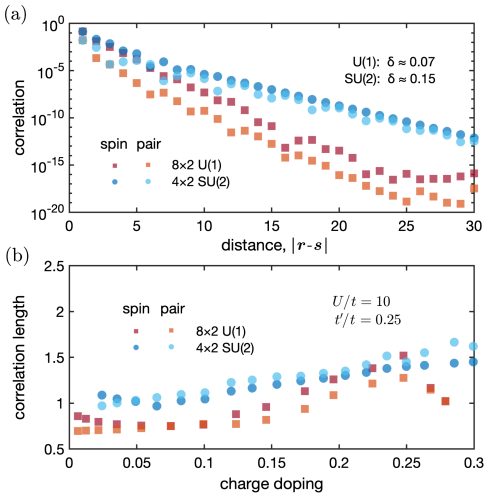


FIG. 3. (a) The long-range spin-spin and pair-pair correlators in the U(1) and SU(2) ground states, respectively. All these correlators exhibit an exponential decay behavior. (b) The corresponding correlation lengths as a function of doping.

units throughout the entire doping range. This suggests no connected long-range spin or pairing orders in both scenarios. Accordingly, a minor local pairing order sufficiently signals weak superconductivity in the stripe states.

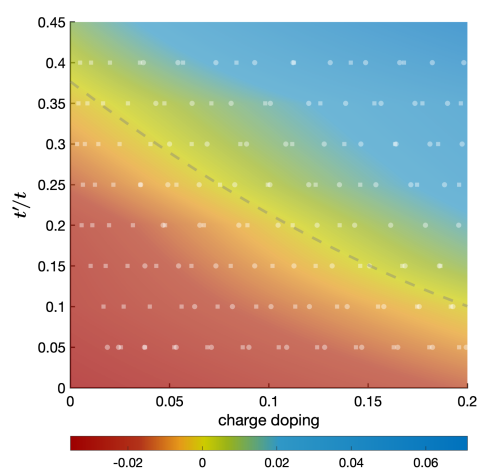


FIG. 4. The ground state phase diagram of the t-t' Hubbard model with respect to doping and t'/t. The color scale indicates $e_1 - e_2$, obtained via linear interpolation from a discrete set of scanning points (white). The gray dashed line marks $e_1 = e_2$.

Phase diagram—Figure 4 presents a schematic ground state *phase diagram* for $t'/t > 0$ derived via linear interpolation from a discrete set of scanning points. The U(1) stripe states are energetically favored in the bottom-left corner, and the SU(2) uniform states in the top-right corner. This is largely consistent with previous studies of the t-t'-J model [60]. Therefore, an increase of either charge doping or the NNN hopping, which both intensify magnetic frustration, will drive the ground state from striped to uniform states. Recall that the uniform ground states are typically accompanied by strong superconductivity. The phase diagram thus supports the conclusion that the enhancement of magnetic frustration helps stabilize superconductivity.

Discussion—In this research, we have studied the ground state properties and the phase diagram of the t-t' Hubbard model via the U(1) and SU(2) symmetric iPEPS method. We discovered an SU(2) uniform state with strong d-wave superconducting orders, with a lower energy than the striped U(1) states at large doping levels. Although the variational space of the U(1) iPEPS is larger than that of the SU(2) iPEPS, the fact that the U(1) iPEPS has so far failed to yield a uniform ground state suggests that the U(1) iPEPS has difficulty handling the subspace devoid of magnetic orders. This highlights the importance of exploring quantum states with several different global symmetries in tensor network simulations. We note, however, that it is possible to recover the SU(2) ground states via a U(1) implementation with *a priori* guidance about the SU(2) compatible settings; see Supplemental Material [78] for more details.

Also, we have demonstrated the interplay between local magnetic orders and superconductivity. The additional NNN interaction terms introduce extra magnetic frustration and help suppress the AFM orders, favoring strong d-wave superconductivity at large doping levels. Besides, a positive t'/t frustrates the domain walls and stimulates pair formation. This suggests that the superconductivity in cuprate materials can be enhanced, and T_c incremented, by elevating the strength of NNN hopping.

Outlook—The novel SU(2) ground state, expressed in terms of an iPEPS tensor network, contains information on dominant contributions from the many-body Hilbert space. Consequently, it is possible to generate various *snapshots* of the type accessible via quantum gas microscopy in the ultracold atom experiments [86,87], enabling a direct comparison with experimental analysis [89,99]. Such information would facilitate further investigations regarding the dopant mobility through high-order correlators [100,101] or string patterns using suitable pattern recognition algorithms [88]. Also, similar SU(2) symmetric tensor techniques can be applied to some thermal tensor network methods, such as finite temperature PEPS [102–104], exponential tensor renormalization group [99,105,106], or tangent space tensor renormalization group [107], to explore physics at finite temperatures where strange metal behavior is observed experimentally.

Acknowledgments—We thank our colleagues Andreas Weichselbaum and Andreas Gleis for stimulating discussions and technical suggestions, which led to a significant speedup of the algorithms. We also thank Philippe Corboz for helpful feedback on a preliminary version of this work, as well as Zi Yang Meng for constructive comments. This research was funded in part by the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy EXC-2111 (Project No. 390814868), and is part of the Munich Quantum Valley, supported by the Bavarian state government through the Hightech Agenda Bayern Plus.

[1] J. G. Bednorz and K. A. Müller, Z. Phys. B **64**, 189 (1986).
[2] B. Keimer, S. A. Kivelson, M. R. Norman, S. Uchida, and J. Zaanen, Nature (London) **518**, 179 (2015).
[3] P. A. Lee, N. Nagaosa, and X.-G. Wen, Rev. Mod. Phys. **78**, 17 (2006).
[4] J. Hubbard, Proc. R. Soc. A **296**, 82 (1967).
[5] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).
[6] R. Blankenbecler, D. J. Scalapino, and R. L. Sugar, Phys. Rev. D **24**, 2278 (1981).
[7] G. Sugiyama and S. Koonin, Ann. Phys. (N.Y.) **168**, 1 (1986).
[8] G. Knizia and Garnet Kin-Lic Chan, Phys. Rev. Lett. **109**, 186404 (2012).
[9] F. Verstraete and J. I. Cirac, arXiv:cond-mat/0407066.
[10] J. Jordan, R. Orús, G. Vidal, F. Verstraete, and J. I. Cirac, Phys. Rev. Lett. **101**, 250602 (2008).
[11] B.-X. Zheng, C.-M. Chung, P. Corboz, G. Ehlers, M.-P. Qin, R. M. Noack, H. Shi, S. R. White, S. Zhang, and G. K.-L. Chan, Science **358**, 1155 (2017).
[12] J. P. F. LeBlanc *et al.*, Phys. Rev. X **5**, 041041 (2015).
[13] J. M. Tranquada, B. J. Sternlieb, J. D. Axe, Y. Nakamura, and S. Uchida, Nature (London) **375**, 561 (1995).
[14] J. M. Tranquada, J. D. Axe, N. Ichikawa, Y. Nakamura, S. Uchida, and B. Nachumi, Phys. Rev. B **54**, 7489 (1996).
[15] R. J. Birgeneau, C. Stock, J. M. Tranquada, and K. Yamada, J. Phys. Soc. Jpn. **75**, 111003 (2006).
[16] G. Ghiringhelli *et al.*, Science **337**, 821 (2012).
[17] R. Comin and A. Damascelli, Annu. Rev. Condens. Matter Phys. **7**, 1 (2015).
[18] T. Wu, H. Mayaffre, S. Krämer, M. Horvatić, C. Berthier, W. N. Hardy, R. Liang, D. A. Bonn, and M.-H. Julien, Nature (London) **477**, 191 (2011).
[19] T. Wu, H. Mayaffre, S. Krämer, M. Horvatić, C. Berthier, W. Hardy, R. Liang, D. Bonn, and M.-H. Julien, Nat. Commun. **6**, 6438 (2015).
[20] A. Mesaros, K. Fujita, S. D. Edkins, M. H. Hamidian, H. Eisaki, S.-i. Uchida, J. C. S. Davis, M. J. Lawler, and E.-A. Kim, Proc. Natl. Acad. Sci. U.S.A. **113**, 12661 (2016).
[21] T. Giamarchi and C. Lhuillier, Phys. Rev. B **43**, 12943 (1991).
[22] E. Dagotto, Rev. Mod. Phys. **66**, 763 (1994).

[23] C. J. Halboth and W. Metzner, Phys. Rev. Lett. **85**, 5162 (2000).
[24] T. A. Maier, M. Jarrell, T. C. Schulthess, P. R. C. Kent, and J. B. White, Phys. Rev. Lett. **95**, 237001 (2005).
[25] M. Capone and G. Kotliar, Phys. Rev. B **74**, 054513 (2006).
[26] D. Eichenberger and D. Baeriswyl, Phys. Rev. B **76**, 180504(R) (2007).
[27] M. Aichhorn, E. Arrigoni, M. Potthoff, and W. Hanke, Phys. Rev. B **76**, 224509 (2007).
[28] L. F. Tocchio, F. Becca, A. Parola, and S. Sorella, Phys. Rev. B **78**, 041101(R) (2008).
[29] S. S. Kancharla, B. Kyung, D. Sénéchal, M. Civelli, M. Capone, G. Kotliar, and A.-M. S. Tremblay, Phys. Rev. B **77**, 184516 (2008).
[30] G. Sordi, P. Sémon, K. Haule, and A. M. S. Tremblay, Phys. Rev. Lett. **108**, 216401 (2012).
[31] E. Gull and A. J. Millis, Phys. Rev. B **86**, 241106(R) (2012).
[32] H. Yokoyama, M. Ogata, Y. Tanaka, K. Kobayashi, and H. Tsuchiura, J. Phys. Soc. Jpn. **82**, 014707 (2012).
[33] J. Kaczmarczyk, J. Spałek, T. Schickling, and J. Bünemann, Phys. Rev. B **88**, 115127 (2013).
[34] E. Gull, O. Parcollet, and A. J. Millis, Phys. Rev. Lett. **110**, 216405 (2013).
[35] K.-S. Chen, Z. Y. Meng, S.-X. Yang, T. Pruschke, J. Moreno, and M. Jarrell, Phys. Rev. B **88**, 245110 (2013).
[36] J. Otsuki, H. Hafermann, and A. I. Lichtenstein, Phys. Rev. B **90**, 235132 (2014).
[37] Y. Deng, E. Kozik, N. V. Prokof'ev, and B. V. Svistunov, Europhys. Lett. **110**, 57001 (2015).
[38] L. F. Tocchio, F. Becca, and S. Sorella, Phys. Rev. B **94**, 195126 (2016).
[39] D. Poilblanc and T. M. Rice, Phys. Rev. B **39**, 9749 (1989).
[40] J. Zaanen and O. Gunnarsson, Phys. Rev. B **40**, 7391 (1989).
[41] S. R. White and D. J. Scalapino, Phys. Rev. Lett. **91**, 136403 (2003).
[42] G. Hager, G. Wellein, E. Jeckelmann, and H. Fehske, Phys. Rev. B **71**, 075108 (2005).
[43] C.-C. Chang and S. Zhang, Phys. Rev. Lett. **104**, 116402 (2010).
[44] H.-H. Zhao, K. Ido, S. Morita, and M. Imada, Phys. Rev. B **96**, 085103 (2017).
[45] E. W. Huang, C. B. Mendl, H.-C. Jiang, B. Moritz, and T. P. Devereaux, npj Quantum Mater. **3**, 22 (2018).
[46] A. S. Darmawan, Y. Nomura, Y. Yamaji, and M. Imada, Phys. Rev. B **98**, 205132 (2018).
[47] T. I. Vanhala and P. Törmä, Phys. Rev. B **97**, 075112 (2018).
[48] K. Ido, T. Ohgoe, and M. Imada, Phys. Rev. B **97**, 045138 (2018).
[49] K. Machida, Physica (Amsterdam) **158C**, 192 (1989).
[50] M. Kato, K. Machida, H. Nakanishi, and M. Fujita, J. Phys. Soc. Jpn. **59**, 1047 (1990).
[51] B. Ponsioen, S. S. Chung, and P. Corboz, Phys. Rev. B **100**, 195141 (2019).
[52] L. F. Tocchio, A. Montorsi, and F. Becca, SciPost Phys. **7**, 021 (2019).
[53] J. M. Tranquada, J. D. Axe, N. Ichikawa, A. R. Moodenbaugh, Y. Nakamura, and S. Uchida, Phys. Rev. Lett. **78**, 338 (1997).
[54] R. Eder, Y. Ohta, and G. A. Sawatzky, Phys. Rev. B **55**, R3414 (1997).
[55] G. B. Martins, R. Eder, and E. Dagotto, Phys. Rev. B **60**, R3716 (1999).
[56] H.-C. Jiang and T. P. Devereaux, Science **365**, 1424 (2019).
[57] B.-X. Zheng and Garnet Kin-Lic Chan, Phys. Rev. B **93**, 035126 (2016).
[58] C.-M. Chung, M. Qin, S. Zhang, U. Schollwöck, and S. R. White, Phys. Rev. B **102**, 041106(R) (2020).
[59] Y.-F. Jiang, J. Zaanen, T. P. Devereaux, and H.-C. Jiang, Phys. Rev. Res. **2**, 033073 (2020).
[60] S. Gong, W. Zhu, and D. N. Sheng, Phys. Rev. Lett. **127**, 097003 (2021).
[61] S. Jiang, D. J. Scalapino, and S. R. White, Proc. Natl. Acad. Sci. U.S.A. **118**, e2109978118 (2021).
[62] S. Jiang, D. J. Scalapino, and S. R. White, Phys. Rev. B **106**, 174507 (2022).
[63] Y.-F. Jiang, T. P. Devereaux, and H.-C. Jiang, arXiv:2303.15541.
[64] H. Xu, C.-M. Chung, M. Qin, U. Schollwöck, S. R. White, and S. Zhang, Science **384**, adh7691 (2024).
[65] H. C. Jiang, Z. Y. Weng, and T. Xiang, Phys. Rev. Lett. **101**, 090603 (2008).
[66] A. Weichselbaum, Ann. Phys. (Amsterdam) **327**, 2972 (2012).
[67] A. Weichselbaum, Phys. Rev. Res. **2**, 023385 (2020).
[68] M. Hirayama, Y. Yamaji, T. Misawa, and M. Imada, Phys. Rev. B **98**, 134501 (2018).
[69] M. Hirayama, T. Misawa, T. Ohgoe, Y. Yamaji, and M. Imada, Phys. Rev. B **99**, 245155 (2019).
[70] P. Corboz, R. Orús, B. Bauer, and G. Vidal, Phys. Rev. B **81**, 165104 (2010).
[71] P. Corboz, J. Jordan, and G. Vidal, Phys. Rev. B **82**, 245119 (2010).
[72] B. Bruognolo, J.-W. Li, J. von Delft, and A. Weichselbaum, SciPost Phys. Lect. Notes **25** (2021).
[73] C. V. Kraus, N. Schuch, F. Verstraete, and J. I. Cirac, Phys. Rev. A **81**, 052338 (2010).
[74] T. Barthel, C. Pineda, and J. Eisert, Phys. Rev. A **80**, 042333 (2009).
[75] F. Verstraete, V. Murg, and J. Cirac, Adv. Phys. **57**, 143 (2008).
[76] P. Corboz, Phys. Rev. B **93**, 045116 (2016).
[77] H.-J. Liao, J.-G. Liu, L. Wang, and T. Xiang, Phys. Rev. X **9**, 031041 (2019).
[78] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.134.116502 for technical details and additional numerical results.
[79] G. Vidal, Phys. Rev. Lett. **98**, 070201 (2007).
[80] P. Corboz, S. R. White, G. Vidal, and M. Troyer, Phys. Rev. B **84**, 041108(R) (2011).
[81] T. Nishino and K. Okunishi, J. Phys. Soc. Jpn. **65**, 891 (1996).
[82] P. Corboz, T. M. Rice, and M. Troyer, Phys. Rev. Lett. **113**, 046402 (2014).
[83] R. Orús and G. Vidal, Phys. Rev. B **80**, 094403 (2009).

[84] A. Weichselbaum, Open Source QSpace 4.0.
[85] J. Koepsell, D. Bourgund, P. Sompet, S. Hirthe, A. Bohrdt, Y. Wang, F. Grusdt, E. Demler, G. Salomon, C. Gross, and I. Bloch, Science **374**, 82 (2021).
[86] J. Koepsell, J. Vijayan, P. Sompet, F. Grusdt, T. A. Hilker, E. Demler, G. Salomon, I. Bloch, and C. Gross, Nature (London) **572**, 358 (2019).
[87] J. Koepsell, S. Hirthe, D. Bourgund, P. Sompet, J. Vijayan, G. Salomon, C. Gross, and I. Bloch, Phys. Rev. Lett. **125**, 010403 (2020).
[88] C. S. Chiu, G. Ji, A. Bohrdt, M. Xu, M. Knap, E. Demler, F. Grusdt, M. Greiner, and D. Greif, Science **365**, 251 (2019).
[89] B.-B. Chen, C. Chen, Z. Chen, J. Cui, Y. Zhai, A. Weichselbaum, J. von Delft, Z. Y. Meng, and W. Li, Phys. Rev. B **103**, L041107 (2021).
[90] J.-W. Li, B. Bruognolo, A. Weichselbaum, and J. von Delft, Phys. Rev. B **103**, 075127 (2021).
[91] G. Baskaran, Z. Zou, and P. Anderson, Solid State Commun. **63**, 973 (1987).
[92] P. W. Anderson, G. Baskaran, Z. Zou, and T. Hsu, Phys. Rev. Lett. **58**, 2790 (1987).
[93] G. B. Martins, J. C. Xavier, L. Arrachea, and E. Dagotto, Phys. Rev. B **64**, 180513(R) (2001).
[94] X. Lu, F. Chen, W. Zhu, D. N. Sheng, and S.-S. Gong, Phys. Rev. Lett. **132**, 066002 (2024).
[95] T. Tohyama and S. Maekawa, Phys. Rev. B **49**, 3596 (1994).
[96] O. Andersen, A. Liechtenstein, O. Jepsen, and F. Paulsen, J. Phys. Chem. Solids **56**, 1573 (1995).
[97] T. Xiang, H. G. Luo, D. H. Lu, K. M. Shen, and Z. X. Shen, Phys. Rev. B **79**, 014524 (2009).
[98] S. Jiang, D. J. Scalapino, and S. R. White, Phys. Rev. B **108**, L161111 (2023).
[99] D.-W. Qu, B.-B. Chen, X. Lu, Q. Li, Y. Qi, S.-S. Gong, W. Li, and G. Su, Phys. Rev. Lett. **133**, 256003 (2024).
[100] A. Bohrdt, Y. Wang, J. Koepsell, M. Kánasz-Nagy, E. Demler, and F. Grusdt, Phys. Rev. Lett. **126**, 026401 (2021).
[101] C. Hubig, A. Bohrdt, M. Knap, F. Grusdt, and I. Cirac, SciPost Phys. **8**, 021 (2020).
[102] P. Czarnik, L. Cincio, and J. Dziarmaga, Phys. Rev. B **86**, 245101 (2012).
[103] P. Czarnik and J. Dziarmaga, Phys. Rev. B **90**, 035144 (2014).
[104] P. Czarnik and J. Dziarmaga, Phys. Rev. B **92**, 035152 (2015).
[105] B.-B. Chen, L. Chen, Z. Chen, W. Li, and A. Weichselbaum, Phys. Rev. X **8**, 031082 (2018).
[106] H. Li, B.-B. Chen, Z. Chen, J. von Delft, A. Weichselbaum, and W. Li, Phys. Rev. B **100**, 045110 (2019).
[107] Q. Li, Y. Gao, Y.-Y. He, Y. Qi, B.-B. Chen, and W. Li, Phys. Rev. Lett. **130**, 226502 (2023).

Supplemental Material – Frustration Induced Superconductivity in the t-t' Hubbard Model

Changkai Zhang (张昌凯),[1] Jheng-Wei Li,[2] Dimitra Nikolaidou,[1] and Jan von Delft[1]

[1]*Arnold Sommerfeld Center for Theoretical Physics, Ludwig-Maximilians-Universität München, 80333 Munich, Germany*
[2]*Université Grenoble Alpes, CEA, Grenoble INP, IRIG, Pheliqs, F-38000 Grenoble, France*

In the Supplemental Material, we discuss (S-I) some technical details for accelerating the optimization of infinite Projected Entangled-Pair State (iPEPS) tensor networks for next-nearest neighbor (NNN) lattice models; (S-II) our optimization strategy; (S-III) the scaling behavior with respect to bond dimensions and error analysis of our U(1) and SU(2) iPEPS algorithms; (S-IV) the consistency checks of the ground state at larger supercell sizes; (S-V) benchmarks of the U(1) and SU(2) iPEPS based on exactly solvable free-fermion models; (S-VI) guided and unguided U(1) iPEPS optimization; (S-VII) artifacts of singlet pairing; (S-VIII) stripes with longer periods; (S-IX) the nearest-neighbor (NN) and on-site contributions; and (S-X) the strategy for selecting optimization methods.

S-I. REFINED SIMPLE UPDATE SCHEME FOR NEXT-NEAREST NEIGHBOR MODELS

In our research, we implement a refined version of simple update based on [71], which leads to tremendous speedup of the algorithm. The update of the nearest neighbor (NN) terms (including on-site) is the same as described in [70]. We modified the NNN update scheme, in a manner suggested by Andreas Weichselbaum as shown in Fig. S1, for further processing.

The major improvement comes from the separation of indices directly involved in the update from the other *spectator* indices not involved in the update. Figure S2 shows the details for treating the diagram in Fig. S1(b). The other diagrams can be performed analogously. Greek letters $\alpha, \beta, \ldots$ label the physical indices. Figure S2(a) describes the construction of tensor Δ_L and the application of corresponding swap gates. Then, we split off the relevant physical and auxiliary indices via singular value decomposition (SVD). Next, we perform analogous actions for Δ_R as depicted in Fig. S2(b). These procedures generate tensors L and R of size $Dd \times D \times d$, where D is the bond dimension of the auxiliary index and d the dimension of local Hilbert space. Hereafter, we apply the Trotter gate, as in Fig. S2(c), and obtain the updated tensors L' and R' via truncated SVD. Then, we restore the original Γ-Λ structure via truncated SVD as shown in Fig. S2(d). Finally, we update all the involved Γ and Λ tensors as described in Fig. S2(e). The last two steps recover the original structure in Fig. S1(b).

The refined simple update scheme has a much better scaling behavior with respect to the bond dimension D. The original version of NNN simple update [71] involves an SVD of complexity $O(D^{11}d^4)$, which is also the dominant contribution to the total complexity of the algorithm. In our refined scheme, the dominant complexity comes from the SVD in Fig. S2(a), which scales as $O(D^7d^3)$, much lower than the complexity of the original simple update scheme.

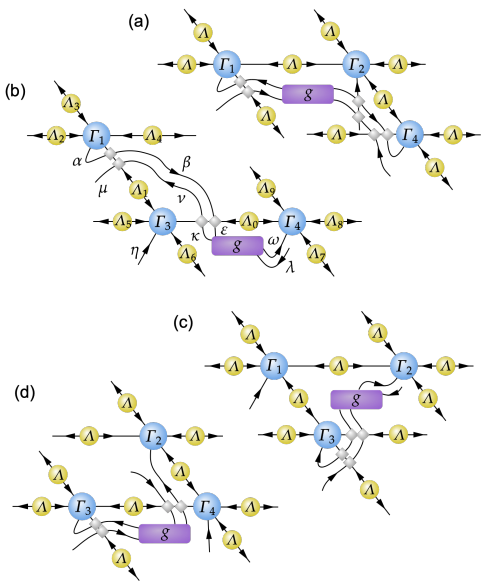


FIG. S1. Tensor network diagrams for the refined simple update scheme for NNN terms. Four diagrams are needed to update the relevant tensors. We use the standard Γ-Λ form of iPEPS tensor network [70]. g is the Trotter gate for the corresponding NNN term. Arrows indicate the flow of quantum numbers [66]. Grey rhombuses depict the fermionic swap gates [70].

S-II. OPTIMIZATION STRATEGY

Our computations start with a random initial iPEPS tensor network state written in the Γ-Λ form [70]. The ground state is attained by repeatedly applying the projector $\exp\{-\tau(\mathcal{H} - \mu N)\}$ (τ is a small number, $\mathcal{H}$ the Hamiltonian (1), μ the chemical potential and N the charge density) to the initial state. The application is broken into a sequence of projectors involving only NN or NNN terms via Suzuki-Trotter decomposition. The Trotter error is of order $O(\tau^2)$. Therefore, we start with a large τ in the beginning so as to approach the ground state quickly, followed by a gradual decrease of τ to improve the accuracy.

Empirically, we find $\tau = 0.1$ a fairly good starting point. During the update iteration, we measure the expectation value of the Trotter gate and compute an estimation of the ground state energy, which serves as an indicator of convergence. τ is reduced

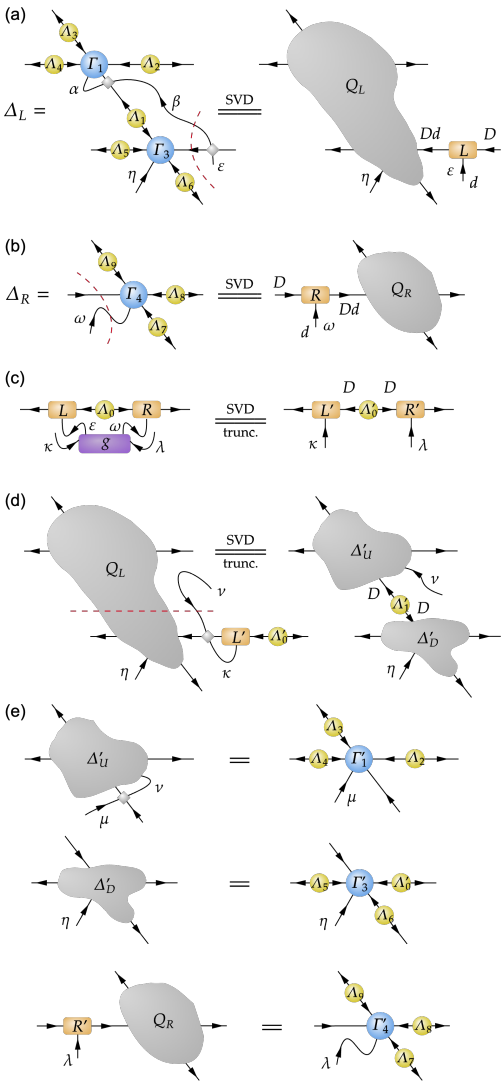


FIG. S2. Realization of the diagram in Fig. S1(b), which updates Λ_0, Λ_1 and all Γs, while leaving all other Λs unchanged. This process *factors out* the environment encoded in Q_L and Q_R, before applying the Trotter gate. (a)(b) Construct tensors Δ_L and Δ_R and split off the relevant physical and auxiliary indices using SVD to obtain tensors L and R. (c) Apply the Trotter gate to tensor L, R and the corresponding Λ. Obtain the updated L', R' and Λ' via truncated SVD. (d) Restore the Γ-Λ structure via truncated SVD. (e) Update all the involved Γ and Λ tensors. Greek labels, in correspondence with Fig. S1(b), keep track of the physical indices.

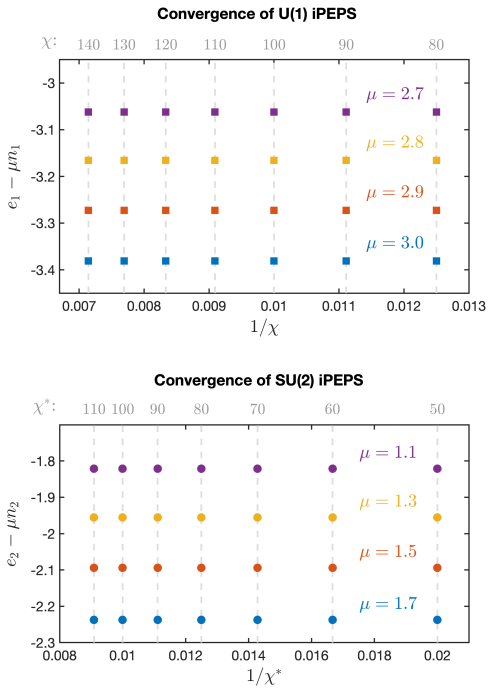


FIG. S3. The convergence characteristics of the U(1) and SU(2) iPEPS with respect to the environmental bond dimension χ and χ^*. The energies per site, $e_1 - \mu n_1$ and $e_2 - \mu n_2$, are well converged starting from $\chi \geqslant 80$ and $\chi^* \geqslant 50$, respectively.

by half once the decrease of estimated ground state energy in an update step drops below τ^2. We regard the convergence to be reached when τ drops below the threshold of 10^{-3}.

The optimization starts with $D = 2$ for U(1) iPEPS and $D^* = 2$ for SU(2) iPEPS. After reaching convergence at the fixed bond dimension, we increment D or D^* by 1 (but see also Sec. S-VI), until we arrive at $D = 12$ for U(1) or $D^* = 7$ for SU(2). A preliminary ground state is obtained for some specific chemical potential μ_0, which serves as the initialization of the optimization for all other values of μ.

The symmetry bookkeeping of the tensors in our algorithm is managed via the QSpace tensor library [66,67,83]. The QSpace library implements a series of tensor operations (e.g. contraction, SVD, etc.) that handle the propagation of symmetries. Therefore, we are able to use identical codes for both our U(1) and SU(2) iPEPS, differing only in the initialization where symmetries are designated. This guarantees the equal reliability of our iPEPS implementation with different symmetries.

The accurate contraction and measurement of the observables are performed via the Corner Transfer Matrix Renormalization Group (CTMRG) scheme [72,79,81], which generates a series

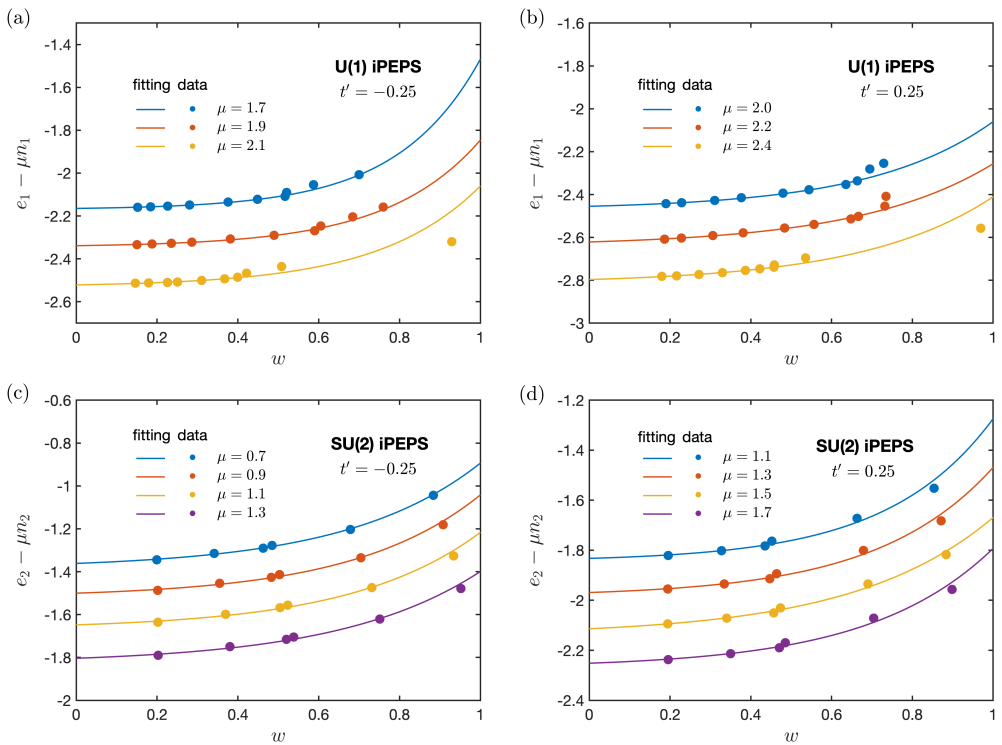


FIG. S4. The scaling behaviors of several typical U(1) (a,b) and SU(2) (c,d) symmetric ground states with respect to the iPEPS truncation error w defined as Eq. (S1) following [76]. The energies and charge densities are measured while reducing the bond dimension from $D = 12$ to 3 for U(1) and from $D^* = 7$ to 2 for SU(2). Since the charge density varies along with the bond dimension, we display the total energy $e_{1,2} - \mu n_{1,2}$ at a fixed chemical potential. Solid lines indicate exponential fits. The scaling behavior demonstrates a sufficient convergence at $D = 12$ and $D^* = 7$.

of environmental tensors. The convergence of the energies per site $e_1 - \mu n_1$ and $e_2 - \mu n_2$ with respect to environmental bond dimension χ at fixed $D = 12$ for U(1) iPEPS or χ^* at fixed $D^* = 7$ for SU(2) iPEPS is shown in Fig. S3. Our findings indicate adequate convergence commencing from $\chi = 80$ and $\chi^* = 50$. To ensure robustness, we set $\chi = 144$ and $\chi^* = 100$ for the results presented in the main text.

S-III. SCALING AND ERROR ANALYSIS

Fig. S4 illustrates the scaling behavior of various typical U(1) and SU(2) ground states. As already pointed out in Ref. [76], the evolution with increasing bond dimensions does not showcase a sufficiently regular behavior to allow a meaningful extrapolation. Consequently, following [76], we opt to present the scaling as a function of the iPEPS truncation error w defined as

$$w = \lim_{\tau \to 0} \mathrm{norm}(|\psi'\rangle - |\Psi\rangle)/\tau, \tag{S1}$$

where $|\Psi\rangle = \exp\{-\tau(\mathcal{H} - \mu N)\}|\psi\rangle$. Here, $|\psi\rangle$ represents the original iPEPS ground state, with $|\Psi\rangle$ and $|\psi'\rangle$ denoting the time-evolved iPEPS prior to and following the truncation. Thus, w serves as a metric for the truncation error and decreases monotonically with increasing D and D^*. This quantity yields a good scaling behavior and furnishes a reliable estimation of the observables in the limit of $w \to 0$ (corresponding to $D, D^* \to \infty$).

The energies and charge densities are measured while reducing the bond dimension from $D = 12$ to $D = 3$ for U(1) and from $D^* = 7$ to $D^* = 2$ for SU(2) (analogous to the knowledge distillation process in the compression of machine learning models). Given the context of an infinite lattice, there is no algorithmic constraint on the charge density. Therefore, the charge density can vary in concert with the bond dimension, and the energies e_1 and e_2 do not display a predictive scaling behavior. Conversely, the total energy $e_{1,2} - \mu n_{1,2}$, which combines the energy and the charge density at a fixed μ, performs much better, as evidenced in Fig. S4.

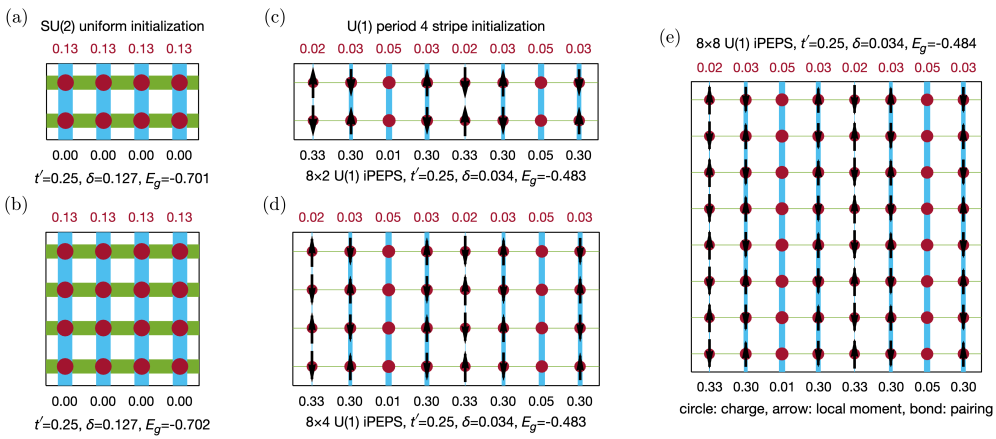


FIG. S5. (a) The 4×2 initialization of the SU(2) iPEPS ground state. (b) The 4×4 SU(2) iPEPS ground state obtained by a continued optimization from a perturbed version of (a). (c) The 8×2 initialization of the U(1) iPEPS ground state. (d,e) The 8×4 (d) and 8×8 (e) U(1) iPEPS ground state derived from a continued optimization via a perturbed initialization (c). The iPEPS simulations at larger supercell sizes in the vertical direction converge back to the initial state, which verifies the consistency of the simulation.

At a specified bond dimension D or D^*, the error of an iPEPS simulation is attributed solely to the truncation in the CTMRG process. As the environmental bond dimension increases, the variation of key observables, e.g. the energy, the charge density, the singlet pairing, etc., lies around 10^{-4}. This serves as an error estimate for these observables. Given that this error magnitude falls below the marker sizes of our figures, we omitted explicit error bars for clarity.

Although the total energy $e_{1,2} - \mu n_{1,2}$ scales well with the truncation error w, the individual observables, such as the energy e_1 and e_2, the charge density and the singlet pairing, do not. This poses challenges in formulating plausible extrapolation towards the limit of $w \to 0$ for a fixed doping. Additionally, the scarcity of data points across varying bond dimensions for such extrapolations further complicates the precision and reliability of extrapolated values and associated error estimates within the iPEPS tensor network framework. Hence, we elect to present data at the largest bond dimensions attainable, i.e. $D = 12$ for U(1) and $D^* = 7$ for SU(2), rather than relying on extrapolated numbers. Fig. S4 employs an exponential fits. In this manner, the discrepancy between relevant observables at the largest bond dimension and their extrapolated counterparts is of order 10^{-3} or less. We deem this level of precision adequate to underpin our conclusions.

S-IV. CONSISTENCY AT LARGER SUPERCELL SIZES

In our main analysis, we conduct iPEPS simulations on the 4×2 and 8×2 supercells with different horizontal and vertical lengths. Here, we extend our simulations to encompass larger supercell sizes and verify if other candidate ground states may emerge. The computational demands escalate significantly with the increase in supercell sizes. Therefore, we adopt a strategic approach to reconcile the challenges.

The essence of our approach is to initiate the simulation from a slightly perturbed state from the already converged ground state, and continue optimizing to scout for any other candidate ground states. Specifically, for simulations on a 4×4 supercell, we populate this supercell with two replicas of 4×2 supercells. Subsequently, we introduce perturbations across all tensors as

$$\begin{aligned} \Gamma_i^\alpha &\to \Gamma_i^\alpha(1 + \xi p^\alpha), \\ \Lambda_i^\alpha &\to \Lambda_i^\alpha(1 + \xi p^\alpha), \end{aligned} \tag{S2}$$

where index i enumerates over all the tensors, and index α iterates over all tensor elements; p^α is assigned a random value in the range of $[-1, 1]$ different for each different α, and ξ is a scaling factor. We set ξ to 0.01 such that it is sufficiently large to explore regions adjacent to the original ground state, yet small enough to avoid a slow convergence. This perturbed state is then subjected to further optimization using the 4×4 iPEPS algorithm to enable the identification of new potential ground states. The 8×4 and 8×8 iPEPS simulations are performed from the 8×2 initializations in a similar manner.

The findings are displayed in Fig. S5. Figs. S5(a,b) show the 4×2 SU(2) initialization and the 4×4 ground states obtained, respectively. The resultant ground state is physically identical to the 4×2 initialization. Fig. S5(c) depicts the 8×2 U(1) initialization, with (d) and (e) presenting the 8×4 and 8×8 ground states after a further optimization. These larger-scale simulations, particularly in the vertical direction, consistently revert to the initial state. This outcome confirms the reliability of our simulation with limited vertical dimensions.

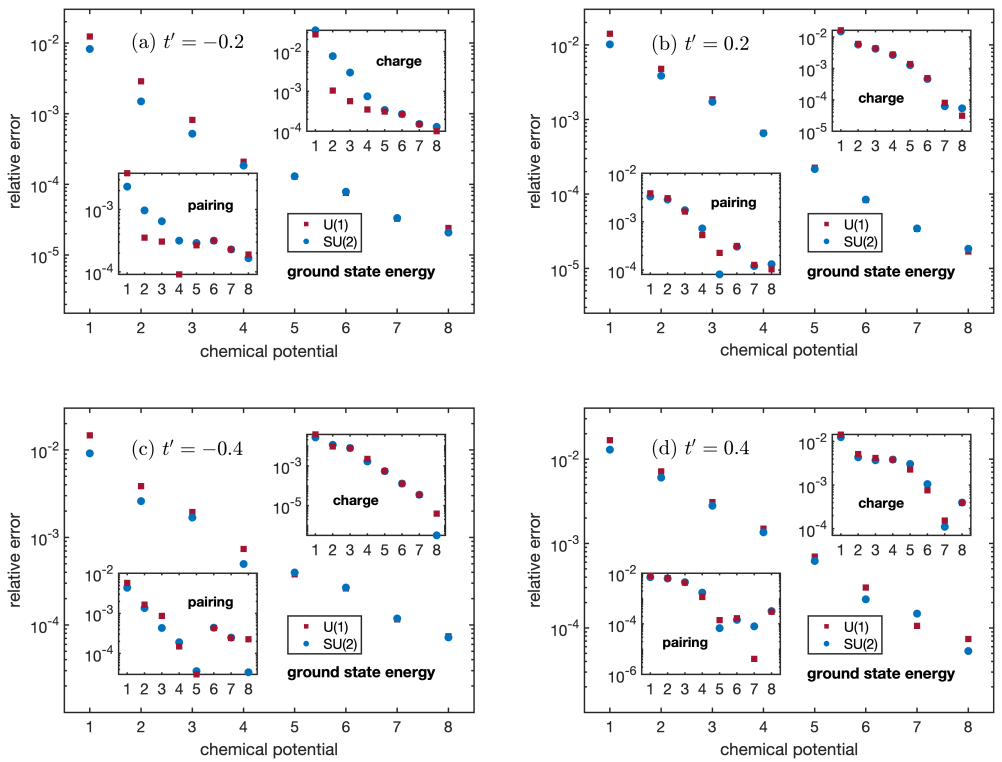


FIG. S6. The error of ground state energy e_0, charge density n_0 and singlet pairing amplitude Δ_s of the free-fermion model obtained by U(1) and SU(2) iPEPS algorithm on a 2×2 supercell at $D = 9$ and $D^*[D] = 6$[9], respectively. We set $\gamma = 1$ and focus on four values of $t' = \pm 0.2, \pm 0.4$. At each t', we vary chemical potential from $\mu = 1$ to 8. The accuracy at large chemical potential is generally better than that at small chemical potential, which is consistent with the fact that the energy gap grows with the chemical potential in this model.

S-V. BENCHMARK

In this section, we present benchmarks of our U(1) and SU(2) iPEPS algorithms for an exactly solvable free-fermion model. The Hamiltonian reads:

$$\begin{aligned}\mathcal{H} = &-t \sum_{\langle i,j\rangle,\sigma} \left[c^\dagger_{i\sigma} c_{j\sigma} + \gamma \Delta_{ij} + \text{h.c.}\right] \\ &- t' \sum_{\langle\langle i,j\rangle\rangle,\sigma} \left[c^\dagger_{i\sigma} c_{j\sigma} + \gamma \Delta_{ij} + \text{h.c.}\right] + \mu \sum_{i,\sigma} c^\dagger_{i\sigma} c_{i\sigma},\end{aligned} \tag{S3}$$

where

$$\Delta_{ij} = c_{i\uparrow} c_{j\downarrow} - c_{i\downarrow} c_{j\uparrow}, \tag{S4}$$

t and t' are the nearest and next-nearest neighbor hopping amplitude, γ the singlet pairing potential, and μ the chemical potential. The ground state energy can be computed via a Fourier transform followed by a Bogoliubov transform:

$$\begin{aligned} e_0 &= \sum_{\mathbf{k}} \left[\xi_{\mathbf{k}} - \sqrt{\xi^2_{\mathbf{k}} + \Delta^2_{\mathbf{k}}}\right], \\ \xi_{\mathbf{k}} &= \mu - 2t(\cos \mathbf{k}\cdot\mathbf{1}_{\text{x}} + \cos \mathbf{k}\cdot\mathbf{1}_{\text{y}}) \\ &\quad - 2t'(\cos \mathbf{k}\cdot\mathbf{1}_{\text{d}} + \cos \mathbf{k}\cdot\mathbf{1}_{\text{o}}), \\ \Delta_{\mathbf{k}} &= 2\gamma t(\cos \mathbf{k}\cdot\mathbf{1}_{\text{x}} + \cos \mathbf{k}\cdot\mathbf{1}_{\text{y}}) \\ &\quad + 2\gamma t'(\cos \mathbf{k}\cdot\mathbf{1}_{\text{d}} + \cos \mathbf{k}\cdot\mathbf{1}_{\text{o}}), \end{aligned} \tag{S5}$$

and $\mathbf{1}_{\text{x}} = [1, 0]$, $\mathbf{1}_{\text{y}} = [0, 1]$, $\mathbf{1}_{\text{d}} = [1, 1]$, $\mathbf{1}_{\text{o}} = [1, -1]$ are the directional vectors of the four hopping directions, respectively. In all our benchmarks, we set $t = 1$ and $\gamma = 1$. And the analytical numerics are performed on a 100×100 finite size lattice, which suffices to approximate the thermodynamic limit with a double precision accuracy.

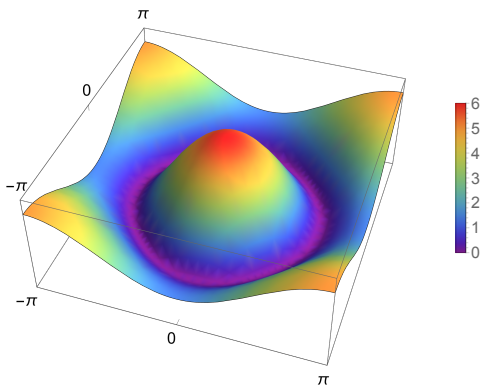


FIG. S7. The dispersion relation for the free-fermion model defined according to Eq. (S3) at $t = 1$, $t' = 0.2$, $\gamma = 1$, and $\mu = 1$. This system demonstrates a gapless nature when $\mu = 0$, and develops an energy gap for $\mu > 0$. The dispersion relation exhibits a characteristic *Mexican hat* shape for $0 < \mu \leqslant 8$. In this range, the energy gap is quantified by the minimal value around the depicted purple region.

In the case of $\gamma = 1$, the dispersion relation $E_{\mathbf{k}} = \sqrt{\xi_{\mathbf{k}}^2 + \Delta_{\mathbf{k}}^2}$ has a simpler form

$$E_{\mathbf{k}} = \sqrt{(\mu - \Delta_{\mathbf{k}})^2 + \Delta_{\mathbf{k}}^2}, \tag{S6}$$

which exhibits a *Mexican hat* shape, as depicted in Fig. S7. The minimum of the dispersion relation satisfies $\partial_{\mathbf{k}} E_{\mathbf{k}} = 0$. For $0 < \mu \leqslant 8$, this condition leads to $\mu = 2\Delta_{\mathbf{k}}$, which yields an energy gap

$$\Delta_E = \mu/\sqrt{2} \tag{S7}$$

Therefore, the energy gap grows as the chemical potential increases. Also, we can calculate analytically the average charge density n_0 and singlet pairing amplitude Δ_s:

$$\begin{aligned} n_0 &= \sum_{\mathbf{k}} \left[1 - \frac{\xi_{\mathbf{k}}}{\sqrt{\xi_{\mathbf{k}}^2 + \Delta_{\mathbf{k}}^2}} \right], \\ \Delta_s &= \sum_{\mathbf{k}} \Delta_{\mathbf{k}}^{\mathrm{nn}} \cdot \frac{\Delta_{\mathbf{k}}}{2\sqrt{\xi_{\mathbf{k}}^2 + \Delta_{\mathbf{k}}^2}}, \end{aligned} \tag{S8}$$

where

$$\Delta_{\mathbf{k}}^{\mathrm{nn}} = 2\gamma t(\cos \mathbf{k}\cdot\mathbf{1}_{\mathrm{x}} + \cos \mathbf{k}\cdot\mathbf{1}_{\mathrm{y}}) \tag{S9}$$

Benchmark results for the U(1) and SU(2) iPEPS algorithms are depicted in Fig. S6, demonstrating their errors relative to the aforementioned analytical values on a 2×2 supercell with $D = 9$ and $D^*[D] = 6[9]$, respectively. The analysis incorporates four selected values of $t' = \pm 0.2, \pm 0.4$, and for each t' value, the chemical potential is varied from $\mu = 1$ to 8.

As the chemical potential increases, there is a discernible enhancement in the accuracy of both the U(1) and SU(2) iPEPS algorithms. This improvement can be attributed to the increase of the energy gap in proportion to the chemical potential, as described by Eq. (S7). An increased energy gap implies diminished long-range correlations, thereby enabling the ground state properties to be better encoded within an iPEPS ansatz at a limited bond dimension.

The SU(2) symmetry is an inherent characteristic of the ground state of this free-fermion model. Thus, the errors produced by both the U(1) and SU(2) iPEPS are comparable. However, in certain instances, the SU(2) iPEPS demonstrates a marginally superior accuracy, which can likely be attributed to the additional SU(2) symmetry constraint that aids in better approximating the true SU(2) symmetric ground state. For large chemical potential, the accuracies of both algorithms reach 10^{-4}, corroborating the efficacy of the algorithms.

S-VI. GUIDED AND UNGUIDED U(1) IPEPS SIMULATIONS

As a mathematical construction, the U(1) iPEPS ansatz encompasses a larger parameter space compared to the SU(2) iPEPS ansatz at the same virtual bond dimensions, since the former is subject to fewer symmetry constraints. Consequently, it might be simplistically perceived that the U(1) iPEPS should be capable of capturing SU(2) symmetric states for the parameter regime where the ground states retain SU(2) symmetry. Yet, from a pragmatic perspective, the U(1) iPEPS rarely converges to an SU(2) symmetric state. This section aims to provide theoretical insights into this artifact.

D^*	D	$S[R]$
1	1	$0[1]$
2	3	$0[1] \oplus \frac{1}{2}[1]$
3	4	$0[2] \oplus \frac{1}{2}[1]$
4	6	$0[2] \oplus \frac{1}{2}[2]$
5	7	$0[3] \oplus \frac{1}{2}[2]$
6	9	$0[3] \oplus \frac{1}{2}[3]$
7	12	$0[3] \oplus \frac{1}{2}[3] \oplus 1[1]$
8	13	$0[4] \oplus \frac{1}{2}[3] \oplus 1[1]$
⋮	⋮	⋮

TABLE I. The D^*-D correspondence and the details of the symmetry sectors. S is the total spin quantum number and R the number of spin-S multiplets, each of dimension $2S + 1$. For example, for $D^* = 5$, there are 5 multiplets, namely 3 singlets and 2 doublets, leading to a total bond dimension of $D = 7\,(= 3\times1+2\times2)$.

During the optimization of tensor networks, specific strategies are usually implemented to foster a more efficient approximation of the ground state. Concretely, the usual approach com-

mences with a small bond dimension D, e.g. $D=2$, followed by optimization to convergence, prior to incrementing D until the pre-determined maximum D is reached. The primary rationale behind this strategy is the circumvention of potential local minimums in the expansive parameter space spanned by a large D. Nevertheless, this strategy engenders complications when searching through the SU(2) symmetric subspace.

In the course of conducting our SU(2) iPEPS simulations, we have identified a distinct relationship between the number of kept multiplets D^* and the corresponding bond dimension D. Particularly, in the context of the t-t' Hubbard model, the D^*-D correspondence and the details of the symmetry sectors are delineated in Table I.

A perusal of the table reveals that not all choices of D are congruous with the SU(2) symmetry. In other words, if one carries out an optimization at $D_u = 2, 5, 8, 10, 11, \cdots$ (the D values not in Table I), the resulting ground state is invariably destined to violate the SU(2) symmetry. Hence, should one want to preserve of SU(2) symmetry within a U(1) iPEPS algorithm, the optimization process should be conducted exclusively following the sequence $D_g = 1, 3, 4, 6, 7, 9, 12, \cdots$. We refer to this approach as *guided* U(1) optimization. Conversely, should one traverse the entire sequence $D = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, \cdots$, the SU(2) symmetry is assured to be broken. We thus refer to this method as *unguided* U(1) optimization.

Figure S8 shows the ground state energy per site obtained from both guided and unguided U(1) iPEPS on the 2×2 supercells, in conjunction with the SU(2) iPEPS on the 4×2 and 2×2 supercells. Remarkably, the 2×2 guided U(1), the 2×2 SU(2), and the 4×2 SU(2) results agree very well with each other. The overlap between 2×2 and 4×2 SU(2) iPEPS substantiates the uniformity of the SU(2) ground states. Furthermore, the alignment between 2×2 guided U(1) and 2×2 SU(2) iPEPS signifies that they are producing states that are physically analogous. This proposition is further supported by comparisons of other observables.

Conversely, the unguided U(1) iPEPS yields non-SU(2) symmetric states characterized by local magnetic orders, which exhibit a higher energy than the states produced by the other three methodologies. This suggests that an iPEPS preserving only the U(1) spin symmetry can become entrenched in specific local minimums in the absence of appropriate guidance or additional symmetry constraints.

An intuitive interpretation of this phenomenon is rooted in the observation that the configuration space of the U(1) iPEPS for the t-t' Hubbard model is non-convex. This is depicted in Fig. S9, a sketch of the energy landscape of such a configuration space. In this particular instance, two local minimums are present (though there could potentially be more, we limit our focus to the two most relevant local minimums): one resides within the SU(2) symmetric subspace; the other stays outside. It becomes evident that when the algorithm adheres to the guided sequence D_g, it remains within the SU(2) symmetric subspace (blue arrow). However, in the absence of such a guidance, the algorithm veers away from the SU(2) subspace, beginning from $D = 2$ (red arrow). Hence, we have demonstrated that guided and unguided U(1) iPEPS can lead to markedly different results.

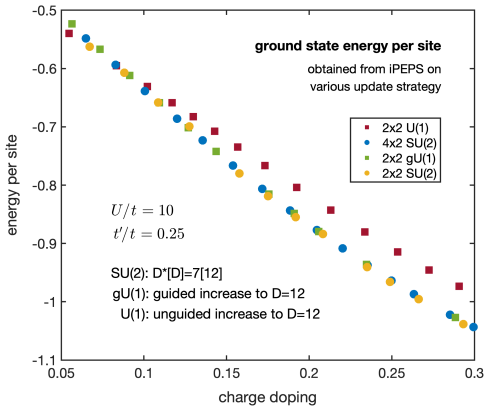


FIG. S8. Ground state energy per site obtained from guided and unguided U(1) iPEPS on the 2×2 supercells, as well as SU(2) iPEPS on the 4×2 and 2×2 supercells respectively as a function of doping. The data points of 2×2 guided U(1), 2×2 SU(2) and 4×2 SU(2) computations overlap with each other, confirming the uniformity of the SU(2) ground states.

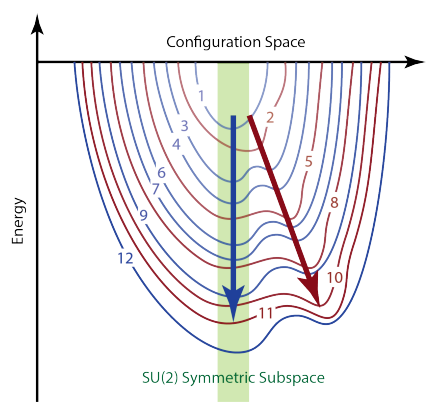


FIG. S9. Schematic free-hand drawing of the energy landscape of the U(1) iPEPS configuration space for the t-t' Hubbard model, specifically within parameter settings where ground states preserve SU(2) symmetry. Each curve depicted in the diagram represents energy at varying points in the configuration space (i.e. the space of all states that can be represented by an iPEPS ansatz), with the numbers indicating the corresponding bond dimension. The blue (red) curves correspond to bond dimensions compatible (incompatible) with the SU(2) symmetry. When optimizing U(1) iPEPS tensor networks, a guided optimization process would follow the direction indicated by the blue arrow, ultimately reaching the SU(2) symmetric ground state. Conversely, an unguided optimization process would follow the direction of the red arrow, leading to entrapment at a local minimum.

S-VII. SINGLET PAIRING PROPERTIES OF THE SU(2) GROUND STATES

Our SU(2) iPEPS simulations encounter a minor artifact at $D^*[D] = 7[12]$ when measuring the singlet pairing amplitudes. Figure S10 displays a 1×2 subcell as a representative of each 4×2 supercell, since the states are uniform in the horizontal direction. For the particular setting of $D^*[D] = 7[12]$, the vertical bond in the middle exhibits a different symmetry structure than the other bonds. This discrepancy significantly impacts the measured singlet pairing properties, as they deviate markedly from the expected d-wave characteristics.

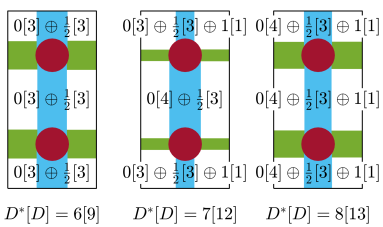


FIG. S10. Singlet pairing properties of the SU(2) ground states at different bond dimensions with $\delta \approx 0.3$ and $t'/t = 0.25$. A 1×2 subcell of each 4×2 supercell is displayed (since the states are uniform in the horizontal direction). At $D^*[D] = 6[9]$ and $D^*[D] = 8[13]$, all horizontal and vertical bonds host the same symmetry sectors. By contrast, at $D^*[D] = 7[12]$, the central vertical bond has one more singlet and one less triplet compared to the other bonds. This results in an enhanced singlet pairing amplitude in the central vertical bonds. Increasing D^* to $D^*[D] = 8[13]$ rectifies this anomaly.

We have confirmed that this anomaly arises exclusively at $D^*[D] = 7[12]$. At adjacent $D^*[D] = 6[9]$ and $D^*[D] = 8[13]$ (as well as all smaller D^*), the symmetry structure remains consistent across all vertical and horizontal bonds, and the singlet pairing amplitudes demonstrate excellent d-wave characteristics, as illustrated in Fig. S10. Thus, we conclude that the ground state should exhibit a d-wave pairing order. In the main text, we present the pairing properties as derived from $D^*[D] = 8[13]$.

S-VIII. STRIPES WITH LONGER PERIODS

In the main text, our focus is predominantly on the period 4 stripe states. Nevertheless, it is important to note that stripes with longer periods might emerge under varying parameter configurations. This section is dedicated to a further exploration of this aspect. As the manifestation of stripes with longer periods necessitates iPEPS on larger supercells, the computational cost can become significantly higher. Consequently, our objective here is not a complete optimization of the iPEPS across all scanning points. Rather, our aim is to estimate the region where the ground state begins to deviate from the charge uniform state.

We initiate the iPEPS simulation with a uniform state, obtained through guided U(1) iPEPS at large doping. Consider, for instance, a period 8 stripe. A 16×2 iPEPS is necessary to capture stripes characterized by period 8 charge orders. We acquire a single ground state via the guided 2×2 U(1) iPEPS at large doping, typically around $\delta \approx 0.3$. Its 2×2 supercell is then replicated 8 times to construct a 16×2 supercell. The corresponding ground state subsequently serves as the initialization for the further 16×2 optimization at lower doping levels.

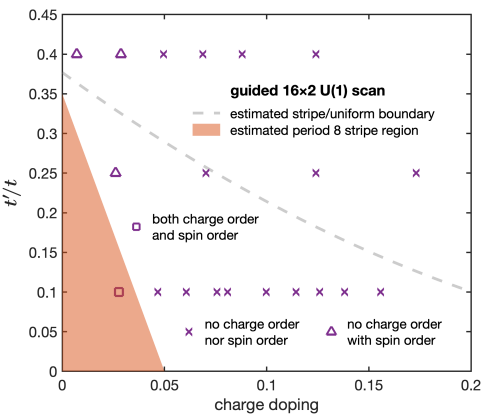


FIG. S11. 16×2 guided U(1) scan at $t'/t = 0.1, 0.25, 0.4$, respectively. Cross markers represent the 16×2 U(1) iPEPS simulations maintaining uniform states devoid of charge and spin orders. Triangle markers denote the 16×2 U(1) iPEPS simulations sustaining uniform in charge orders but displaying spin orders. For even lower doping, charge orders start to emerge, indicated by an orange shaded region.

The outcome of this process is illustrated in Fig. S11. At the majority of scan points where $\delta > 0.05$, the ground state derived from further 16×2 U(1) iPEPS optimization remains physically the same as the uniform state produced by guided 2×2 U(1) iPEPS (cross markers). As the doping level decreases, spin orders begin to appear (triangle markers), followed by the charge orders (square marker). The pairing order is suppressed once the spin orders are fully developed. The estimated region for period 8 stripes is indicated as an orange shaded area in Fig. S11. This region clearly resides beneath the estimated stripe/uniform boundary as derived in the main text. Consequently, this boundary can be confidently considered as a non-superconducting/superconducting demarcation for $\delta > 0.05$.

S-IX. NEAREST-NEIGHBOR VS. ONSITE CONTRIBUTIONS

The NN hopping and on-site terms are the minimal components of the Hubbard model. Our primary objective was to assess the impact of the additional NNN hopping terms. Therefore, Fig. 2(a) of the main text shows only the sum of NN hopping and on-site contributions as functions of charge doping.

For completeness, Fig. S12 shows these two contributions separately. Our findings indicate that: (a) the NN contribution decreases faster for the $\mathrm{SU}(2)$ states compared to the $\mathrm{U}(1)$ states, while (b) the on-site contribution decreases faster for the $\mathrm{U}(1)$ states than for the $\mathrm{SU}(2)$ states. Consequently, the sum of the NN and on-site contributions exhibits a roughly equivalent rate of decrease for both $\mathrm{SU}(2)$ and $\mathrm{U}(1)$ states. This relationship is detailed in Fig. 2(a) of the main text.

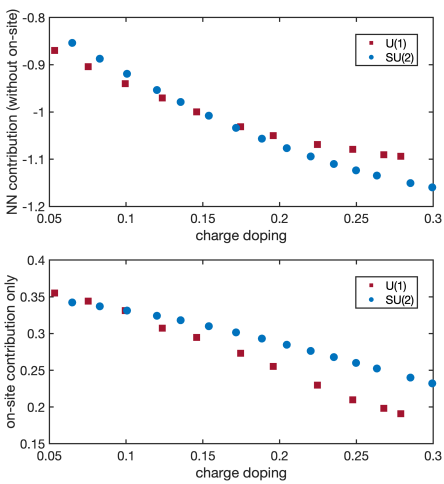


FIG. S12. The NN hopping and on-site contributions to the ground state energy at $t'/t = 0.25$ displayed separately. The NN hopping contribution decreases faster for the $\mathrm{SU}(2)$ states compared to the $\mathrm{U}(1)$ states, while the on-site contribution decreases faster for the $\mathrm{U}(1)$ states than for the $\mathrm{SU}(2)$ states.

S-X. STRATEGY FOR THE OPTIMIZATION METHOD

This research adopts a time-evolution-based simple update (SU) as the optimization method for iPEPS tensor networks. There are numerous alternative optimization methods available, such as the variation-based automatic differentiation (AD) [77], which have been demonstrated to provide greater accuracy. This section justifies the selection of this superficially less accurate simple update method by contrasting it with automatic differentiation as a representative of more precise algorithms.

Accuracy is undeniably a critical attribute of any optimization algorithm. However, computational complexity is another essential factor that influences the choice of optimization strategy. Automatic differentiation is recognized as a more accurate algorithm when compared to the simple update, but this assertion is valid only under the condition of equivalent bond dimensions and incurs significantly higher computational complexity. In practical settings, where computational resources are limited, the use of automatic differentiation would necessitate a reduced bond dimension D. Consequently, a pragmatic comparison involves evaluating the accuracy of the simple update at a larger D against automatic differentiation at a smaller D.

Figure S13 illustrates the practical evaluation of automatic differentiation's performance across various bond dimensions for $\mathrm{U}(1)$ and $\mathrm{SU}(2)$ iPEPS, respectively. The AD optimization, based on the L-BFGS algorithm and initiated from an SU-optimized ground state, was conducted for more than 10 iterations until achieving satisfactory convergence. The results indicate that for a fixed bond dimension D, automatic differentiation yields marginally lower energies for $\mathrm{U}(1)$ iPEPS ground states. However, the maximum bond dimension attainable by AD optimization is $D = 7$, and the lowest energy achieved at this level is still significantly higher than the optimal result obtained through the simple update method, which can accommodate a bond dimension of $D = 12$. Furthermore, the reduction in energy achieved through a continued AD optimization is generally less significant than that obtained by an incrementation of the bond dimension. This pattern is similarly observed in the case of $\mathrm{SU}(2)$ iPEPS, where the marginal accuracy benefits of automatic differentiation do not surpass the advantages conferred by a larger bond dimension. Thus, at the current stage, maximizing the bond dimension remains the most effective strategy for achieving lower ground state energies.

Additionally, our current AD optimization, based upon 1-site [77] L-BFGS algorithm, primarily lowers the variational energy without modifying the symmetry structure of the tensors. In parameter regimes where the ground states exhibit $\mathrm{SU}(2)$ symmetry, the AD algorithm is incapable of reinstating the $\mathrm{SU}(2)$ symmetry for a $\mathrm{U}(1)$ iPEPS in the absence of the appropriate guidance outlined in Section S-VI. Consequently, employing the AD algorithm does not yield any physics distinct from those obtained through the simple update method.

In conclusion, although the simple update method may not be the most competitive optimization algorithm in terms of accuracy, its high efficiency enables the attainment of significantly larger bond dimensions compared to other, more accurate alternatives. The ability to access a larger bond dimension not only facilitates exploration of a more extensive variational space but also yields considerably lower ground state energies. Therefore, given the current computational constraints, the simple update algorithm remains the most advantageous optimization method for the purpose of this research.

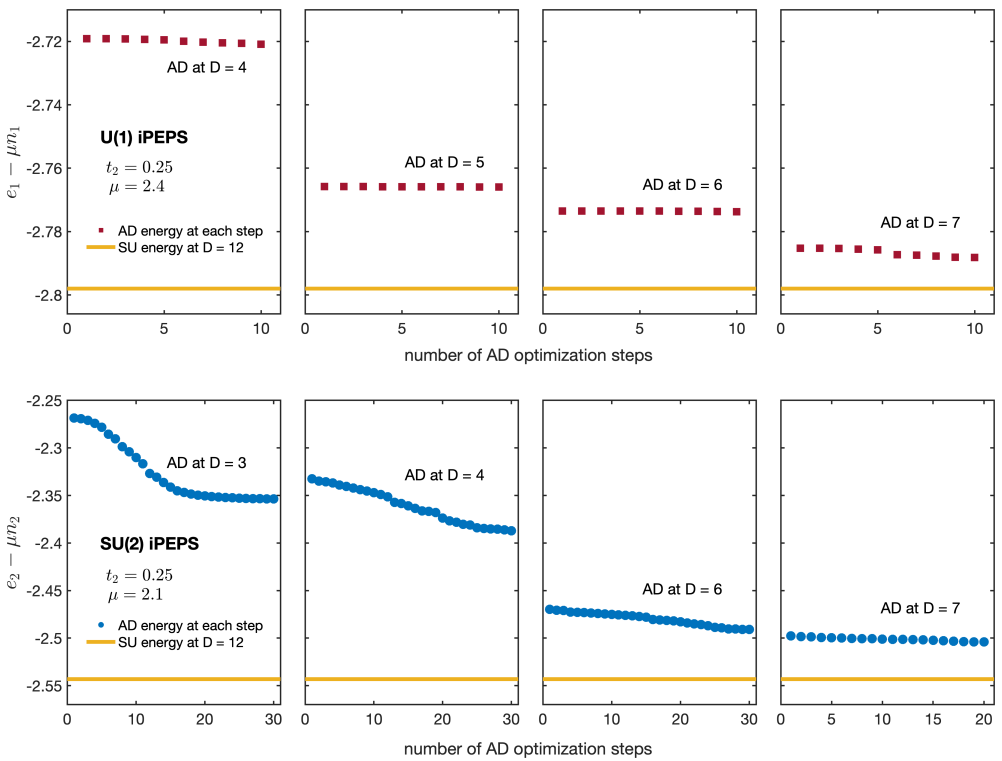


FIG. S13. Comparison of total energy $e_{1,2} - \mu n_{1,2}$ from automatic differentiation (AD) (red or blue symbols) optimization at various bond dimensions with the optimal result from the simple update (SU) method at $D = 12$ (yellow line), computed while enforcing $\mathrm{U}(1)$ (upper row) or $\mathrm{SU}(2)$ (lower row) symmetry. Each AD optimization was initialized using an SU-optimized state at the respective bond dimension. While AD optimization yields lower energies at equivalent bond dimensions, it does not surpass the performance of the SU method when computational resources are equalized. This is attributed to SU's superior efficiency, which facilitates a substantially larger bond dimensions.

3.2 Finite-temperature Phase Diagram

The long-standing challenge in connecting Hubbard physics to cuprate phenomenology is to chart, at *finite temperature*, how competing tendencies — AFM order, pseudogap behavior, and SC correlations — emerge and fade with temperature and doping. Direct numerical access to this regime is notoriously difficult: sign problems cripple standard Monte Carlo at finite doping, while many TNS studies have focused on ground states. This paper advances the program by developing an enhanced XTRG scheme and deploying it to the 2D Hubbard model across a broad temperature and doping range. The central conceptual contribution is to turn thermal simulation into a controlled exponential cooling process that can be pushed into the ultra-cold regime where meaningful comparisons to zero-temperature iPEPS become possible, thereby providing a continuous bridge from thermal equilibrium to ground-state physics and to ultracold atom experiments.

At the heart of the work is a refined $1s^+$ XTRG update scheme, which augments the single-site (1s) variational manifold so as to recover two-site (2s) update accuracy at essentially 1s computational cost. Technically, the innovation is to migrate the CBE ideas, previously successful for MPS-based ground-state and time-evolution algorithms, to MPO-MPO products within XTRG. Since XTRG constructs the low-temperature thermal density matrix by fast exponentiation of a high-temperature initial density matrix with compression, the quality-cost tradeoff of the MPO product becomes the algorithm's linchpin; enriching the variational subspace under fine control helps maintain accuracy while accelerating cooling. In practice, the method consistently reaches the low-temperature window where thermal states are sufficiently pure to be confronted with iPEPS ground states. This constitutes a broadly applicable upgrade to thermal TN simulations of correlated fermions.

The thermal properties extracted with $1s^+$ XTRG yield a coherent narrative for the doped Hubbard model. SC orders — quantified via nearest-neighbor singlet pair correlators — grow systematically upon lowering temperature and, at fixed temperature, are enhanced for *positive* NNN hopping t'/t. This mirrors ground-state trends reported in the previous iPEPS and cylinder DMRG calculations, while being demonstrated directly in the thermal ensemble. This further underscores the inconsistency of the DFT-proposed t'/t settings for hole-doped cuprates with the numerically observed pairing propensities, underscoring further investigations into the underlying pairing mechanism. In addition, the spin susceptibility delineates a pseudogap regime whose onset temperature falls with doping, in qualitative accord with cuprate phenomenology; intriguingly, a narrow underdoped window exhibits a low-temperature upturn of the uniform susceptibility, signifying the emergence of a Nagaoka polaron.

Beyond methodological and physical insights, the study facilitates a curated ensemble of *snapshots* (analogous to what is obtained via quantum state tomography) covering underdoped to overdoped regimes and spanning high, intermediate, and low temperatures. The dataset is well-suited for AI-driven analyses of patterns and cor-

relations in the minimal Hubbard model and for ultracold atom cross-checks, where site-resolved quantum gas microscopy produces comparable snapshots in optical-lattice emulators. In this sense, this technology supplies a pivotal benchmark basis that enables subsequent AI workflows and experiment-theory comparisons.

Finite-Temperature Study of the Hubbard Model via Enhanced Exponential Tensor Renormalization Group

Changkai Zhang (张昌凯) and Jan von Delft
Arnold Sommerfeld Center for Theoretical Physics, Center for NanoScience, and Munich Center for Quantum Science and Technology, Ludwig-Maximilians-Universität München, 80333 Munich, Germany
(Dated: October 31, 2025)

The two-dimensional (2D) Hubbard model has long attracted interest for its rich phase diagram and its relevance to high-T_c superconductivity. However, reliable finite-temperature studies remain challenging due to the exponential complexity of many-body interactions. Here, we introduce an enhanced 1s$^+$ eXponential Tensor Renormalization Group algorithm that enables efficient finite-temperature simulations of the 2D Hubbard model. By exploring an expanded space, our approach achieves two-site update accuracy at the computational cost of a one-site update, and delivers up to 50% acceleration for Hubbard-like systems, which enables simulations down to $T \approx 0.004t$. This advance permits a direct investigation of superconducting order over a wide temperature range and facilitates a comparison with zero-temperature infinite Projected Entangled Pair State simulations. Finally, we compile a comprehensive dataset of snapshots spanning the relevant region of the phase diagram, providing a valuable reference for Artificial Intelligence-driven analyses of the Hubbard model and a comparison with cold-atom experiments.

I. INTRODUCTION

High-T_c superconductivity in cuprate materials has attracted immense research interest ever since its discovery. [1–5] The two-dimensional (2D) Hubbard model [6, 7] on a square lattice is widely believed to capture the essential physics of these systems, particularly the effects of strong electron correlations. In recent years, extensive researches have been carried out to explore the ground-state (zero-temperature) properties of the Hubbard model [8–47] or its simplified counterpart, t-J model [48–65]. In particular, long-range or quasi-long-range pairing orders, indicative of robust superconductivity, have been observed in the presence of a positive next-nearest-neighbor hopping amplitude t' [37, 44, 47, 57–59, 63–65]. This naturally raises the question of how much such superconducting behavior extends to higher temperatures.

However, finite-temperature studies of the Hubbard model remain challenging due to the limitations of traditional numerical techniques. For example, the finite-temperature Lanczos method is constrained to small system sizes [66–68], while Quantum Monte Carlo (QMC) is plagued by the notorious sign problem at finite doping [69, 70]. In this context, thermal tensor network algorithms have emerged as promising alternatives. [71–89]. In particular, the eXponential Tensor Renormalization Group (XTRG) [78] method excels in performing exponentially rapid cooling of the system, albeit at the cost of relatively high computational complexity.

In this work, we apply the Controlled Bond Expansion (CBE) technique [90–92] originally designed for Matrix Product State (MPS) methods such as Density Matrix Renormalization Group (DMRG) [91] and Time-Dependent Variational Principle [92] to products of Matrix Product Operators (MPOs), thereby introducing an enhanced 1s$^+$ XTRG algorithm. By exploring an enlarged variational space, our scheme attains 2-site update accuracy at a much lower computational complexity, delivering up to a 50% acceleration for Hubbard-like systems. This improvement enables cooling down to $T/t = 1/256$ (approximately 20 K for $t = 0.3 \sim 0.5$ eV) and facilitates quantitative comparison with zero-temperature infinite Projected Entangled Pair State (iPEPS) results [47]. Pairing correlations are found to be enhanced at positive next-nearest neighbor (NNN) hopping ratio t'/t, and the pseudogap behavior, together with a possible Nagaoka polaron, is identified through the temperature dependence of the spin susceptibility. Finally, we generate and validate a comprehensive dataset of snapshots spanning the underdoped, intermediate-doped, and overdoped regimes across high, medium, and low temperatures. This dataset offers valuable resources for future Artificial Intelligence (AI)-driven analyses [93, 94] of the Hubbard model and for comparison or calibration of cold-atom experiments [83, 94–108].

II. MODEL AND METHOD

In this paper, we consider the 2D t-t' Hubbard model on an 8×8 square lattice with periodic boundary conditions (PBC) on the y direction and open boundary conditions (OBC) on the x direction unless stated otherwise. The Hamiltonian is defined as

$$\mathcal{H} = -\sum_{i,j,\sigma} t_{ij} \left[c^\dagger_{i\sigma} c_{j\sigma} + \text{h.c.} \right] + U \sum_i n_{i\uparrow} n_{i\downarrow}. \quad (1)$$

Here, $t_{ij} = t$ or t' for nearest neighbor (NN) or NNN hopping amplitude, respectively, and zero otherwise; U measures the on-site Coulomb repulsion. Throughout this paper, we use $U/t = 10$, as established to be realistic for cuprate materials [109, 110], and set $t = 1$ for convenience.

A comparison of band structure between cuprate superconductors and the Hubbard model suggests a positive t'/t ratio for electron-doped cuprates and a negative t'/t for hole-doped cuprates [109, 110]. However, we note that this identification leads to inconsistencies between model predictions and experimental observations for crucial order parameters, as reported previously [47, 88, 111] and further corroborated in the present work (see Section VI on pair-pair correlations).

The XTRG algorithm [78] provides an efficient framework for constructing the thermal density matrix $\rho = \exp(-\beta\mathcal{H})$ of a quantum many-body system in the form of a MPO, where $\beta = 1/T$ is the inverse temperature. Specifically, an initial density matrix $\rho(\beta_0)$ is prepared at a high temperature (small $\beta_0 = 2^{-n_0}$) via the series expansion

$$\rho(\beta_0) = \mathbb{1} - \beta_0\mathcal{H} + \frac{1}{2!}\beta_0^2\mathcal{H}^2 + \cdots . \tag{2}$$

The density matrix at half the temperature, or equivalently, at inverse temperature 2β, is then obtained by squaring the density matrix at β:

$$\rho(2\beta) = \exp\{-2\beta\mathcal{H}\} = \rho(\beta)\cdot\rho(\beta). \tag{3}$$

After n iterations, one reaches an inverse temperature of $2^n\beta_0$, realizing an exponentially fast cooling of the system. The principal computational bottleneck in XTRG thus lies in the efficient MPO-MPO product.

Analogous to the DMRG, XTRG supports both 1-site and 2-site update schemes for MPO-MPO products. The 1-site update is computationally cheaper but limited by a restricted variational space. The 2-site update overcomes this limitation by exploring a larger variational manifold, at the cost of much higher computational complexity.

The Controlled Bond Expansion (CBE) technique [90–92] offers a balanced compromise: by modestly enlarging the variational space, CBE achieves near 2-site update accuracy while maintaining a cost closer to that of the 1-site update. In the following section, we demonstrate how the CBE technique can be adapted to the MPO-MPO product, leading to a 1s$^+$ scheme that significantly accelerates the XTRG algorithm.

Our simulations employ the state-of-the-art QSpace tensor library [112–115], which preserves the $\mathrm{U}(1)_{\mathrm{charge}}\times\mathrm{SU}(2)_{\mathrm{spin}}$ symmetry in the MPO representation of the thermal density matrix. This allows us to retain up to $D^* = 1500$ multiplets (approximately $D \simeq 4000$ individual states), ensuring sufficient convergence of the XTRG algorithm.

III. 1S$^+$ MPO-MPO PRODUCT

The MPO-MPO product constitutes a fundamental, and often the most computationally demanding, kernel in the XTRG algorithm. Both computing powers of $\mathcal{H}$ in the series expansion of $\rho(\beta_0)$ and performing the cooling step entail repeated MPO-MPO multiplications. A naive contraction of two MPOs with bond dimension D yields an MPO of bond dimension D^2; compressing it back to D via a straightforward singular value decomposition (SVD) incurs a cost of $\mathcal{O}(D^6)$. This overhead naturally motivates the development of variational compression schemes.

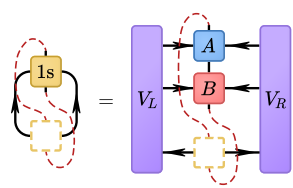
In the variational approach, one posits an ansatz for the target MPO and optimizes its proximity to the exact product. Concretely, given MPOs A and B with product $A\cdot B$, we seek an MPO C that minimizes the squared Frobenius norm of the discrepancy

$$\|C - A\cdot B\|_F^2 = C^\dagger C - C^\dagger(A\cdot B) + \text{h.c.} + \text{const.}, \tag{4}$$

where $\cdot$ denotes the MPO-MPO product. The optimization proceeds by sweeping along the chain and alternately updating one tensor at a time (1-site update, or 1s) or two adjacent tensors (2-site update, or 2s), while keeping all others fixed. The 1-site update requires solving the linear system

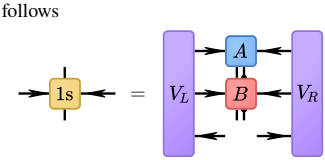
$$\frac{\partial C^\dagger C}{\partial C_i^*} = \frac{\partial C^\dagger(A\cdot B)}{\partial C_i^*} \tag{5}$$

iteratively for $i = 1, 2, \ldots, L$, where L is the length of the MPO chain. For canonicalized MPOs, the left-hand side collapses to the tensor C_i to be updated (denoted as 1s below). Diagrammatically, the 1-site update reads



(6)

To circumvent the clutter introduced by twisty lines for physical indices, we streamline the notation by re-arranging the physical indices as follows



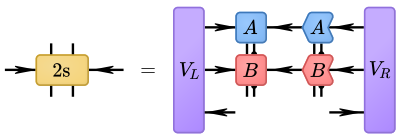
(7)

where the environment tensors V_L and V_R are pre-computed and held fixed during the 1-site update. The semicircles at the contact points between the physical index and tensor B indicate that the physical index passes through without contraction.

A prevalent flaw of the 1-site update is its inability to enlarge the variational space in the presence of symmetry constraints. As implied by Eq. (7), the updated tensor remains confined to the symmetry sector of the original tensor. This constraint can be alleviated by the 2-site update, which simultaneously optimizes two neighboring tensors, i.e.

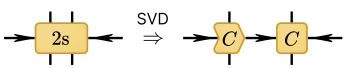
$$\frac{\partial C^\dagger C}{\partial (C_i C_{i+1})^*} = \frac{\partial C^\dagger(A\cdot B)}{\partial (C_i C_{i+1})^*} . \tag{8}$$

Diagrammatically, the 2-site update reads



(9)

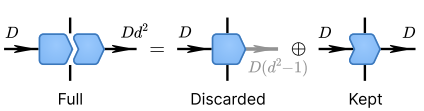
The meaning of the polygon-shaped tensor will be clarified later. The optimized 2-site tensor is then decomposed via an SVD to produce the updated MPO tensors, i.e.



(10)

Note that a full SVD in Eq. (10) generates in total Dd^2 components, where d is the physical Hilbert-space dimension, whereas only the leading D components are ultimately retained. Consequently, most components are computed only to be discarded, squandering computational resources. Our 1s$^+$ scheme mitigates this inefficiency by injecting only a small fraction of *additional components*, thereby enlarging the accessible variational space without incurring the full cost of a 2-site update.

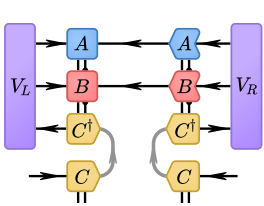
In the spirit of [90], a full MPO tensor admits a decomposition into a discarded-space component and a kept-space component



(11)

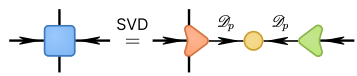
In our notation, tensors are depicted as (composite) polygons. The discarded component is denoted by a pentagon with a wedge, and the kept component by a polygon with both a wedge and a notch. Gray lines indicate components from the discarded space. The protruding face indicates the direction toward the orthogonality center of the MPO, or equivalently, the opposite direction to the normalization. In the example above, the MPO tensors are left-normalized, placing the orthogonality center to the right of the tensor. Directions of the arrows encode the flow of quantum numbers, following the convention of [114].

Similar to the CBE technique [91], the discrepancy between the variational spaces accessible to the 1-site and 2-site updates can be identified by the following construction



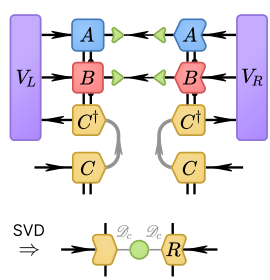
(12)

The 2-site discarded-space tensor spans the variational sector accessible to the 2-site update but excluded from the 1-site manifold. An exact construction of this tensor would incur a computational cost $\mathcal{O}(D^4d^4)$ comparable to a full 2-site update. The dominant cost resides in the contraction of the left and right part of the above 2-site tensor. Therefore, for practical purposes, we instead distill its principal components by performing the following SVD



(13)

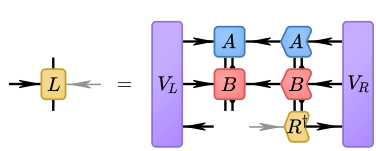
and retain only the leading $\mathscr{D}_p$ components. Here, $\mathscr{D}_p$ serves as a hyperparameter that controls the fidelity of the 2-site discarded-space approximation. In our simulations, we choose $\mathscr{D}_p = \text{round}(\sqrt{D})$ to balance accuracy against computational overhead. With this construction, the green tensor acts as a projector that compresses the bond dimension further down to $\mathscr{D}_p$. The principal components of the 2-site discarded-space tensor are then extracted by inserting these projectors as



(14)

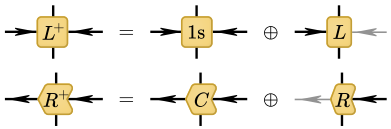
We emphasize that this procedure yields only an *approximation* to the principal components. In practice, exact identification of the principal components are unnecessary: the goal is to efficiently isolate the principal subspace that complements the 1-site update, after which subsequent optimization sweeps would converge to the desired solution.

We now outline the ensuing steps for a left-to-right optimization sweep; the right-to-left sweep proceeds analogously. The 2-site discarded-space tensor is decomposed and truncated to $\mathscr{D}_c$ via an SVD (see Eq. (14)), yielding the isometry R tailored for the left-to-right sweep. The tensor R thus serves as the supplement to the MPO tensor on the right. In our computations, we set $\mathscr{D}_c = \text{round}(\sqrt{D})$. Next, the left supplement tensor, L, is obtained by



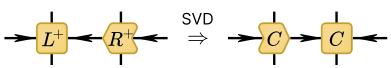
(15)

Tensors L and R thus offers the *additional components* that augment the variational space. Equipped with these ingredients, we assemble the expanded MPO tensors L^+ and R^+ as follows



(16)

Contracting L^+ and R^+ produces a tensor analogous to the 2-site object in Eq. (10), albeit with substantially fewer auxiliary components. The updated MPO tensors are then obtained by applying an SVD and truncating to the leading D components



(17)

The canonical 2-site update exhibits a computational complexity $\mathcal{O}(D^4d^4)$, where d denotes the physical dimension of the MPO tensors; this scaling is dominated by constructing the updated 2-site tensors. By contrast, the 1s$^+$ scheme reduces the complexity to the 1-site level, i.e., $\mathcal{O}(D^4d^2)$, thereby shifting the computational bottleneck to constructing the environmental tensors V_L and V_R. Consequently, the overall complexity of the 1s$^+$ scheme is $\mathcal{O}(D^4d^2 + D^4d^3)$. For Hubbard-like systems, the physical dimension $d = 4$, rendering a maximum $4\times$ speedup relative to the 2-site update. In practice, we observe speedups of up to 50% for the Hubbard model, owing to the overheads from the additional operations. The realized acceleration also depends on the specific choice of $\mathscr{D}_p$ and $\mathscr{D}_c$.

IV. BENCHMARKS

In this section, we present benchmarks of the 1s$^+$ XTRG algorithm on an analytically solvable free-fermion model. We adopt the following Hamiltonian defined on a one-dimensional open-boundary chain of length L:

$$\mathcal{H} = -t\sum_{i=1}^{L-1}\left[c_i^\dagger c_{i+1} + c_{i+1}^\dagger c_i\right]. \tag{18}$$

This Hamiltonian is exactly diagonalizable, yielding eigenvalues $\epsilon_k = -2t\cos(k\pi/(L+1))$ for $k = 1,\ldots,L$. Accordingly, the free energy can be computed as

$$F = -T\sum_{k=1}^{L}\ln(1+e^{-\beta\epsilon_k}). \tag{19}$$

In our benchmarks, we initialize the density matrix at a high temperature of $\beta = 2^{-12}$ and perform 20 1s$^+$ XTRG iterations to cool the system down to $\beta = 2^8$. Fig. 1 displays the relative error of the free energy F as a function of the inverse temperature β, computed via the 1s$^+$ XTRG algorithm with bond dimensions $D = 400, 600, 800$, respectively. The numerics indicate that the 1s$^+$ XTRG algorithm achieves accuracy comparable to the 2-site update scheme [78], while delivering a substantial speedup.

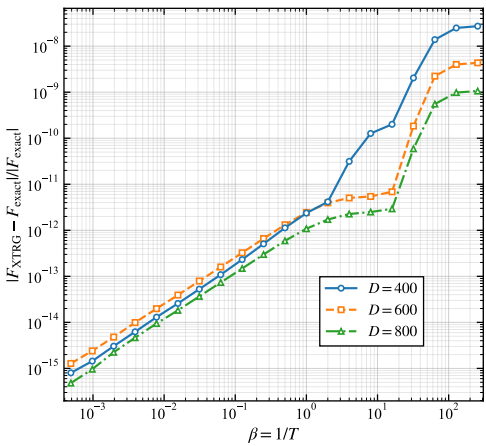


FIG. 1. The error of the free energy F of the free-fermion model relative to the exact value as a function of the inverse temperature β, obtained via the 1s$^+$ XTRG algorithm with bond dimension $D = 400, 600, 800$, respectively.

V. ENERGETICS OF THE HUBBARD MODEL

In this section, we juxtapose the energetics of the t-t' Hubbard model obtained from an iPEPS ground-state search with those from XTRG cooling. Our enhanced 1s$^+$ XTRG algorithm enables a cooling of the system down to a temperature of $T/t = 1/256$, sufficiently low to warrant a comparison with zero-temperature ground state studies.

The iPEPS ground state search is performed on an infinite lattice, achieved by PBC on both x and y directions, with a 4×2 (uniform) or 8×2 (striped) supercell, following [47]. Ground states of distinct characteristics are targeted by constraining the spin symmetry of the tensor network: imposing $\mathrm{SU}(2)_{\mathrm{spin}}$ yields a uniform state, whereas enforcing $\mathrm{U}(1)_{\mathrm{spin}}$ produces a striped state. The uniform states retain $D^* = 7$ multiplets, equivalent to $D = 12$ individual states; for consistency, the striped states are fixed to a bond dimension of $D = 12$.

The XTRG runs start from a high temperature of $\beta = 2^{-12}$ on an 8×8 lattice with PBC along the y direction, and the system is cooled down to $\beta = 2^8$ via 20 1s$^+$ XTRG iterations. Throughout, we preserve $\mathrm{U}(1)_{\mathrm{charge}} \times \mathrm{SU}(2)_{\mathrm{spin}}$ symmetry in the MPO representation of the thermal density matrix, allowing us to retain up to $D^* = 1500$ multiplets (approximately $D \simeq 4000$ individual states). We consider only the $\mathrm{SU}(2)_{\mathrm{spin}}$ symmetry here since continous symmetry breaking is precluded at finite temperature by the Mermin-Wagner theorem [116,117].

The iPEPS ground states and XTRG density matrices are defined on lattices with differing boundary conditions, which in turn modifies the count of kinetic terms in the Hamiltonian. Specifically, an 8×8 lattice with PBC along both x and y direc-

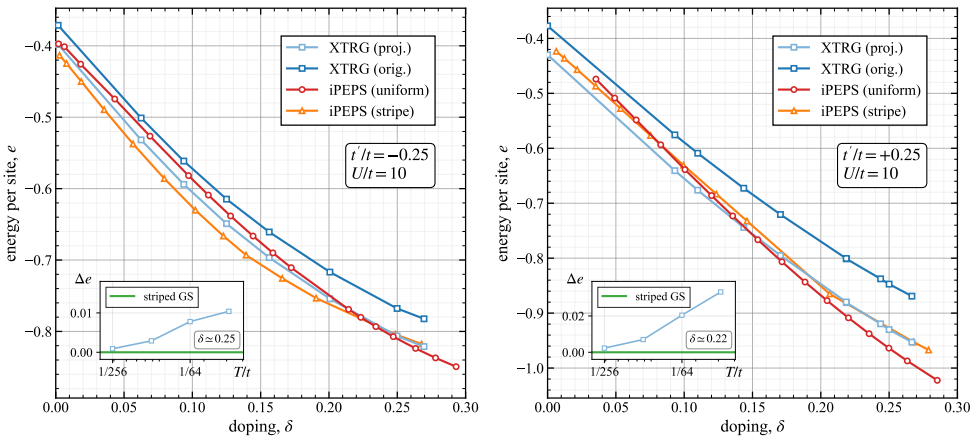


FIG. 2. The ground state energy per site (red for uniform and orange for striped states) obtained from iPEPS simulations, and the energy at $T/t = 1/256$ (blue) obtained from the $1s^+$ XTRG algorithm. The iPEPS ground states are acquired on an infinite lattice with 4×2 (uniform) or 8×2 (striped) supercell and the XTRG density matrices are generated on an 8×8 lattice with PBC on the y direction. The light blue curves mark the projected energies for the XTRG density matrices with PBC on both directions. The insets show the difference Δe between the projected XTRG energies and the striped iPEPS ground state (GS) energies with respect to temperature for two representative doping levels.

tions contains 128 nearest neighbor (NN) hopping terms and 128 next-nearest neighbor (NNN) hopping terms, whereas the same lattice with PBC only along the y direction comprises 120 NN terms and 112 NNN terms. To reconcile these discrepancies, we posit bulk-averaged energies for the missing boundary contributions. This leads to a projected (kinetic) energy with

$$\frac{e^{\text{kin}}_{\text{proj}}}{e^{\text{kin}}_{\text{orig}}} = \frac{128 \cdot e_{\text{NN}} + 128 \cdot e_{\text{NNN}}}{120 \cdot e_{\text{NN}} + 112 \cdot e_{\text{NNN}}}, \tag{20}$$

where e_{NN} and e_{NNN} denote the mean energies of the NN and NNN hopping terms, respectively. This projected energy furnishes a more commensurate reference for comparison with the iPEPS results.

Fig. 2 displays the ground state energy (red for uniform and orange for striped states) obtained from iPEPS simulations, alongside the energy at $T/t = 1/256$ (blue) derived from the $1s^+$ XTRG algorithm. The light blue curves indicate the projected energies for the XTRG density matrices with PBC along both directions.

For both $t'/t = -0.25$ and $t'/t = +0.25$, the projected XTRG energies align well with the iPEPS ground state energies, showing a similar overall doping dependence. In particular, they lie quite close to the energies of the striped states, which may be attributed to stripe formation induced by the OBC along the x direction in the XTRG simulations. This agreement indicates that our XTRG algorithm attains sufficient purity upon approaching the ultra-cold regime, i.e. the density matrix $\rho \simeq |\text{GS}\rangle\langle\text{GS}|$ at lowest temperature, where $|\text{GS}\rangle$ is the ground state.

VI. PAIR-PAIR CORRELATIONS

In this section, we investigate the dependence of pairing correlation, which signifies the strength of superconductivity, on doping, temperature and the sign of t'/t. Rigorously speaking, anomalous local pairing amplitudes are forbidden by the one-dimensionality of the tensor network together with the conserved $\text{U}(1)_{\text{charge}}$ symmetry. We therefore focus on the pair-pair correlations, which quantify the propensity for pair propagation.

Here, we primarily examine the nearest neighbor singlet pairing correlation

$$\rho_S(\mathbf{r}, \boldsymbol{\alpha}|\mathbf{s}, \boldsymbol{\beta}) = \langle \Delta^\dagger_{\mathbf{r},\mathbf{r}+\boldsymbol{\alpha}} \Delta_{\mathbf{s},\mathbf{s}+\boldsymbol{\beta}} \rangle, \tag{21}$$

where the singlet pairing operator $\Delta_{\mathbf{r},\mathbf{r}+\boldsymbol{\alpha}}$ is defined as

$$\Delta_{\mathbf{r},\mathbf{r}+\boldsymbol{\alpha}} = c_{\mathbf{r},\uparrow} c_{\mathbf{r}+\boldsymbol{\alpha},\downarrow} - c_{\mathbf{r},\downarrow} c_{\mathbf{r}+\boldsymbol{\alpha},\uparrow}, \tag{22}$$

and $\boldsymbol{\alpha} = \mathbf{x}, \mathbf{y}$ denotes the horizontal or vertical unit vector. Following [60,118], we zero-out the contributions whenever $\mathbf{r}$, $\mathrm{r}+\boldsymbol{\alpha}$, $\mathbf{s}$, or $\mathbf{s}+\boldsymbol{\beta}$ are not all distinct.

The correlator $\rho_S(\mathbf{r}, \boldsymbol{\alpha}|\mathbf{s}, \boldsymbol{\beta})$, viewed as a matrix with composite indices $(\mathbf{r}, \boldsymbol{\alpha})$ and $(\mathbf{s}, \boldsymbol{\beta})$, can be diagonalized, and one can take the dominant eigenvalue as a scalar indicator of the pairing strength [60,118,119].

Figure 3 presents the dominant eigenvalue of ρ_S as a function of doping δ and temperature T/t, obtained via the $1s^+$ XTRG for $t'/t = -0.25$ (left) and $t'/t = +0.25$ (right). The colored contour map shows the interpolation between discrete data points

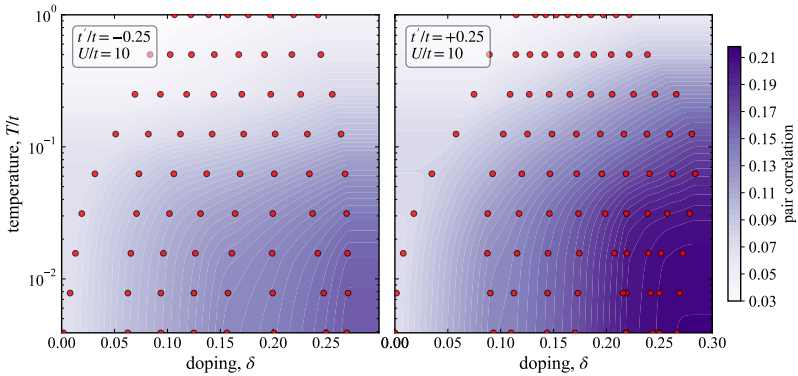


FIG. 3. The pair correlation indicator as a function of doping δ and temperature T/t, obtained from the 1s$^+$ XTRG for $t'/t = -0.25$ (left) and $t'/t = +0.25$ (right). Red circles mark the positions with exact data, and the colored contour map displays the interpolation between the data points. The pairing correlation is found strengthened at large doping, low temperature and a positive t'/t ratio.

marked by the red circles.

We observe from Fig. 3 that pairing correlations strengthen with increasing doping δ and decreasing temperature T/t, consistent with the general trends seen in iPEPS ground state simulations [47] and in cuprate phenomenology. Moreover, the correlations are systematically larger for positive t'/t, in line with various ground-state or finite-temperature studies of the t–J [57–59, 63–65, 89] and Hubbard models [37, 44, 47, 88, 111, 120]. However, if one adopts the band-structure-based identification that for cuprate superconductors $t'/t > 0$ or < 0 corresponds to electron or hole doping, respectively, our finding that pairing is stronger for $t'/t > 0$ than for $t'/t < 0$ would then be at odds with the experimental observations that pairing is weaker for electron- than hole-doped systems. This emphasizes the necessity for further investigations regarding the appropriate parameter settings or additional terms in the effective models [42, 58, 111, 121].

VII. SPIN SUSCEPTIBILITY AND PSEUDOGAP

The pseudogap — manifest as a partial depletion in the electronic density of states — is a hallmark of underdoped cuprate superconductors. Its onset suppresses the low-energy spectral weight and, concomitantly, reduces the uniform spin susceptibility χ_S. In practice, as temperature decreases, χ_S initially follows a Curie-Weiss-like increase, reaches a maximum at a characteristic scale T^*, and subsequently decreases as the pseudogap develops. The locus of this maximum thus provides a practical proxy for the pseudogap onset temperature T^*.

We compute the spin susceptibility by augmenting the Hamiltonian with a Zeeman term, $H_Z = -h \sum_i S_i^z$. The uniform susceptibility is then approximated as

$$\chi_S = -\langle S^z \rangle / h \tag{23}$$

for a sufficiently small external field $h = 0.01$, where $S^z = \sum_i S_i^z / L$. To probe the response to a magnetic field, we relax the $\mathrm{SU}(2)_{\mathrm{spin}}$ symmetry to $\mathrm{U}(1)_{\mathrm{spin}}$ in these simulations, and retain a maximum bond dimension of $D = 1500$.

Figure 4 compiles χ_S across doping δ and temperature T/t, obtained via the 1s$^+$ XTRG for $t'/t = -0.25$ (left) and $t'/t = +0.25$ (right). The colored contour shading interpolates between discrete data points as indicated by red markers. The dashed curve tracks the ridge line of the susceptibility, identifying the onset scale T^* at which the pseudogap begins to emerge.

From the figure, T^* decreases monotonically with increasing δ, with a more pronounced reduction for positive t'/t. These trends are broadly consistent with results for the t-J model [89] and with observations in various cuprate compounds [122–124]. We note that the temperature axis is logarithmic, which visually attenuates the apparent variation of T^* with doping. Besides, the overall magnitude of χ_S is smaller for positive t'/t, indicating a more enhanced pseudogap effect in this regime.

Moreover, for positive t'/t we observe an anomalous low-temperature enhancement of χ_S near $\delta \simeq 0.02$. On an 8×8 lattice, this doping approximately corresponds to a single hole introduced into the half-filled system. The upturn can therefore plausibly be attributed to the formation of a Nagaoka polaron [125–128], which engenders an underlying ferromagnetic tendency and exhibits a strong response to the external magnetic field.

VIII. SAMPLING OF THE DENSITY MATRIX

In this section, we elaborate a straight-forward sampling [129] scheme for the thermal density matrix ρ in the format of an MPO. This procedure can be used to generate an ensemble of snapshots (projections onto basis states of the Fock space) for each density

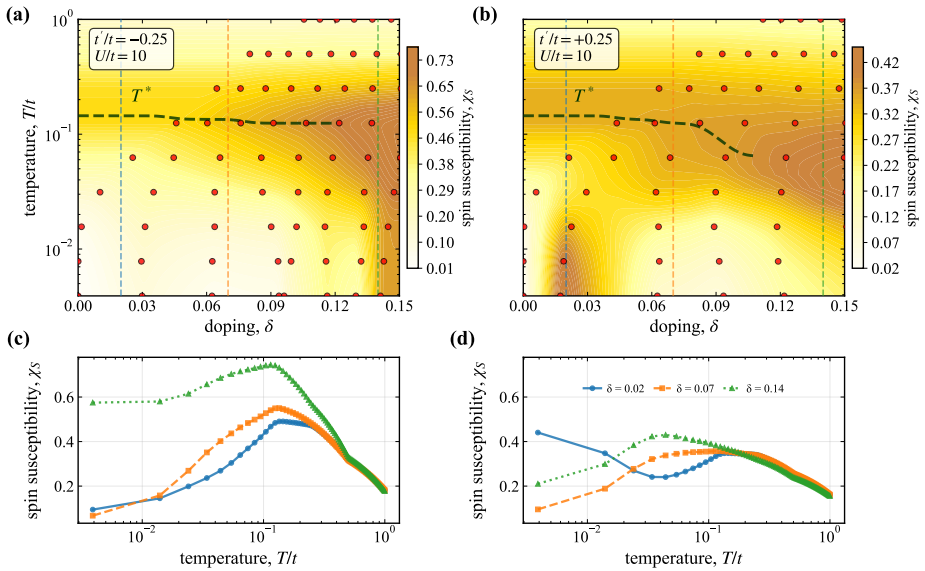
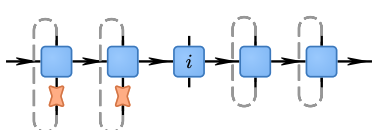


FIG. 4. The spin susceptibility (a,b) as a function of doping δ and temperature T/t, and (c,d) as a function of temperature T/t for three representative doping levels (via interpolation), obtained from the 1s$^+$ XTRG for (a,c) $t'/t = -0.25$ and (b,d) $t'/t = +0.25$. Red circles mark the positions with exact data, and the colored contour map displays the interpolation between the data points. The thick dashed line indicates the locus of the susceptibility peak, signaling the onset temperature T^* where pseudogap starts to develop.

matrix at any specified location in phase space. These ensembles can then be assembled into datasets for subsequent Artificial Intelligence (AI) analyses aimed at uncovering novel, non-trivial features of the Hubbard model.

We employ sequential sampling. Specifically, for site i, assuming all sites $\iota < i$ have already been sampled, the single-site reduced density matrix ρ_i can be constructed via the partial trace:

$\rho_i =$  (24)

where the local state projector is defined as

 $= |\sigma\rangle\langle\sigma|$ (25)

with σ denoting the sampled local state of the corresponding site. The diagonal entries of ρ_i thus furnish the probability distribution for the local state σ_i at site i. Sampling proceeds by drawing σ_i according to this distribution and then repeating the procedure for the next site $i + 1$.

In this work, we generate snapshot ensembles for the minimal Hubbard model ($t' = 0$) on an 8×8 open-boundary lattice intended for future comparisons with modern ultra-cold atom experiments [83, 97, 99, 100, 103, 105, 108]. The thermal density matrix MPOs are produced using the 1s$^+$ XTRG method across the doping range of $0 < \delta < 0.25$ and down to a lowest temperature of $T/t = 1/256$. To resolve the spin orientation, we relax the symmetry here from $\mathrm{SU}(2)_{\mathrm{spin}}$ down to $\mathrm{U}(1)_{\mathrm{spin}}$. Applying the sampling protocol described above, we draw 1000 snapshots for each combination of doping and temperature.

Figure 5 documents the convergence of the hole density n_h and double occupancy n_d as a function of sample size at $\delta \simeq 0.1694$ and $T/t = 1/16$. Green dashed lines indicate reference values extracted directly from the density matrix. We find that satisfactory convergence is achieved for $\gtrsim 200$ samples; accordingly, a sample size of 1000 is sufficient to furnish a representative ensemble for the corresponding point in the phase space.

IX. SUMMARY AND OUTLOOK

In this work, we introduced an enhanced 1s$^+$ XTRG algorithm that accelerates the cooling of the two-dimensional Hubbard model at finite temperature. By enlarging the accessible variational manifold, the method attains accuracy comparable to a two-site update while incurring computational cost near that of a one-site update, delivering speedups of up to 50% for

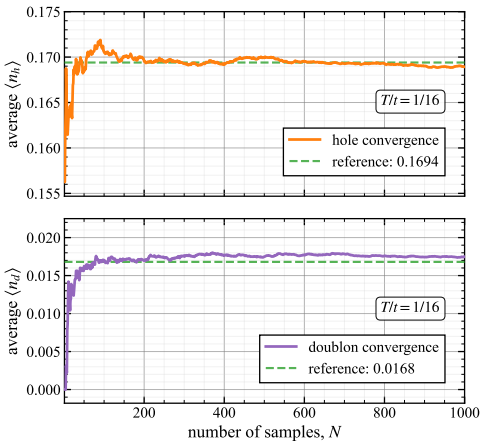


FIG. 5. Convergence behavior of hole density n_h and double occupancy n_d as a function of sample size at $\delta \simeq 0.1694$ and $T/t = 1/16$. Green dashed lines indicate reference values extracted directly from the density matrix.

Hubbard-like systems. This improvement enables cooling down to $T/t \approx 0.004$ and supports direct comparisons with zero-temperature iPEPS simulations. We demonstrate that the projected XTRG energies and pairing-correlation characteristics are in close accord with iPEPS ground-state results, indicating that our XTRG approach achieves sufficient purity upon entering the ultra-cold regime.

Our finite-temperature simulations provide a systematic characterization of singlet pairing correlations and spin susceptibilities across a broad range of dopings and temperatures. We find that pairing correlations are enhanced at large doping, low temperature, and for positive t'/t; the latter trend accords with numerous prior numerical studies yet contrasts with behaviors observed in cuprate materials. The pseudogap onset temperature T^* decreases with increasing doping for both signs of t'/t, in line with experimental observations. Finally, we identify an anomalous low-temperature enhancement of the spin susceptibility within a narrow underdoped regime, which may be attributable to the formation of a Nagaoka polaron.

In addition, we leveraged a sequential sampling scheme that generates snapshot ensembles for the minimal Hubbard model, each comprising 1000 samples and shown to be statistically representative of the corresponding point in phase space. These snapshots furnish a data resource for analyzing thermal properties and correlations, opening the door to future AI-driven investigations of the Hubbard model. A comprehensive analysis of this dataset will be presented in a forthcoming publication.

ACKNOWLEDGEMENTS

We acknowledge fruitful discussions with Annabelle Bohrdt, Yuan Gao, Ming Huang, Qiaoyi Li, Wei Li, and Dai-Wei Qu. This research was funded in part by the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy EXC-2111 (Project No. 390814868), and is part of the Munich Quantum Valley, supported by the Bavarian state government through the Hightech Agenda Bayern Plus.

[1] J. G. Bednorz and K. A. Müller, Z. Phys. B: Condens. Matter **64**, 189 (1986).
[2] B. Keimer, S. A. Kivelson, M. R. Norman, S. Uchida, and J. Zaanen, Nature **518**, 179 (2015).
[3] C. C. Tsuei and J. R. Kirtley, Rev. Mod. Phys. **72**, 969 (2000).
[4] P. A. Lee, N. Nagaosa, and X.-G. Wen, Rev. Mod. Phys. **78**, 17 (2006).
[5] C. Proust and L. Taillefer, Annu. Rev. Condens. Matter Phys. **10**, 409 (2019).
[6] J. Hubbard, Proc. R. Soc. London, Ser. A **296**, 82 (1967).
[7] D. P. Arovas, E. Berg, S. A. Kivelson, and S. Raghu, Annu. Rev. Condens. Matter Phys. **13**, 239 (2022).
[8] T. Giamarchi and C. Lhuillier, Phys. Rev. B **43**, 12943 (1991).
[9] E. Dagotto, Rev. Mod. Phys. **66**, 763 (1994).
[10] C. J. Halboth and W. Metzner, Phys. Rev. Lett. **85**, 5162 (2000).
[11] C. Huscroft, M. Jarrell, Th. Maier, S. Moukouri, and A. N. Tahvildarzadeh, Phys. Rev. Lett. **86**, 139 (2001).
[12] S. R. White and D. J. Scalapino, Phys. Rev. Lett. **91**, 136403 (2003).
[13] G. Hager, G. Wellein, E. Jeckelmann, and H. Fehske, Phys. Rev. B **71**, 075108 (2005).
[14] A. Macridin, M. Jarrell, T. Maier, P. R. C. Kent, and E. D'Azevedo, Phys. Rev. Lett. **97**, 036401 (2005).
[15] T. A. Maier, M. Jarrell, T. C. Schulthess, P. R. C. Kent, and J. B. White, Phys. Rev. Lett. **95**, 237001 (2005).
[16] M. Capone and G. Kotliar, Phys. Rev. B **74**, 054513 (2006).
[17] D. J. Scalapino, J. Supercond. Novel Magn. **19**, 195 (2006).
[18] C.-C. Chang and S. Zhang, Phys. Rev. Lett. **104**, 116402 (2010).
[19] G. Sordi, P. Sémon, K. Haule, and A.-M. S. Tremblay, Phys. Rev. Lett. **108**, 216401 (2012).
[20] H. Yokoyama, M. Ogata, Y. Tanaka, K. Kobayashi, and H. Tsuchiura, J. Phys. Soc. Jpn. **82**, 014707 (2012).
[21] E. Gull, O. Parcollet, and A. J. Millis, Phys. Rev. Lett. **110**, 216405 (2013).
[22] J. Otsuki, H. Hafermann, and A. I. Lichtenstein, Phys. Rev. B **90**, 235132 (2014).
[23] J. P. F. LeBlanc, A. E. Antipov, F. Becca, I. W. Bulik, G. K.-L. Chan, C.-M. Chung, Y. Deng, M. Ferrero, T. M. Henderson, C. A. Jiménez-Hoyos, E. Kozik, X.-W. Liu, A. J. Millis, N. V. Prokof'ev, M. Qin, G. E. Scuseria, H. Shi, B. V. Svistunov, L. F. Tocchio, I. S. Tupitsyn, S. R. White, S. Zhang, B.-X. Zheng, Z. Zhu, E. Gull, and S. C. o. t. M.-E. Problem, Phys. Rev. X **5**, 041041 (2015).

[24] Y. Deng, E. Kozik, N. V. Prokof'ev, and B. V. Svistunov, Epl-europhys Lett **110**, 57001 (2015).
[25] P. Corboz, Phys. Rev. B **93**, 045116 (2016).
[26] L. F. Tocchio, F. Becca, and S. Sorella, Phys. Rev. B **94**, 195126 (2016).
[27] B.-X. Zheng and G. K.-L. Chan, Phys. Rev. B **93**, 035126 (2016).
[28] B.-X. Zheng, C.-M. Chung, P. Corboz, G. Ehlers, M.-P. Qin, R. M. Noack, H. Shi, S. R. White, S. Zhang, and G. K.-L. Chan, Science **358**, 1155 (2017).
[29] H.-H. Zhao, K. Ido, S. Morita, and M. Imada, Phys. Rev. B **96**, 085103 (2017).
[30] A. S. Darmawan, Y. Nomura, Y. Yamaji, and M. Imada, Phys. Rev. B **98**, 205132 (2018).
[31] T. I. Vanhala and P. Törmä, Phys. Rev. B **97**, 075112 (2018).
[32] B. Ponsioen, S. S. Chung, and P. Corboz, Phys. Rev. B **100**, 195141 (2019).
[33] H.-C. Jiang and T. P. Devereaux, Science **365**, 1424 (2019).
[34] N. J. Robinson, P. D. Johnson, T. M. Rice, and A. M. Tsvelik, Rep. Prog. Phys. **82**, 126501 (2019).
[35] Y.-F. Jiang, J. Zaanen, T. P. Devereaux, and H.-C. Jiang, Phys. Rev. Res. **2**, 33073 (2020).
[36] M. Qin, C.-M. Chung, H. Shi, E. Vitali, C. Hubig, U. Schollwöck, S. R. White, S. Zhang, and S. C. o. t. M.-E. Problem, Phys. Rev. X **10**, 031016 (2020).
[37] C.-M. Chung, M. Qin, S. Zhang, U. Schollwöck, and S. R. White, Phys. Rev. B **102**, 041106 (2020).
[38] X. Dong, L. Del Re, A. Toschi, and E. Gull, Proc. Natl. Acad. Sci. **119**, e2205048119 (2022).
[39] X. Dong, E. Gull, and A. J. Millis, Nat. Phys. **18**, 1293 (2022).
[40] M. Qin, T. Schäfer, S. Andergassen, P. Corboz, and E. Gull, Annu. Rev. Condens. Matter Phys. **13**, 275 (2022).
[41] H. Xu, H. Shi, E. Vitali, M. Qin, and S. Zhang, Phys. Rev. Res. **4**, 13239 (2022).
[42] S. Jiang, D. J. Scalapino, and S. R. White, Phys. Rev. B **108**, L161111 (2023).
[43] B. Xiao, Y.-Y. He, A. Georges, and S. Zhang, Phys. Rev. X **13**, 011007 (2023).
[44] Y.-F. Jiang, T. P. Devereaux, and H.-C. Jiang, Phys. Rev. B **109**, 85121 (2024).
[45] H. Xu, C.-M. Chung, M. Qin, U. Schollwöck, S. R. White, and S. Zhang, Science **384**, eadh7691 (2024).
[46] S. d. A. Sousa-Júnior, N. C. Costa, and R. R. dos Santos, Phys. Rev. B **109**, 165102 (2024).
[47] C. Zhang, J.-W. Li, D. Nikolaidou, and J. von Delft, Phys. Rev. Lett. **134**, 116502 (2025).
[48] S. R. White and D. J. Scalapino, Phys. Rev. Lett. **80**, 1272 (1998).
[49] W.-J. Hu, F. Becca, and S. Sorella, Phys. Rev. B **85**, 081110 (2012).
[50] P. Corboz, T. M. Rice, and M. Troyer, Phys. Rev. Lett. **113**, 046402 (2014).
[51] J. F. Dodaro, H.-C. Jiang, and S. A. Kivelson, Phys. Rev. B **95**, 155116 (2017).
[52] H.-C. Jiang, Z.-Y. Weng, and S. A. Kivelson, Phys. Rev. B **98**, 140505 (2018).
[53] C. Hubig, A. Bohrdt, M. Knap, F. Grusdt, and I. Cirac, SciPost Phys. **8**, 021 (2020).
[54] H.-C. Jiang, S. Chen, and Z.-Y. Weng, Phys. Rev. B **102**, 104512 (2020).
[55] H.-C. Jiang and S. A. Kivelson, Phys. Rev. Lett. **127**, 097002 (2021).
[56] J.-W. Li, B. Bruognolo, A. Weichselbaum, and J. von Delft, Phys. Rev. B **103**, 075127 (2021).
[57] S. Gong, W. Zhu, and D. N. Sheng, Phys. Rev. Lett. **127**, 097003 (2021).
[58] S. Jiang, D. J. Scalapino, and S. R. White, Proc. Natl. Acad. Sci. U.S.A. **118**, e2109978118 (2021).
[59] S. Jiang, D. J. Scalapino, and S. R. White, Phys. Rev. B **106**, 174507 (2022).
[60] A. Wietek, Phys. Rev. Lett. **129**, 177001 (2022).
[61] H.-C. Jiang, S. A. Kivelson, and D.-H. Lee, Phys. Rev. B **108**, 54505 (2023).
[62] X. Lu, D.-W. Qu, Y. Qi, W. Li, and S.-S. Gong, Phys. Rev. B **107**, 125114 (2023).
[63] X. Lu, F. Chen, W. Zhu, D. N. Sheng, and S.-S. Gong, Phys. Rev. Lett. **132**, 66002 (2024).
[64] X. Lu, J.-X. Zhang, S.-S. Gong, D. N. Sheng, and Z.-Y. Weng, Phys. Rev. B **110**, 165127 (2024).
[65] X. Lu, H. Guo, W.-Q. Chen, D. N. Sheng, and S.-S. Gong, Phys. Rev. B **111**, 035139 (2025).
[66] J. Jaklič and P. Prelovšek, Phys. Rev. B **49**, 5065 (1994).
[67] J. Jaklič and P. Prelovšek, Phys. Rev. Lett. **77**, 892 (1996).
[68] J. Jaklič and P. Prelovšek, Adv. Phys. **49**, 1 (2000).
[69] K.-S. Chen, Z. Y. Meng, S.-X. Yang, T. Pruschke, J. Moreno, and M. Jarrell, Phys. Rev. B **88**, 245110 (2013).
[70] Y.-Y. He, M. Qin, H. Shi, Z.-Y. Lu, and S. Zhang, Phys. Rev. B **99**, 45108 (2019).
[71] S. R. White, Phys. Rev. Lett. **102**, 190601 (2009).
[72] E. M. Stoudenmire and S. R. White, New J. Phys. **12**, 55026 (2010).
[73] W. Li, S.-J. Ran, S.-S. Gong, Y. Zhao, B. Xi, F. Ye, and G. Su, Phys. Rev. Lett. **106**, 127202 (2011).
[74] P. Czarnik, L. Cincio, and J. Dziarmaga, Phys. Rev. B **86**, 245101 (2012).
[75] P. Czarnik and J. Dziarmaga, Phys. Rev. B **90**, 035144 (2014).
[76] P. Czarnik and J. Dziarmaga, Phys. Rev. B **92**, 035152 (2015).
[77] P. Czarnik, M. M. Rams, and J. Dziarmaga, Phys. Rev. B **94**, 235142 (2016).
[78] B.-B. Chen, L. Chen, Z. Chen, W. Li, and A. Weichselbaum, Phys. Rev. X **8**, 031082 (2018).
[79] L. Chen, D.-W. Qu, H. Li, B.-B. Chen, S.-S. Gong, J. von Delft, A. Weichselbaum, and W. Li, Phys. Rev. B **99**, 140404 (2019).
[80] P. Czarnik, J. Dziarmaga, and P. Corboz, Phys. Rev. B **99**, 35115 (2019).
[81] P. Czarnik and P. Corboz, Phys. Rev. B **99**, 245107 (2019).
[82] H. Li, B.-B. Chen, Z. Chen, J. von Delft, A. Weichselbaum, and W. Li, Phys. Rev. B **100**, 045110 (2019).
[83] B.-B. Chen, C. Chen, Z. Chen, J. Cui, Y. Zhai, A. Weichselbaum, J. von Delft, Z. Y. Meng, and W. Li, Phys. Rev. B **103**, L041107 (2021).
[84] A. Wietek, R. Rossi, F. Šimkovic, M. Klett, P. Hansmann, M. Ferrero, E. M. Stoudenmire, T. Schäfer, and A. Georges, Phys. Rev. X **11**, 41013 (2021).
[85] A. Wietek, Y.-Y. He, S. R. White, A. Georges, and E. M. Stoudenmire, Phys. Rev. X **11**, 31007 (2021).
[86] X. Lin, B.-B. Chen, W. Li, Z. Y. Meng, and T. Shi, Phys. Rev. Lett. **128**, 157201 (2022).
[87] A. Sinha, M. M. Rams, P. Czarnik, and J. Dziarmaga, Phys. Rev. B **106**, 195105 (2022).
[88] Q. Li, Y. Gao, Y.-Y. He, Y. Qi, B.-B. Chen, and W. Li, Phys. Rev. Lett. **130**, 226502 (2023).

[89] D.-W. Qu, Q. Li, S.-S. Gong, Y. Qi, W. Li, and G. Su, Phys. Rev. Lett. **133**, 256003 (2024).
[90] A. Gleis, J.-W. Li, and J. von Delft, Phys. Rev. B **106**, 195138 (2022).
[91] A. Gleis, J.-W. Li, and J. von Delft, Phys. Rev. Lett. **130**, 246402 (2023).
[92] J.-W. Li, A. Gleis, and J. von Delft, Phys. Rev. Lett. **133**, 26401 (2024).
[93] A. Bohrdt, C. S. Chiu, G. Ji, M. Xu, D. Greif, M. Greiner, E. Demler, F. Grusdt, and M. Knap, Nat. Phys. **15**, 921 (2019).
[94] C. S. Chiu, G. Ji, A. Bohrdt, M. Xu, M. Knap, E. Demler, F. Grusdt, M. Greiner, and D. Greif, Science **365**, 251 (2019).
[95] A. Mazurenko, C. S. Chiu, G. Ji, M. F. Parsons, M. Kanász-Nagy, R. Schmidt, F. Grusdt, E. Demler, D. Greif, and M. Greiner, Nature **545**, 462 (2017).
[96] C. S. Chiu, G. Ji, A. Mazurenko, D. Greif, and M. Greiner, Phys. Rev. Lett. **120**, 243201 (2018).
[97] J. Koepsell, J. Vijayan, P. Sompet, F. Grusdt, T. A. Hilker, E. Demler, G. Salomon, I. Bloch, and C. Gross, Nature **572**, 358 (2019).
[98] G. Salomon, J. Koepsell, J. Vijayan, T. A. Hilker, J. Nespolo, L. Pollet, I. Bloch, and C. Gross, Nature **565**, 56 (2019).
[99] J. Koepsell, S. Hirthe, D. Bourgund, P. Sompet, J. Vijayan, G. Salomon, C. Gross, and I. Bloch, Phys. Rev. Lett. **125**, 010403 (2020).
[100] J. Koepsell, D. Bourgund, P. Sompet, S. Hirthe, A. Bohrdt, Y. Wang, F. Grusdt, E. Demler, G. Salomon, C. Gross, and I. Bloch, Science **374**, 82 (2021).
[101] P. Sompet, S. Hirthe, D. Bourgund, T. Chalopin, J. Bibo, J. Koepsell, P. Bojović, R. Verresen, F. Pollmann, G. Salomon, C. Gross, T. A. Hilker, and I. Bloch, Nature **606**, 484 (2022).
[102] S. Hirthe, T. Chalopin, D. Bourgund, P. Bojović, A. Bohrdt, E. Demler, F. Grusdt, I. Bloch, and T. A. Hilker, Nature **613**, 463 (2023).
[103] M. Xu, L. H. Kendrick, A. Kale, Y. Gang, G. Ji, R. T. Scalettar, M. Lebrat, and M. Greiner, Nature **620**, 971 (2023).
[104] T. Chalopin, P. Bojović, S. Wang, T. Franz, A. Sinha, Z. Wang, D. Bourgund, J. Obermeyer, F. Grusdt, A. Bohrdt, L. Pollet, A. Wietek, A. Georges, T. Hilker, and I. Bloch (2024), arXiv:2412.17801 [cond-mat].
[105] G. Pasqualetti, O. Bettermann, N. Darkwah Oppong, E. Ibarra-García-Padilla, S. Dasgupta, R. T. Scalettar, K. R. A. Hazzard, I. Bloch, and S. Fölling, Phys. Rev. Lett. **132**, 83401 (2024).
[106] D. Bourgund, T. Chalopin, P. Bojović, H. Schlömer, S. Wang, T. Franz, S. Hirthe, A. Bohrdt, F. Grusdt, I. Bloch, and T. A. Hilker, Nature **637**, 57 (2025).
[107] T. Chalopin, P. Bojović, D. Bourgund, S. Wang, T. Franz, I. Bloch, and T. Hilker, Phys. Rev. Lett. **134**, 53402 (2025).
[108] M. Xu, L. H. Kendrick, A. Kale, Y. Gang, C. Feng, S. Zhang, A. W. Young, M. Lebrat, and M. Greiner, Nature **642**, 909 (2025).
[109] M. Hirayama, Y. Yamaji, T. Misawa, and M. Imada, Phys. Rev. B **98**, 134501 (2018).
[110] M. Hirayama, T. Misawa, T. Ohgoe, Y. Yamaji, and M. Imada, Phys. Rev. B **99**, 245155 (2019).
[111] W. O. Wang and T. P. Devereaux (2025), arXiv:2510.16616 [cond-mat].
[112] A. Weichselbaum, Ann. Phys. **327**, 2972 (2012).
[113] A. Weichselbaum, Phys. Rev. Res. **2**, 023385 (2020).
[114] A. Weichselbaum, SciPost Phys. Codebases , 40 (2024).
[115] A. Weichselbaum, Open Source QSpace tensor library v4.0, https://bitbucket.org/qspace4u.
[116] N. D. Mermin and H. Wagner, Physical Review Letters **17**, 1307 (1966).
[117] P. C. Hohenberg, Physical Review **158**, 383 (1967).
[118] N. Baldelli, H. Karlsson, B. Kloss, M. Fishman, and A. Wietek, npj Quantum Materials **10**, 22 (2025).
[119] O. Penrose and L. Onsager, Phys. Rev. **104**, 576 (1956).
[120] F. Chen, F. D. M. Haldane, and D. N. Sheng, Proceedings of the National Academy of Sciences **122**, e2420963122 (2025).
[121] T. Xiang, H. G. Luo, D. H. Lu, K. M. Shen, and Z. X. Shen, Phys. Rev. B **79**, 014524 (2009).
[122] B. Loret, S. Sakai, S. Benhabib, Y. Gallais, M. Cazayous, M. A. Méasson, R. D. Zhong, J. Schneeloch, G. D. Gu, A. Forget, D. Colson, I. Paul, M. Civelli, and A. Sacuto, Physical Review B **96**, 094525 (2017).
[123] F. Boschini, M. Zonno, E. Razzoli, R. P. Day, M. Michiardi, B. Zwartsenberg, P. Nigge, M. Schneider, E. H. da Silva Neto, A. Erb, S. Zhdanovich, A. K. Mills, G. Levy, C. Giannetti, D. J. Jones, and A. Damascelli, npj Quantum Materials **5**, 6 (2020).
[124] G. Grissonnanche, O. Cyr-Choinière, J. Day, R. Liang, D. A. Bonn, W. N. Hardy, N. Doiron-Leyraud, and L. Taillefer, Physical Review X **13**, 031010 (2023).
[125] Y. Nagaoka, Physical Review **147**, 392 (1966).
[126] H. Tasaki, Progress of Theoretical Physics **99**, 489 (1998).
[127] M. Lebrat, M. Xu, L. H. Kendrick, A. Kale, Y. Gang, P. Seetharaman, I. Morera, E. Khatami, E. Demler, and M. Greiner, Nature **629**, 317 (2024).
[128] R. Samajdar and R. N. Bhatt, Physical Review B **109**, 235128 (2024).
[129] D.-W. Qu, Q. Li, S.-S. Gong, Y. Qi, W. Li, and G. Su (2023), arXiv:2211.06322v2 [cond-mat.str-el].

3.3 Analysis via Artificial Intelligence

This work advocates a shift in how we interrogate the 2D Hubbard model and, more broadly, strongly correlated quantum matter: rather than adding yet another solver or tuning yet another Hamiltonian parameter, we propose interpretable AI workflows as analysis tools for system-level snapshot datasets produced by modern TN methods and ultracold-atom experiments. The central premise lies in the capability of attention-based architectures — now mature in their ability to capture high-order correlations — to extract latent regularities from thermal states across temperature and doping, while potentially revealing insights into what the model has learned. The effective capture of correlations by the AI model furnishes a universal *omnimeter* that can infer calibrated quantities (e.g., temperature, charge doping, etc.) for ultracold atom experiments based on the perceived correlations from quantum gas microscopy.

Central to this study is a streamlined *core* architecture designed expressly for correlation analysis and interpretability. It shares the same input encoding, output projection and multi-head attention backbone as the encoder-only transformer (referred to as a *pro* variant), but defers most non-linearity to a single feed-forward module placed after the entire attention stack and avoids interleaved layer normalizations. This leads to a semi-linear attention pipeline whose intermediate computations are simpler to interpret in physical terms (e.g., as effective mixing or propagation of learned local observables across the lattice), while training and inference benefit from enhanced parallelization. Despite its simplicity, this core model achieves performance comparable to the pro variant in the production runs when attention is fully functional. In the practical regime, the core thus achieves better computational efficiency and internal transparency without sacrificing meaningful predictive power.

The proficiency of the core AI model in perceiving correlations is validated through two analytical diagnostics. First, the confusion analysis shows that misclassifications cluster within the same doping level and systematically diminish as temperature lowers. This reduction is consistent with physics: as thermal fluctuations subside, quantum correlations sharpen and become more recognizable, so the classifier errs less since the underlying signals are intrinsically clearer. Second, in the small-scale omnimeter test, where classifier logits are mapped to a calibrated posterior over thermodynamic conditions, we observe selective failures in temperature estimates where correlation patterns were unseen in the training dataset. Also, the inferred charge doping deviates toward the phase with analogous correlation regularities. This indicates that decisions are governed by the many-body correlations learned from data, rather than by spurious cues or simple interpolations.

The performance of the omnimeter can reach a practical level once a comprehensive reference library of snapshots is established. Rather than fitting manufactured correlators for the dataset, the AI omnimeter provides posterior-based readouts of thermodynamic variables directly from microscopy snapshots. This yields a substantial efficiency gain in information retrieval compared with the traditional spin-

correlation thermometer, since the model aggregates all correlation cues present in the snapshots rather than committing to a single hand-crafted observable. Side-by-side tests illustrate that, in regimes where simple spin correlators are weak, noisy, or non-monotonic in temperature, the omnimeter maintains stable, calibrated estimates with transparent uncertainty quantification. In this sense, the discriminative model already functions as a practical metrological tool for ultracold-atom platforms, turning correlation perception into an efficient, quantitative experimental utility.

Interpretable Artificial Intelligence (AI) Analysis of Strongly Correlated Electrons

Changkai Zhang (张昌凯) and Jan von Delft
Arnold Sommerfeld Center for Theoretical Physics, Center for NanoScience, and Munich Center for Quantum Science and Technology, Ludwig-Maximilians-Universität München, 80333 Munich, Germany
(Dated: October 31, 2025)

Artificial Intelligence (AI) has become an exceptionally powerful tool for analyzing scientific data. In particular, attention-based architectures have demonstrated a remarkable capability to capture complex correlations and to furnish interpretable insights into latent, otherwise inconspicuous patterns. This progress motivates the application of AI techniques to the analysis of strongly correlated electrons, which remain notoriously challenging to study using conventional theoretical approaches. Here, we propose novel AI workflows for analyzing snapshot datasets from tensor-network simulations of the two-dimensional (2D) Hubbard model over a broad range of temperature and doping. The 2D Hubbard model is an archetypal strongly correlated system, hosting diverse intriguing phenomena including Mott insulators, anomalous metals, and high-T_c superconductivity. Our AI techniques yield fresh perspectives on the intricate quantum correlations underpinning these phenomena and facilitate universal omnimetry for ultracold-atom simulations of the corresponding strongly correlated systems.

I. INTRODUCTION

The invention of Artificial Intelligence (AI) has revolutionized the way we interrogate and interpret scientific data. The attention scheme [1–5] has been proven highly effective in transduction tasks in conjunction with recurrent or convolutional networks. Afterwards, the transformer [6] — an architecture built solely upon the attention scheme — was demonstrated to be compelling in capturing global dependencies in sequential data. Over the past decade, transformer-like architectures have dramatically enhanced the capability of the AI models across various domains, including natural language processing (NLP) [7–12], computer vision [13–16], bioinformatics [17–19], and numerous other areas [20–23]. Compared with alternative designs, the attention mechanism in the transformer excels particularly at encoding the correlation structure of the input data. This feature motivates the application of the transformer models in studying strongly correlated electrons.

Strongly correlated systems [24–27] are governed by considerably strong interactions, inducing collective behaviors that defy descriptions hinged on individual (quasi-) particles. Canonical examples include Mott insulators [28–30], high-T_c superconductors [31–34], heavy-fermion materials [35–38], fractional quantum Hall systems [39–41], spin liquids [42–45], and quark-gluon plasmas [46–49]. The consequent high levels of quantum entanglement and correlations render these systems notoriously challenging for conventional theoretical approaches. With advances in computational hardware, a handful of numerical algorithms — among them Quantum Monte Carlo (QMC) [50–53], Dynamical Mean-Field Theory (DMFT) [54–57], Density Matrix Renormalization Group (DMRG) [58–62], and various ground-state [63–70] or finite-temperature [71–79] Tensor Network (TN) methods — have been devised to tackle strongly correlated systems. Moreover, quantum simulation apparatuses based on ultra-cold atoms [80–85] have achieved substantial progress in emulating strongly interacting lattice systems. Together, these techniques offer valuable many-body data from which the AI models can learn and distill meaningful insights.

Among the plethora of strongly correlated electron systems, the two-dimensional (2D) Hubbard model [86, 87] stands out as a paradigmatic arena for a variety of intriguing phenomena, such as Mott physics, anomalous metals, and high-T_c superconductivity. The Hubbard model encapsulates the essential physics of itinerant electrons on a lattice with strong on-site Coulomb repulsion. Over the past few decades, the 2D Hubbard model has been subject to intensive investigations both numerically [52, 53, 70, 78, 79, 88–98] and experimentally [80–85, 99–108]. Robust anti-ferromagnetic (AFM) orders have been confirmed near half-filling [70, 93, 97, 109–112], while in the doped regime — especially with carrier hopping beyond neighboring sites — diverse charge and spin orders, often coexisting with or competing against pairing tendencies, have been identified [70, 78, 93–95, 113]. These properties broadly echo the observations in the cuprate superconductors.

Despite this decent progress, the majority of the existing researches focused on local and low-order spin and/or charge correlations, especially two-point correlators. However, mounting evidence indicates that high-order [108, 114, 115], non-local [116, 117], polaronic [82, 100, 118, 119] or otherwise string-like [99, 120, 121] correlations play a pivotal role in deciphering the complicated phase diagram of the 2D Hubbard model. This recognition highlights the promise of the AI techniques for the global vision of the underlying quantum correlations.

In this Article, we study the 2D Hubbard model using specifically designed AI models. We start with assembling a dedicated dataset of snapshots across categories of temperatures and doping levels by sampling the thermal density matrix via TN simulations. Then, we propose two AI architectures classifying snapshots into the respective categories: the *pro architecture*, an analog of the encoder-only transformer, and the *core architecture*, a

streamlined variant that attains comparable performance, better support for parallelism and improved interpretability.

Next, we perform multiple analyses on the trained core model. We use a confusion analysis to measure the quality of the classification tasks and obtain insights into the aggregate strength of quantum correlations in each category. Exploiting the *semi-linear* structure of the attention stack in the core architecture, we propose an interpretation in terms of an effective Markovian dynamics, demonstrating the alignment of the attention design with intrinsic features of the physical system. Further examinations on the orthogonality relationships of the embedding and the attention maps are provided in the supplemental material [122].

Finally, we demonstrate an application of our core AI models as a universal omnimeter for ultracold-atom quantum simulations. The AI classifiers produce probablistic scores (logits) for each category, which serve as posterior likelihoods conditioned on a snapshot acquired in the experiment. Averaging these outputs over a snapshot ensemble from repeated observations thus provides an empirical probability distribution over the categories. Once the categories are calibrated with pre-determined physical quantities, the expectation values weighted by the probability distribution yield an accurate estimate of the corresponding quantities for the ensemble.

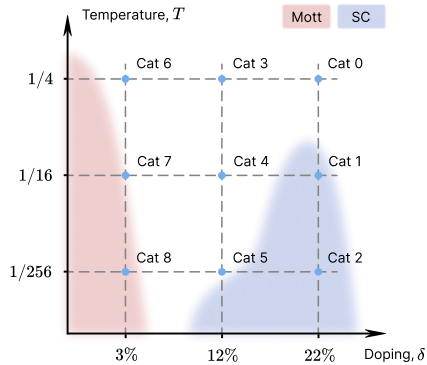


FIG. 1. A schematic depiction of the locations in phase space for the nine categories (Cat), created by combining three choices of temperatures (high, medium, and low) with three doping regimes (over-doped, medium-doped, and under-doped). The red and blue freehand-shaded areas mark the AFM Mott insulating phase and the high-T_c superconducting phase, respectively, as expected for the Hubbard model. The charge doping varies with temperature (see also Fig. 5); precise values are provided in the supplemental material [122].

II. LATTICE MODEL & DATASET

Lattice models serve as common platforms for the physics of crystalline materials, wherein charge carriers reside on and hop between discrete lattice sites. In many materials of interest, itinerant electrons predominantly occupy the outer-most s orbital for transport. Consequently, the local Hilbert space at each lattice site is spanned by four basis: empty $|\varnothing\rangle$, spin-up $|\uparrow\rangle$, spin-down $|\downarrow\rangle$, and doubly occupied $|\uparrow\downarrow\rangle$ state.

In our study, we focus on the quintessential 2D Hubbard model on an 8×8 square lattice with open boundary conditions, defined via the following Hamiltonian

$$\mathcal{H} = -\sum_{i,j,\sigma} t_{ij} \left[c^\dagger_{i\sigma} c_{j\sigma} + \text{h.c.} \right] + U \sum_i n_{i\uparrow} n_{i\downarrow}. \tag{1}$$

Here, $c^\dagger_{i\sigma}$ ($c_{i\sigma}$) creates (annihilates) an electron with spin σ on site i, and $n_{i\sigma} = c^\dagger_{i\sigma} c_{i\sigma}$ denotes the corresponding number operator. The first term in Eq. (1) describes the kinetic energy associated with electron hopping between sites i and j with amplitude t_{ij}, while the second term accounts for the on-site Coulomb repulsion with strength U. Throughout this work, we consider the minimal Hubbard model where $t_{ij} = 1$ for nearest-neighbor pairs and $t_{ij} = 0$ otherwise. Also, we set $U = 10$ as established to be realistic for cuprate materials [123, 124].

In many elemental metals, electron interactions are effectively weak ($U \approx 0$) due to electric-field screening by the surrounding lattice ions, yielding conventional metallic behavior at half-filling (one electron per site). By contrast, in materials like high-T_c cuprates, the on-site Coulomb repulsion becomes abnormally strong for electrons, opening a large energy gap that penalizes double occupancy. Hence, electron motion is substantially hindered by a large potential barrier and the system exhibits Mott insulating behavior at the macroscopic level.

Extra charge carriers can be introduced by adding (electron doping) or removing (hole doping) electrons relative to half-filling. Upon sufficient doping, charge transport in the material sets in and superconductivity may emerge, signified by enhanced pairing correlations at low temperatures. This evolution underlies the schematic phase diagram as shown in Fig. 1, where red and blue shaded areas mark the AFM Mott insulating phase and the high-T_c superconducting dome, respectively.

The (unnormalized) thermal density matrix $\rho = e^{-\beta\mathcal{H}}$ characterizes the statistical state of the lattice system, with inverse temperature $\beta = 1/T$. Note that ρ admits a Taylor expansion at high temperature (small β) for a given Hamiltonian, and that

$$\rho(2\beta) = \rho(\beta) \cdot \rho(\beta). \tag{2}$$

Accordingly, one may cool the system down by repeatedly squaring the thermal density matrix starting from a high-temperature construct. This idea underlies the eXponential Tensor Renormalization Group (XTRG) method [76, 77, 79, 101, 125], which offers a comprehensive thermal description of the lattice system over a broad temperature range. We thus employ XTRG to produce thermal density matrices at high, medium, and low temperatures at over-doped, medium-doped, and under-doped regions (hole-doped), yielding nine categories as indicated in Fig. 1.

We then perform standard site-wise sampling on each thermal density matrix [79] to obtain snapshots (Fock bases of the many-

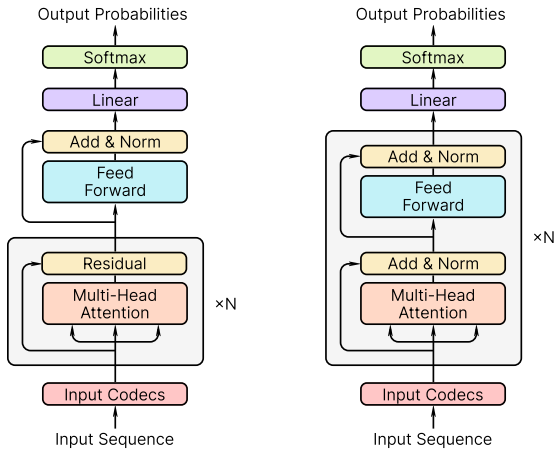


FIG. 2. Schematic illustrations of the core (left) and the pro (right) architecture for classification of sequential inputs. Both architectures comprise input codecs, multi-head attention blocks, feed-forward networks and a final linear classification head. The pro architecture is an analog of the encoder-only transformer, while the core architecture leaves out the feed-forward networks between attention blocks which enhances parallelism and improves interpretability.

body Hilbert space) of the lattice system. The snapshots, each consisting of 8×8 cells containing either $|\varnothing\rangle$, $|\uparrow\rangle$, $|\downarrow\rangle$, or $|\uparrow\downarrow\rangle$, are next flattened according to row-major order into a sequence of 64 elements. For each location in the phase space, we generate 1000 snapshots, yielding a dataset of 9 categories and 9000 snapshots in total. This dataset is randomly partitioned into training (90%) and test (10%) subsets for the subsequent AI workflows.

III. ARCHITECTURES

Our AI architectures originate from the transformer paradigm while being tailored for categorical classification. Given an input snapshot, the objective is to infer a probability distribution over categories. Rather than appending a dedicated `CLS` token [8, 16] as a label, we adopt a streamlined design that endows the involved attention mechanism with a distinctive and physically meaningful interpretation.

Figure 2 depicts schematic layouts of the models deployed in this study. Both architectures share the same foundational components — input codecs, a stack of multi-head attention, feed-forward networks, and a terminal linear classification head. The pro architecture (right) is inspired by an encoder-only transformer, whereas the core architecture (left) leaves out the feed-forward networks between attention blocks, thereby giving enhanced parallelism and improved interpretability.

Our exposition below assumes familiarity with standard components and techniques in the practices of the transformer architecture [6], and hence will focus on our distinctive designs. Full technical details can be found in the supplemental material [122].

A. Tokenization & Input Codecs

The tokenization is straightforward in our setting, as the *vocabulary* (local Hilbert space) comprises only four distinct *words* (local states). We therefore assign `0`, `1`, `2`, and `3` to the empty, spin-up, spin-down, and doubly occupied states, respectively. Under this encoding, each snapshot in the dataset now becomes a sequence of integers (tokens). Formally, let $\mathcal{S} = \{0, 1, 2, 3\}$ denote the tokenized local state space. The input sequence becomes $\vec{\sigma} \in \mathcal{S}^L$, where L is the flattened sequence length.

The input codecs accept and map each tokenized sequence into the model's latent parameter space in two stages. Each token σ is first transformed into a d_{model}-dimensional embedding vector $\boldsymbol{e}(\sigma) \in \mathbb{R}^{d_{\text{model}}}$ via a learnable embedding module; this embedding depends solely on the token (local state) and is agnostic to its location in the sequence. To preserve positional information, we then add a positional vector $\boldsymbol{\varrho}_i \in \mathbb{R}^{d_{\text{model}}}$ for site i to each embedding vector. For this purpose, we employ the sinusoidal positional encoding [6], which has proved effective across a wide range of applications.

The input codecs thus assemble a feature matrix Σ with elements $\Sigma_i^\mu = e^\mu(\sigma_i) + \varrho_i^\mu$ for each snapshot, where i indexes lattice sites and μ indexes latent dimensions. Contingent on the dataset under consideration, d_{model} should be adjusted for a balance of expressivity against overfitting. Note that in [6] the input embeddings are multiplied by a factor of $\sqrt{d_{\text{model}}}$ to scale up the weights; in contrast, this operation is empirically detrimental in our application domain, plausibly due to the exceedingly small vocabulary size relative to the sequence length.

B. Locality-Biased Attention

The (multi-head) attention mechanism plays a central role in harnessing global correlation awareness for both architectural designs. We adopt the prevalent scaled dot-product attention scheme [1, 6] to acquire raw attention scores, and subsequently impose a locality bias for an improved training profile. Even though positional information has been encoded amid the input codecs, we find that the prototypical attention setup, which is primarily designed for 1D sequences, struggles in perceiving 2D spatial relationships. Hence, an explicit locality bias assists in this regard.

We start with linear projections of the input embeddings into query, key, and value vectors

$$\boldsymbol{\mathcal{Q}}_i = \boldsymbol{\Sigma}_i W_Q, \quad \boldsymbol{\mathcal{K}}_i = \boldsymbol{\Sigma}_i W_K, \quad \boldsymbol{\mathcal{V}}_i = \boldsymbol{\Sigma}_i W_V \tag{3}$$

with W_Q, W_K, and W_V being learnable weight matrices. Here, we suppress the latent-space index μ and take matrix multiplications implicit. The attention between site i and j thus reads

$$A_{ij} = \mathrm{softmax}_j(\boldsymbol{\mathcal{Q}}_i \boldsymbol{\mathcal{K}}_j / \mathfrak{T}), \tag{4}$$

where $\mathfrak{T}$ denotes the *model temperature* (conceptually distinct from the physical temperature) which controls the sharpness of the attention distribution. In our exercises, $\mathfrak{T} = \sqrt{d_k}$ (see below for the definition of d_k) works reasonably well.

The multi-head attention is realized by partitioning the latent space into h subspaces, each with dimension $d_k = d_{\text{model}}/h$. Attention is computed independently within each head, after which the head outputs are concatenated and linearly transformed to yield the final raw attention. In our actual practice, the multi-head configurations fail to outperform their single-head counterpart, likely attributable to the limited size of the training set.

Many realistic physical systems exhibit locality: objects only significantly influence their immediate neighbors, leading to a decay of the interactions and correlations with spatial separation. Accordingly, we apply a locality bias to the raw attention scores

$$\mathcal{A}_{ij} = \mathrm{softmax}_j(A_{ij} \circ G_{ij}), \tag{5}$$

where G_{ij} is a hand-crafted bias function that decays with the physical distance d_{ij} between sites i and j, and the circle $\circ$ denotes element-wise (Hadamard) multiplication. This locality bias encourages the mechanism to focus on nearby sites, effectively accelerating the convergence during the training process. In our implementation, we choose a Gaussian kernel

$$G_{ij} = \exp\left\{-d_{ij}^2 / 2\varsigma^2\right\}, \tag{6}$$

with standard deviation $\varsigma = \lambda/2$ and λ a characteristic length scale of the system (e.g., $\lambda = 8$ for the 8×8 square lattice considered here). The eventual performance is not highly affected by the specific choice of the bias function G_{ij}. For instance, a power-law decay kernel works almost equally well.

Finally, as a standard technique to stabilize the gradient propagation, we apply a residual connection [126] by adding a $\boldsymbol{\Sigma}_i$ to the output of the attention block as

$$\mathrm{attn}(\boldsymbol{\Sigma}_i) = \boldsymbol{\Sigma}_i + \textstyle\sum_j \mathcal{A}_{ij} \boldsymbol{\mathcal{V}}_j. \tag{7}$$

Afterwards, layer normalization [127] is employed in the pro architecture, whereas the core architecture omits this step for reasons that will become clear in the ensuing interpretation.

C. Feed-Forward & Classification

The feed-forward networks (FFNs) constitute one of the principal sources of non-linearity in the model. Each FFN is a site-wise fully-connected three-layer perceptron comprising an input layer, a hidden layer, and an output layer. The input and output layers have width d_{model}, while the hidden layer has width d_{hidden}. A ReLU activation is applied between the two affine maps to introduce non-linearity. Concretely, the FFN reads

$$\mathrm{FFN}(\boldsymbol{\Sigma}_i) = \mathrm{ReLU}(\boldsymbol{\Sigma}_i W_1 + \boldsymbol{b}_1) W_2 + \boldsymbol{b}_2, \tag{8}$$

where W_1, W_2, $\boldsymbol{b}_1$, and $\boldsymbol{b}_2$ are learnable weights and biases. The same linear maps are shared across all sites i, whereas different FFN blocks carry independent parameters. A residual connection and layer normalization follow each FFN.

The arrangement of FFNs constitutes the pivotal difference between the pro and core architectures. The pro variant inserts an FFN of dimension $d_{\text{hidden}} = d_{\text{ff}}$ after each attention block, while the core variant defers non-linearity to a single FFN of dimension $d_{\text{hidden}} = N \times d_{\text{ff}}$ after the entire stack of N attention blocks. We call the latter design *semi-linear* attention stack (for the reason that will become clear in the impending interpretation).

Under this parametrization, the core and pro architectures contain an almost equivalent amount of learnable model parameters. However, the postponed feed-forward network markedly reduces the depth of non-linearity within the core model, which benefits the upcoming interpretation since physical objects commonly propagate linearly. Also, centralizing the FFN boosts parallelism during both training and inference.

The classification head ingests the abstract embedding $\boldsymbol{\Sigma}_i^{\text{ff}}$ processed through the preceding attention and feed-forward networks, and returns a categorical distribution over the target labels. Concretely, the logit $y_{i,c}$ and the corresponding probability $p_{i,c}$ are computed as

$$y_{i,c} = \boldsymbol{\Sigma}_i^{\text{ff}} W_c + \boldsymbol{b}_c, \quad p_{i,c} = \mathrm{softmax}_c(y_{i,c}) \tag{9}$$

where W_c and $\boldsymbol{b}_c$ denote the learnable weight and bias associated with category c. Each site i produces its own distribution; this is warranted by the attention mechanism, which injects into each site contextual information aggregated from all other sites. The overall prediction is obtained by the argmax of the averaged per-site distributions $p_c = \mathrm{avg}_i\, p_{i,c}$ over all lattice sites.

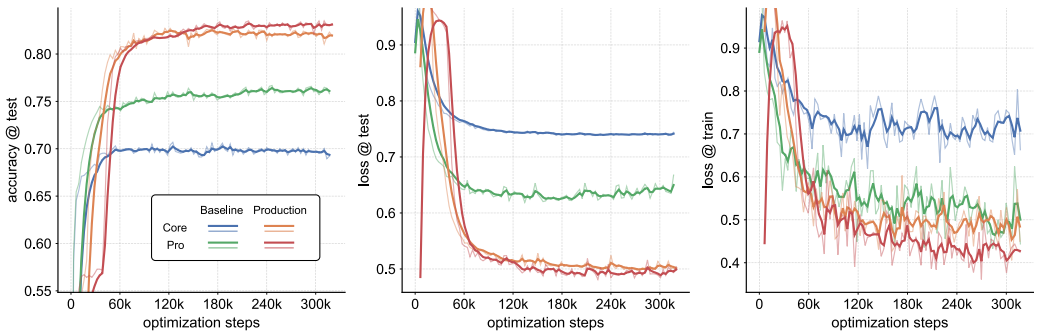


FIG. 3. Training profiles and benchmarks of the production and baseline models following the core and pro architectures. Metrics are displayed as raw data (thin lines with muted color) and with an exponential smoothing factor $\alpha = 0.4$ (thick lines with deep color). For baseline models (with trivialized attention), the pro architecture consistently outperforms the core variant across all metrics, congruous with the anticipated benefits of elevated non-linearity in the pro model. By contrast, for production models (with full-functional attention), the core architecture achieves merely negligible gaps in performance, indicating an alignment of the semi-linear attention with the intrinsic properties of the dataset.

D. Hyperparameters

For the XTRG-generated snapshot dataset of the Hubbard model, we adopt the following hyperparameters: embedding dimension $d_{\text{model}} = 128$; single-head attention $h = 1$; feed-forward dimension $d_{\text{ff}} = 1024$; and $N = 2$ attention blocks. Increasing either the number of heads or the number of attention blocks empirically induces severe overfitting and should be considered only after expanding the dataset.

Furthermore, we implement an ablation toggle that trivializes all attention blocks by hard-setting $A_{ij} = 1$ when enabled. Under this switch, the effective attention reduces to the fixed locality kernel, thereby delegating the classification task entirely to the FFNs. This toggle is useful both as a baseline and during the warm-up phase of training.

IV. TRAINING & BENCHMARKS

Hardware. — All the production and baseline models were trained on an NVIDIA 3090 GPU. The project utilizes PyTorch version 2.2.2 and CUDA 11.8.

Initialization. — All learnable parameters are initialized with Xavier (Glorot) initialization [128]. For production runs, the first 200 epochs serve as a warm-up phase during which the ablation toggle is enabled, effectively pre-training the FFNs with trivialized attention.

Batch & Epochs. — We use a batch size of 256 and train for a total of 10,000 epochs. Checkpoints are saved every 100 epochs, and the one with the lowest validation loss is selected as the final deliverable.

Objective. — We optimize the ubiquitous cross-entropy loss as our optimization objective. The loss is computed as the negative log-likelihood averaged over all sites

$$\text{loss} = \text{avg}_i \left[-\sum_c \xi_c \log p_{i,c} \right], \tag{10}$$

where ξ_c denotes the ground-truth one-hot label for category c (broadcasted across all sites), or specifically

$$\xi_c = \begin{cases} 1 & \text{for correct category } c, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Optimizer. — We use the standard Adam optimizer [129] with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-9}$. The learning rate is fixed to 5×10^{-6}. We find that both architectural designs are sensitive to this setting: materially larger or smaller values tend to induce premature plateaus at elevated loss.

Regularization. — We apply dropouts [130] with a rate of 0.01 to input codecs, attention blocks, FFNs and all residual connections. Contrary to the common practice in the NLP applications, we disable the label smoothing [131] as over-confidence is not a primary concern for a physically-generated dataset.

Training Yield. — Owing to stochastic initialization, training outcomes exhibit variability. Empirically, roughly one in seven attempts attains a top-performing model.

Benchmarks. — Figure 3 summarizes the training profiles of four models we trained: the production and baseline models under the core and pro architectures. The baselines are trained with the aforementioned ablation toggle switched on, such that their attention mechanisms are effectively disabled.

Benchmarks are reported as test-set accuracy and loss, along with training loss. Test accuracy is defined as the fraction of cor-

rectly classified snapshots in the held-out test set. All four models achieve stable convergence given sufficient training, and generalization is satisfactory as indicated by the proximity of training and test losses.

Moreover, the production models consistently surpass their baseline counterparts, attesting to the efficacy of the attention mechanism. For the baselines, the pro variant outperforms the core across all metrics, consistent with the anticipated benefits of enhanced non-linearity. By contrast, in production runs, the core model closes the gap to within negligible differences, indicating that FFNs interleaved between attention blocks are largely redundant and that the semi-linear attention stack aligns well with the intrinsic structure of the dataset.

V. INTERPRETATION

For decades, achieving a principled interpretation of the internal mechanism of AI models has been one of the highest endeavors in the field [132]. Unlike opaque black-box approaches [133], the attention mechanism offers a natural lens on a model's focus and decision-making. The calculated attention scores $\mathcal{A}_{ij}$ are typically construed as a measure of *importance* of token j to token i (or equivalently, the *attention* paid by token i to token j) [2,3]. This heuristic has been widely utilized in a variety of application domains [4, 134–142] for analysis, diagnosis and debugging.

However, a comprehensive study [143] showed that attention weights often fail to provide consistent or exclusive explanations of model predictions; in particular, alternative attention patterns can yield essentially identical performance. These observations have ignited a prolonged debate [143–151] on whether — or to what extent — the attention meaningfully reveals a model's reasoning process.

Despite the ongoing controversy, consensus remains that attention maps furnish at least *an* (if not *the*) explanation for the inner workings of the model [146]. Thereafter, further aggregation methods, such as attention rollout [152–154] and attention flow [152, 155, 156], have been proposed to propagate attention scores across multiple layers. The attention rollout, in particular, essentially performs a layer-wise matrix multiplication of the (residual-augmented) attention matrices. Considering the fact that these matrices are all row-stochastic, it becomes natural to interpret them as Markovian transition kernels [153]. This probabilistic viewpoint forms the basis of our interpretation.

Our interpretation focuses exclusively on the core architecture, as it depends critically on the *semi-linear* nature of the attention stack (the precise meaning of which will be clarified in the upcoming subsections).

A. Classical & Quantum Markov Process

Before heading to the interpretation, we first formalize the relevant constructs for both classical and quantum Markov processes on the lattice system. For the classical scenario, consider a discrete-time Markov dynamics in which, at each update, the state s_i at site i may overwrite the state s_j at site j; the corresponding transition probability is $\mathcal{A}_{ij}$. All lattice sites update synchronously in one time step.

Suppose that a collection of observables is associated with each local state, and write Σ_i^μ for the μ-th observable evaluated on the state at site i. Under the Markov evolution described above, the observables update after one step as

$$\Sigma_i^\mu \leftarrow \textstyle\sum_j \mathcal{A}_{ij} \Sigma_j^\mu. \tag{12}$$

For the quantum scenario, the local state can be associated with a (pure) local density matrix $\rho_i = |\sigma_i\rangle\langle\sigma_i| \equiv |i\rangle\langle i|$. The quantum Markov process is specified by a completely-positive trace-preserving (CPTP) map $\mathcal{E}$ (also known as a quantum channel) comprising a set of Kraus operators

$$K_j = \textstyle\sum_\imath \sqrt{\mathcal{A}_{\imath j}}\, |j\rangle\langle\imath|, \tag{13}$$

whose action on the local state is

$$\mathcal{E}(\rho_i) = \textstyle\sum_j K_j \rho_i K_j^\dagger = \sum_j \mathcal{A}_{ij} \rho_j. \tag{14}$$

Assign to each ρ_i the same family of observables Σ_i^μ. Under the channel $\mathcal{E}$, these observables also evolve according to Eq. (12). Therefore, in both classical and quantum constructs, Eq. (12) captures the one-step evolution of observables under the Markov process with the transition kernel $\mathcal{A}_{ij}$.

B. Interpretation of the Attention Stack

We now make explicit the link between the attention stack in the core architecture and the Markovian description above. Viewing the embedded features as observables, the right-hand side of Eq. (12) coincides with the application of the attention matrix $\mathcal{A}_{ij}$ to the input embedding Σ_i^μ. In conjunction with the embedding projector W_V and the residual connection, the output of a stack of $N=2$ attention blocks can be expressed as

$$\begin{aligned}
\boldsymbol{\Sigma}_i^{\text{attn}} &= \boldsymbol{\Sigma}_i + \textstyle\sum_k \mathcal{A}_{ik}^{(1)} \boldsymbol{\Sigma}_k W_V^{(1)} \\
&\quad + \textstyle\sum_j \mathcal{A}_{ij}^{(2)} \left[\boldsymbol{\Sigma}_j + \sum_k \mathcal{A}_{jk}^{(1)} \boldsymbol{\Sigma}_k W_V^{(1)} \right] W_V^{(2)} \\
&= \boldsymbol{\Sigma}_i + \textstyle\sum_j \mathcal{A}_{ij}^{(1)} \boldsymbol{\Sigma}_j W_V^{(1)} + \sum_j \mathcal{A}_{ij}^{(2)} \boldsymbol{\Sigma}_j W_V^{(2)} \\
&\quad + \textstyle\sum_{j,k} \mathcal{A}_{ij}^{(2)} \mathcal{A}_{jk}^{(1)} \boldsymbol{\Sigma}_k W_V^{(1)} W_V^{(2)},
\end{aligned} \tag{15}$$

where $\mathcal{A}_{ij}^{(\ell)}$ and $W_V^{(\ell)}$ denote, respectively, the attention matrix and the embedding projector from the ℓ-th attention block. On the right-hand side of the final equality, four terms appear: the first term carries the original embedding; the second and third terms propagate features according to the first and second attention kernels; and the final term captures the cascaded propagation through both blocks. These terms can be interpreted as four dis-

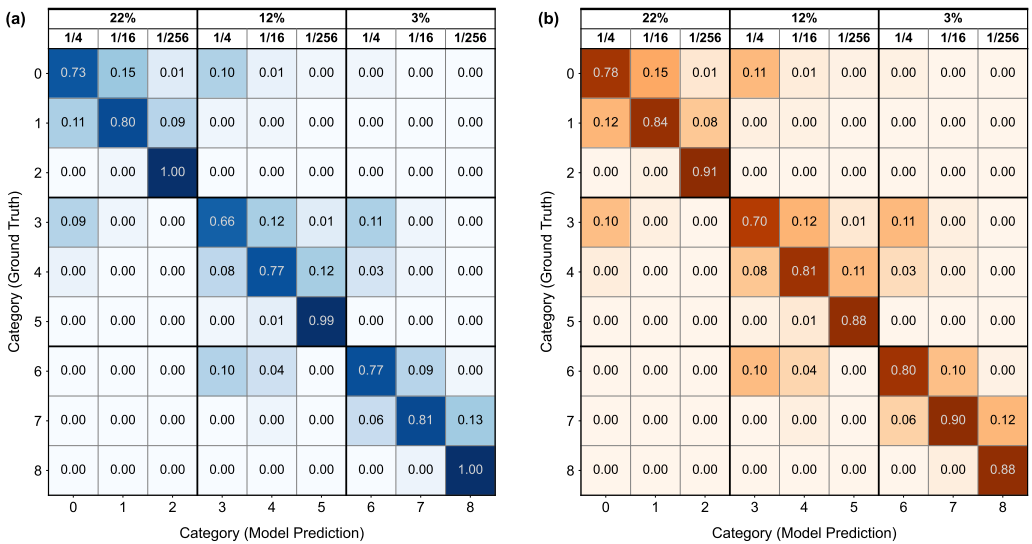


FIG. 4. (a) Sensitivity matrix (row-normalized confusion matrix) and (b) precision matrix (column-normalized confusion matrix) for the core model. Color intensity indicates the degree of sensitivity (a) and precision (b), with exact values written within each cell. Top rows indicate the charge doping and temperature of the corresponding categories. Diagonal entries show (a) the probability of correct classification for each category and (b) the probability of a predicted category being correct.

tinct *Markovian propagation modes* acting on the input $\boldsymbol{\Sigma}_i$. The same expansion generalizes analogously to deeper stacks.

For trivial embedding projectors $W_V^{(\ell)} = \mathbb{1}$ (where $\mathbb{1}$ is the identity matrix), Eq. (15) reduces to

$$\boldsymbol{\Sigma}_i^{\text{attn}} = \sum_j \mathcal{R}_{ij} \boldsymbol{\Sigma}_j = \sum_j \left[\prod_\ell (\mathbb{1} + \mathcal{A}^{(\ell)})_{ij} \right] \boldsymbol{\Sigma}_j, \qquad (16)$$

where $\mathcal{R}_{ij}$ is precisely the standard attention rollout [152] (without normalization). We refer to this limiting case as a *linear* attention stack. However, in practice, the embedding projectors are generally non-trivial, yielding the *semi-linear* attention stack. This is the unique feature of the core architecture; by contrast, in the pro variant, the layer normalization and the FFN introduce non-linearity between attention blocks.

This perspective furnishes an interesting interpretation of the core architecture. First, the input codecs learn a feature embedding whose components can be construed as physically relevant observables attached to each local state. The attention stack then effects a superposition of Markovian propagation modes that evolve these observables across the lattice, while residual connections preserve the original features. Finally, the classification head operates on the resultant evolved features (observables). In essence, the model learns an *effective* Markov dynamics — encoded in the attention kernels and embedding projections — that best aligns the propagated observables with the downstream classification objective.

VI. CONFUSION ANALYSIS

The core (production) model attains an overall 83% accuracy on the test subset, as delineated in Fig. 3. However, this aggregate metric masks substantial variation across categories; a nuanced assessment requires a full-scale confusion analysis of per-category sensitivity (true positive rate) and precision (positive predictive value).

We construct the confusion matrix Ξ, where each entry $\Xi_{cc'}$ counts snapshots whose ground-truth category is c but are classified as c'. Row-normalizing Ξ yields the sensitivity matrix, which estimates the probability $p(c'|c)$ that a snapshot from category c is predicted as c'. Column-normalizing Ξ produces the precision matrix, which estimates the probability $p(c|c')$ that a snapshot predicted as c' actually originates from category c.

Figures 4(a) and 4(b) display the sensitivity and precision matrices, respectively. The 9×9 matrices are partitioned into a 3×3 (doping) block of 3×3 (temperature) cells and exhibit a pronounced block-diagonal structure, indicating that misclassifications occur predominantly within the same doping level. Moreover, sensitivity increases systematically as temperature decreases (reaching almost 100% at the lowest temperature), whereas precision does not exhibit an equally monotonic trend.

To rationalize these tendencies, we identify two principal sources of randomness in the dataset: thermal and quantum fluctuations. Thermal fluctuations are essentially structureless and

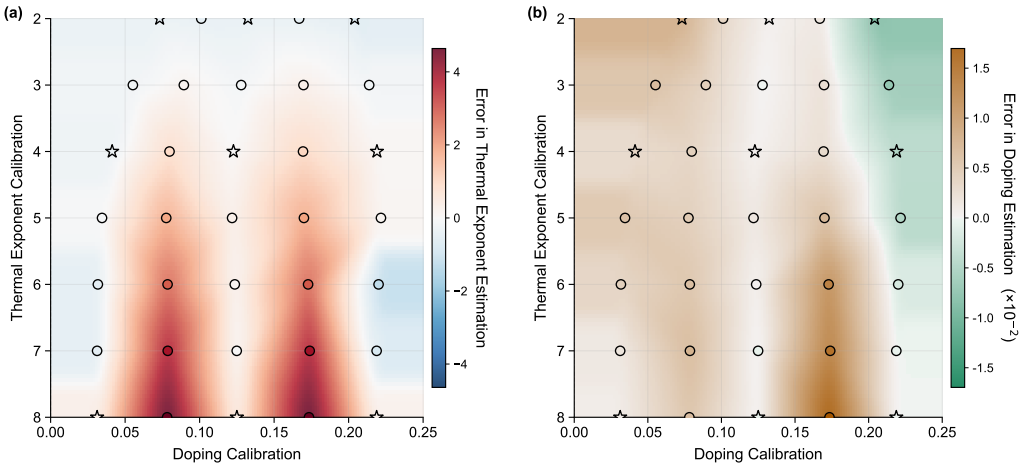


FIG. 5. Error in the omnimeter estimation of (a) the thermal exponent and (b) charge doping. Open circles/pentagrams mark the locations in phase space of the snapshot ensembles under evaluation, with pentagrams (circles) indicating data included (not included) in the training set. Color scales are obtained via interpolation. Overall performance is good, except for the bands at the unseen doping levels (around 7% and 17% in (a) and around 17% in (b)).

uncorrelated, while quantum fluctuations can admit non-trivial quantum correlations. The latter effectively enhance the system's entanglement entropy, thereby providing additional information that aids discrimination across categories.

Under this perspective, sensitivity may be interpreted as the *prominence* of a category's correlation pattern — greater correlation strength raises the likelihood of correct classification; whereas precision reflects the *uniqueness* of that pattern — greater distinctiveness reduces the chance that snapshots from other categories are misattributed to it.

Consequently, the observed rise in sensitivity at lower temperatures suggests increasingly prominent correlation structures, consistent with the suppressed thermal noise (and relatively accentuated quantum correlations). By contrast, the more modest gains in precision at the lowest temperatures imply that low-temperature patterns also occur at a higher temperature, in line with the known persistence of Mott physics and superconducting correlations into a moderate-temperature regime.

Lastly, we remark that misclassification is not catastrophic in our context, since the dataset is itself intrinsically stochastic and exhibits substantial randomness. Therefore, an argmax-based decision rule may mislabel even under a Bayes-optimal classifier. For instance, let's suppose a perfect model assigns a snapshot 50% probability to category c and 40% to category c'; the argmax strategy will thereby deterministically predict category c, although this snapshot may quite plausibly originate from category c'. Hence, one should regard the probability distribution p_c as the faithful output of the model.

VII. UNIVERSAL OMNIMETRY

One of the straightforward applications of our AI classifier is to perform measurements on an arbitrary ensemble of snapshots. Each category in the dataset is affiliated with a set of known observables, and the classifier outputs a distribution over categories, thereby inducing an estimate of the corresponding observables. We refer to this procedure as *omnimetry*, since all affiliated observables are inferred simultaneously.

For a demonstration of this new technique, we augment the dataset with additional snapshots drawn from regions of phase space that were not included in the training set (i.e., *unseen* by the model). A sufficiently generalizable model should then produce a distribution over categories that reflects the resemblance of the correlation patterns in the input ensemble against those learned during training. Thus, the performance of the omnimeter serves as a probe of the underlying correlation structure across distinct partitions of phase space.

The workflow starts with a *calibration* of observables for all categories in the training set. Let $\omega_c^{(\alpha)}$ denote the α-th observable affiliated with category c. In our study of the finite-temperature Hubbard model, the affiliated observables are temperature and charge doping. As XTRG algorithm produces thermal density matrices at temperatures $T = 1/2^{n_T}$, we utilize the *thermal exponent* n_T as a representative observable in place of temperature. Additional observables can, of course, be accommodated, provided the dataset supports and a corresponding calibration is available.

Next, given an ensemble of snapshots $\{x\}$ generated under fixed conditions, the core classifier returns, for each snapshot, a categorical distribution $p_c(x)$ over c. Averaging these distributions across the ensemble yields a collective distribution measuring the probability that this ensemble corresponds to category c. The target observable for the ensemble is then estimated by the distribution-weighted average of the calibration

$$\langle \omega^{(\alpha)} \rangle = \sum_c \omega_c^{(\alpha)} \, \mathrm{avg}_x p_c(x). \tag{17}$$

Figure 5(a) shows the error of the omnimeter's estimates of the thermal exponent n_T for ensembles drawn from various locations in phase space. Overall performance is satisfactory — particularly at doping levels partially covered in the training set. Notably, two red bands appear at unseen doping levels (around 7% and 17%), with inferred temperatures systematically higher than the ground truth. This bias is plausibly attributed to correlation patterns in those regions that were absent during training; the model interprets these as signatures of elevated thermal fluctuations and hence predicts higher temperature. This behavior, in turn, corroborates that the classifier has genuinely learned to associate correlation patterns with thermodynamic conditions.

Figure 5(b) reports the error in estimating charge doping. Performance is again strong, aside from a distinct band at the unseen doping of around 17%. A similar mechanism applies: the model finds that the unseen patterns resemble those at around 22% doping, consistent with both doping levels lying within the superconducting regime.

These artifacts can be eliminated by augmenting the training data in the relevant portions of phase space. As a trailer, we can announce that a 25-category classifier which includes the problematic doping levels in its training set substantially mitigates these issues, reducing the relative error in both temperature and doping to below 10%. Further details will be available in the supplemental material [122] as well as an upcoming dedicated technical report.

Thermometry remains a central challenge in ultracold-atom experiments [85, 106], and our AI omnimeter offers a competitive upgrade. Modern quantum gas microscopy [82, 83] produces ensembles of site-resolved snapshots of the analog cold-atom simulator, which can be compared directly with our numerical snapshot dataset. Whereas current thermometers often rely on matching hand-selected metrics and correlators [85, 106, 107, 157], our approach automatically discovers discriminative patterns and aggregates them into robust temperature estimates. We therefore anticipate that this approach can materially enhance the reliability of thermometry in ultracold-atom platforms.

VIII. SUMMARY & OUTLOOK

In this Article, we establish an end-to-end technological stack for AI-assisted analysis of strongly correlated electron systems on a lattice. The workflow starts with tensor-network simulations that generate thermal density matrices and, in turn, an extensive snapshot dataset. This dataset is then processed by our tailored AI architectures featuring locality-biased, semi-linear attention with principled interpretability grounded in effective Markovian dynamics and strong capacity to capture latent correlation patterns. The trained model is subsequently subjected to a comprehensive confusion analysis, revealing the prominence and uniqueness of correlation structures across thermodynamic conditions. Finally, the model is deployed as an omnimeter to infer multiple observables from arbitrary ensembles of snapshots.

Our research demonstrates the viability of bespoke AI technologies for interrogating challenging strongly correlated quantum systems. The approach is versatile and readily extends to lattice models for diverse physical scenarios. Moreover, the tailored architectures are especially useful for sequential datasets with small *vocabulary* but long *sequence length*, a characteristic feature of many scientific data such as DNA/RNA chains, protein structures, or other high-resolution imaging. The observation that the core model attains performance comparable to the pro variant suggests further opportunities to optimize the transformer architecture. Besides, the universal omnimetry furnishes a generic measurement methodology for quantum many-body experiments equipped with site-resolved quantum microscopy [82, 83]. Beyond classifiers, alternative AI paradigms — e.g., generative models — merit applications in e.g. autoregressive neural quantum states [158–162] or neural transformer backflow [163–165]. We therefore anticipate that this work opens a promising new avenue for the study of strongly correlated systems and will motivate further researches along these lines.

ACKNOWLEDGEMENTS

We thank Annabelle Bohrdt, Fabian Döschl, Hannah Lange, Roger Melko and Henning Schlömer for helpful discussions and feedback. This research was funded in part by the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy EXC-2111 (Project No. 390814868), and is part of the Munich Quantum Valley, supported by the Bavarian state government through the Hightech Agenda Bayern Plus.

[1] D. Bahdanau, K. Cho, and Y. Bengio, in *International Conference on Learning Representations* (2014).

[2] T. Rocktäschel, E. Grefenstette, K. M. Hermann, T. Kočiský, and P. Blunsom, in *International Conference on Learning Representations* (2015).

[3] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, in *Proceedings of the 32nd International Conference on Machine Learning* (PMLR, 2015) pp. 2048–2057.

[4] Y. Wang, M. Huang, X. Zhu, and L. Zhao, in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, edited by J. Su, K. Duh, and X. Carreras (Association for Computational Linguistics, Austin, Texas, 2016) pp. 606–615.

[5] Y. Kim, C. Denton, L. Hoang, and A. M. Rush, in *International Conference on Learning Representations* (2017).

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, in *Advances in Neural Information Processing Systems*, Vol. 30 (Curran Associates, Inc., 2017).

[7] L. Dong, S. Xu, and B. Xu, in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2018) pp. 5884–5888.

[8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (long and Short Papers)*, edited by J. Burstein, C. Doran, and T. Solorio (Association for Computational Linguistics, Minneapolis, Minnesota, 2019) pp. 4171–4186.

[9] Y. Ren, Y. Ruan, X. Tan, T. Qin, S. Zhao, Z. Zhao, and T.-Y. Liu, in *Advances in Neural Information Processing Systems*, Vol. 32 (Curran Associates, Inc., 2019).

[10] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, in *Advances in Neural Information Processing Systems*, Vol. 33 (Curran Associates, Inc., 2020) pp. 1877–1901.

[11] A. Gulati, J. Qin, C.-C. Chiu, N. Parmar, Y. Zhang, J. Yu, W. Han, S. Wang, Z. Zhang, Y. Wu, and R. Pang, in *Interspeech* (2020).

[12] S. Latif, A. Zaidi, H. Cuayahuitl, F. Shamshad, M. Shoukat, M. Usama, and J. Qadir, in *Computer Science Review* (2023).

[13] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, in *Computer Vision – ECCV 2020*, edited by A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm (Springer International Publishing, Cham, 2020) pp. 213–229.

[14] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, in *International Conference on Learning Representations* (arXiv, 2020).

[15] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, in *IEEE International Conference on Computer Vision* (arXiv, 2021) pp. 9992–10002.

[16] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, ACM Comput, Surv, **54**, 200:1 (2022).

[17] A. Jaegle, F. Gimeno, A. Brock, A. Zisserman, O. Vinyals, and J. Carreira, in *International Conference on Machine Learning* (arXiv, 2021).

[18] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, Nature **596**, 583 (2021).

[19] A. Rives, J. Meier, T. Sercu, S. Goyal, Z. Lin, J. Liu, D. Guo, M. Ott, C. L. Zitnick, J. Ma, and R. Fergus, Proc. Natl. Acad. Sci. **118**, e2016239118 (2021).

[20] L. Chen, K. Lu, A. Rajeswaran, K. Lee, A. Grover, M. Laskin, P. Abbeel, A. Srinivas, and I. Mordatch, in *Neural Information Processing Systems* (2021) pp. 15084–15097.

[21] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan, and L. Sun, in *International Joint Conference on Artificial Intelligence* (arXiv, 2022).

[22] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, in *AAAI Conference on Artificial Intelligence* (arXiv, 2020).

[23] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, in *International Conference on Learning Representations* (2018).

[24] E. Dagotto, Science **309**, 257 (2005).

[25] A. Avella and F. Mancini, eds., *Strongly Correlated Systems: Theoretical Methods*, Springer Series in Solid-State Sciences, Vol. 171 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012).

[26] A. Avella and F. Mancini, eds., *Strongly Correlated Systems: Numerical Methods*, Springer Series in Solid-State Sciences, Vol. 176 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013).

[27] A. Avella and F. Mancini, eds., *Strongly Correlated Systems: Experimental Techniques*, Springer Series in Solid-State Sciences, Vol. 180 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2015).

[28] N. F. Mott, Proc. Phys. Soc. London, Sect. A **62**, 416 (1949).

[29] D. B. McWhan, A. Menth, J. P. Remeika, W. F. Brinkman, and T. M. Rice, Phys. Rev. B **7**, 1920 (1973).

[30] M. Imada, A. Fujimori, and Y. Tokura, Rev. Mod. Phys. **70**, 1039 (1998).

[31] J. G. Bednorz and K. A. Müller, Z. Phys. B: Condens. Matter **64**, 189 (1986).

[32] M. K. Wu, J. R. Ashburn, C. J. Torng, P. H. Hor, R. L. Meng, L. Gao, Z. J. Huang, Y. Q. Wang, and C. W. Chu, Phys. Rev. Lett. **58**, 908 (1987).

[33] Y. Tokura, H. Takagi, and S. Uchida, Nature **337**, 345 (1989).

[34] A. Schilling, M. Cantoni, J. D. Guo, and H. R. Ott, Nature **363**, 56 (1993).

[35] J. Kondo, Prog. Theor. Phys. **32**, 37 (1964).

[36] K. Andres, J. E. Graebner, and H. R. Ott, Phys. Rev. Lett. **35**, 1779 (1975).

[37] F. Steglich, J. Aarts, C. D. Bredl, W. Lieke, D. Meschede, W. Franz, and H. Schäfer, Phys. Rev. Lett. **43**, 1892 (1979).

[38] Y. Kamihara, T. Watanabe, M. Hirano, and H. Hosono, J. Am. Chem. Soc. **130**, 3296 (2008).

[39] D. C. Tsui, H. L. Stormer, and A. C. Gossard, Phys. Rev. Lett. **48**, 1559 (1982).

[40] R. B. Laughlin, Phys. Rev. Lett. **50**, 1395 (1983).

[41] H. L. Stormer, D. C. Tsui, and A. C. Gossard, Rev. Mod. Phys. **71**, S298 (1999).

[42] P. W. Anderson, Mater. Res. Bull. **8**, 153 (1973).

[43] F. D. M. Haldane, Phys. Rev. Lett. **50**, 1153 (1983).

[44] Y. Shimizu, K. Miyagawa, K. Kanoda, M. Maesato, and G. Saito, Phys. Rev. Lett. **91**, 107001 (2003).

[45] P. Szirmai, C. Mézière, G. Bastien, P. Wzietek, P. Batail, E. Martino, K. Mantulnikovs, A. Pisoni, K. Riedl, S. Cottrell, C. Baines, L. Forró, and B. Náfrádi, Proc. Natl. Acad. Sci. **117**, 29555 (2020).

[46] D. J. Gross and F. Wilczek, Phys. Rev. Lett. **30**, 1343 (1973).

[47] H. David Politzer, Phys. Rep. **14**, 129 (1974).

[48] P. Koch, B. Müller, and J. Rafelski, Phys. Rep. **142**, 167 (1986).

[49] K. G. Wilson, Nucl. Phys. B Proc. Suppl. LATTICE 2004, **140**, 3 (2005).

[50] R. Blankenbecler, D. J. Scalapino, and R. L. Sugar, Phys. Rev. D **24**, 2278 (1981).

[51] G. Sugiyama and S. Koonin, Ann. Phys. **168**, 1 (1986).
[52] M. Qin, C.-M. Chung, H. Shi, E. Vitali, C. Hubig, U. Schollwöck, S. R. White, S. Zhang, and S. C. o. t. M.-E. Problem, Phys. Rev. X **10**, 031016 (2020).
[53] H. Xu, C.-M. Chung, M. Qin, U. Schollwöck, S. R. White, and S. Zhang, Science **384**, eadh7691 (2024).
[54] M. Capone and G. Kotliar, Phys. Rev. B **74**, 054513 (2006).
[55] G. Knizia and G. K.-L. Chan, Phys. Rev. Lett. **109**, 186404 (2012).
[56] B.-X. Zheng and G. K.-L. Chan, Phys. Rev. B **93**, 035126 (2016).
[57] T. I. Vanhala and P. Törmä, Phys. Rev. B **97**, 075112 (2018).
[58] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).
[59] S. R. White and D. J. Scalapino, Phys. Rev. Lett. **80**, 1272 (1998).
[60] U. Schollwöck, Ann. Phys. **326**, 96 (2011).
[61] E. Stoudenmire and S. R. White, Annu. Rev. Condens. Matter Phys. **3**, 111 (2012).
[62] A. Gleis, J.-W. Li, and J. von Delft, Phys. Rev. Lett. **130**, 246402 (2023).
[63] F. Verstraete and J. I. Cirac, Arxiv 10.48550/arxiv.cond-mat/0407066 (2004).
[64] J. Jordan, R. Orús, G. Vidal, F. Verstraete, and J. I. Cirac, Phys. Rev. Lett. **101**, 250602 (2008).
[65] H. C. Jiang, Z. Y. Weng, and T. Xiang, Phys. Rev. Lett. **101**, 090603 (2008).
[66] T. Barthel, C. Pineda, and J. Eisert, Phys. Rev. A **80**, 042333 (2009).
[67] C. V. Kraus, N. Schuch, F. Verstraete, and J. I. Cirac, Phys. Rev. A **81**, 052338 (2010).
[68] P. Corboz, R. Orús, B. Bauer, and G. Vidal, Phys. Rev. B **81**, 165104 (2010).
[69] P. Corboz, J. Jordan, and G. Vidal, Phys. Rev. B **82**, 245119 (2010).
[70] C. Zhang, J.-W. Li, D. Nikolaidou, and J. von Delft, Phys. Rev. Lett. **134**, 116502 (2025).
[71] S. R. White, Phys. Rev. Lett. **102**, 190601 (2009).
[72] E. M. Stoudenmire and S. R. White, New J. Phys. **12**, 55026 (2010).
[73] W. Li, S.-J. Ran, S.-S. Gong, Y. Zhao, B. Xi, F. Ye, and G. Su, Phys. Rev. Lett. **106**, 127202 (2011).
[74] P. Czarnik, L. Cincio, and J. Dziarmaga, Phys. Rev. B **86**, 245101 (2012).
[75] P. Czarnik and J. Dziarmaga, Phys. Rev. B **92**, 035152 (2015).
[76] B.-B. Chen, L. Chen, Z. Chen, W. Li, and A. Weichselbaum, Phys. Rev. X **8**, 031082 (2018).
[77] H. Li, B.-B. Chen, Z. Chen, J. von Delft, A. Weichselbaum, and W. Li, Phys. Rev. B **100**, 045110 (2019).
[78] Q. Li, Y. Gao, Y.-Y. He, Y. Qi, B.-B. Chen, and W. Li, Phys. Rev. Lett. **130**, 226502 (2023).
[79] C. Zhang and J. von Delft (2025), arXiv:2510.25022.
[80] A. Mazurenko, C. S. Chiu, G. Ji, M. F. Parsons, M. Kanász-Nagy, R. Schmidt, F. Grusdt, E. Demler, D. Greif, and M. Greiner, Nature **545**, 462 (2017).
[81] C. S. Chiu, G. Ji, A. Mazurenko, D. Greif, and M. Greiner, Phys. Rev. Lett. **120**, 243201 (2018).
[82] J. Koepsell, J. Vijayan, P. Sompet, F. Grusdt, T. A. Hilker, E. Demler, G. Salomon, I. Bloch, and C. Gross, Nature **572**, 358 (2019).
[83] J. Koepsell, S. Hirthe, D. Bourgund, P. Sompet, J. Vijayan, G. Salomon, C. Gross, and I. Bloch, Phys. Rev. Lett. **125**, 010403 (2020).
[84] T. Chalopin, P. Bojović, D. Bourgund, S. Wang, T. Franz, I. Bloch, and T. Hilker, Phys. Rev. Lett. **134**, 53402 (2025).
[85] M. Xu, L. H. Kendrick, A. Kale, Y. Gang, C. Feng, S. Zhang, A. W. Young, M. Lebrat, and M. Greiner, Nature **642**, 909 (2025).
[86] J. Hubbard, Proc. R. Soc. London, Ser. A **296**, 82 (1967).
[87] P. W. Anderson, Science **235**, 1196 (1987).
[88] J. P. F. LeBlanc, A. E. Antipov, F. Becca, I. W. Bulik, G. K.-L. Chan, C.-M. Chung, Y. Deng, M. Ferrero, T. M. Henderson, C. A. Jiménez-Hoyos, E. Kozik, X.-W. Liu, A. J. Millis, N. V. Prokof'ev, M. Qin, G. E. Scuseria, H. Shi, B. V. Svistunov, L. F. Tocchio, I. S. Tupitsyn, S. R. White, S. Zhang, B.-X. Zheng, Z. Zhu, E. Gull, and S. C. o. t. M.-E. Problem, Phys. Rev. X **5**, 041041 (2015).
[89] B.-X. Zheng, C.-M. Chung, P. Corboz, G. Ehlers, M.-P. Qin, R. M. Noack, H. Shi, S. R. White, S. Zhang, and G. K.-L. Chan, Science **358**, 1155 (2017).
[90] N. J. Robinson, P. D. Johnson, T. M. Rice, and A. M. Tsvelik, Rep. Prog. Phys. **82**, 126501 (2019).
[91] M. Qin, T. Schäfer, S. Andergassen, P. Corboz, and E. Gull, Annu. Rev. Condens. Matter Phys. **13**, 275 (2022).
[92] H.-C. Jiang and T. P. Devereaux, Science **365**, 1424 (2019).
[93] Y.-F. Jiang, J. Zaanen, T. P. Devereaux, and H.-C. Jiang, Phys. Rev. Res. **2**, 33073 (2020).
[94] C.-M. Chung, M. Qin, S. Zhang, U. Schollwöck, and S. R. White, Phys. Rev. B **102**, 041106 (2020).
[95] Y.-F. Jiang, T. P. Devereaux, and H.-C. Jiang, Phys. Rev. B **109**, 85121 (2024).
[96] B. Ponsioen, S. S. Chung, and P. Corboz, Phys. Rev. B **100**, 195141 (2019).
[97] A. Wietek, Y.-Y. He, S. R. White, A. Georges, and E. M. Stoudenmire, Phys. Rev. X **11**, 31007 (2021).
[98] A. Bohrdt, C. S. Chiu, G. Ji, M. Xu, D. Greif, M. Greiner, E. Demler, F. Grusdt, and M. Knap, Nat. Phys. **15**, 921 (2019).
[99] C. S. Chiu, G. Ji, A. Bohrdt, M. Xu, M. Knap, E. Demler, F. Grusdt, M. Greiner, and D. Greif, Science **365**, 251 (2019).
[100] G. Salomon, J. Koepsell, J. Vijayan, T. A. Hilker, J. Nespolo, L. Pollet, I. Bloch, and C. Gross, Nature **565**, 56 (2019).
[101] B.-B. Chen, C. Chen, Z. Chen, J. Cui, Y. Zhai, A. Weichselbaum, J. von Delft, Z. Y. Meng, and W. Li, Phys. Rev. B **103**, L041107 (2021).
[102] J. Koepsell, D. Bourgund, P. Sompet, S. Hirthe, A. Bohrdt, Y. Wang, F. Grusdt, E. Demler, G. Salomon, C. Gross, and I. Bloch, Science **374**, 82 (2021).
[103] P. Sompet, S. Hirthe, D. Bourgund, T. Chalopin, J. Bibo, J. Koepsell, P. Bojović, R. Verresen, F. Pollmann, G. Salomon, C. Gross, T. A. Hilker, and I. Bloch, Nature **606**, 484 (2022).
[104] S. Hirthe, T. Chalopin, D. Bourgund, P. Bojović, A. Bohrdt, E. Demler, F. Grusdt, I. Bloch, and T. A. Hilker, Nature **613**, 463 (2023).
[105] M. Xu, L. H. Kendrick, A. Kale, Y. Gang, G. Ji, R. T. Scalettar, M. Lebrat, and M. Greiner, Nature **620**, 971 (2023).
[106] T. Chalopin, P. Bojović, S. Wang, T. Franz, A. Sinha, Z. Wang, D. Bourgund, J. Obermeyer, F. Grusdt, A. Bohrdt, L. Pollet, A. Wietek, A. Georges, T. Hilker, and I. Bloch (2024), arXiv:2412.17801 [cond-mat].
[107] G. Pasqualetti, O. Bettermann, N. Darkwah Oppong, E. Ibarra-García-Padilla, S. Dasgupta, R. T. Scalettar, K. R. A. Hazzard, I. Bloch, and S. Fölling, Phys. Rev. Lett. **132**, 83401 (2024).
[108] D. Bourgund, T. Chalopin, P. Bojović, H. Schlömer, S. Wang, T. Franz, S. Hirthe, A. Bohrdt, F. Grusdt, I. Bloch, and T. A.

Hilker, Nature **637**, 57 (2025).
[109] X. Dong, E. Gull, and A. J. Millis, Nat. Phys. **18**, 1293 (2022).
[110] H. Xu, H. Shi, E. Vitali, M. Qin, and S. Zhang, Phys. Rev. Res. **4**, 13239 (2022).
[111] B. Xiao, Y.-Y. He, A. Georges, and S. Zhang, Phys. Rev. X **13**, 011007 (2023).
[112] S. d. A. Sousa-Júnior, N. C. Costa, and R. R. dos Santos, Phys. Rev. B **109**, 165102 (2024).
[113] S. Jiang, D. J. Scalapino, and S. R. White, Phys. Rev. B **108**, L161111 (2023).
[114] A. Bohrdt, Y. Wang, J. Koepsell, M. Kánasz-Nagy, E. Demler, and F. Grusdt, Phys. Rev. Lett. **126**, 026401 (2021).
[115] C. Miles, A. Bohrdt, R. Wu, C. Chiu, M. Xu, G. Ji, M. Greiner, K. Q. Weinberger, E. Demler, and E.-A. Kim, Nat. Commun. **12**, 3905 (2021).
[116] L. W. Cheuk, M. A. Nichols, K. R. Lawrence, M. Okan, H. Zhang, E. Khatami, N. Trivedi, T. Paiva, M. Rigol, and M. W. Zwierlein, Science **353**, 1260 (2016).
[117] T. A. Hilker, G. Salomon, F. Grusdt, A. Omran, M. Boll, E. Demler, I. Bloch, and C. Gross, Science **357**, 484 (2017).
[118] F. Grusdt, M. Kánasz-Nagy, A. Bohrdt, C. S. Chiu, G. Ji, M. Greiner, D. Greif, and E. Demler, Phys. Rev. X **8**, 11046 (2018).
[119] F. Grusdt and L. Pollet, Phys. Rev. Lett. **125**, 256401 (2020).
[120] Z. Y. Weng, D. N. Sheng, Y.-C. Chen, and C. S. Ting, Phys. Rev. B **55**, 3894 (1997).
[121] T.-L. Ho, Proc. Natl. Acad. Sci. **117**, 26141 (2020).
[122] See supplemental material for an introduction to the transformer architecture for physicists, additional technical details, benchmarks and analysis, and a performance preview of a 25-category omnimeter.
[123] M. Hirayama, Y. Yamaji, T. Misawa, and M. Imada, Phys. Rev. B **98**, 134501 (2018).
[124] M. Hirayama, T. Misawa, T. Ohgoe, Y. Yamaji, and M. Imada, Phys. Rev. B **99**, 245155 (2019).
[125] X. Lin, B.-B. Chen, W. Li, Z. Y. Meng, and T. Shi, Phys. Rev. Lett. **128**, 157201 (2022).
[126] K. He, X. Zhang, S. Ren, and J. Sun, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016) pp. 770–778.
[127] J. L. Ba, J. R. Kiros, and G. E. Hinton (2016), arXiv:1607.06450 [stat].
[128] X. Glorot and Y. Bengio, in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (JMLR Workshop and Conference Proceedings, 2010) pp. 249–256.
[129] D. P. Kingma and J. Ba, in *International Conference on Learning Representations* (2014).
[130] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, J. Mach. Learn. Res. **15**, 1929 (2014).
[131] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, in *Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Las Vegas, NV, USA, 2016) pp. 2818–2826.
[132] A. Bibal and B. Frénay, in *The European Symposium on Artificial Neural Networks* (2016).
[133] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, ACM Comput, Surv, **51**, 93:1 (2018).
[134] E. Choi, M. T. Bahadori, J. A. Kulas, A. Schuetz, W. F. Stewart, and J. Sun, in *Neural Information Processing Systems* (arXiv, 2016) pp. 3504–3512.
[135] A. Martins and R. Astudillo, in *Proceedings of the 33rd International Conference on Machine Learning* (PMLR, 2016) pp. 1614–1623.
[136] J. Lee, J.-H. Shin, and J.-S. Kim, in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, edited by L. Specia, M. Post, and M. Paul (Association for Computational Linguistics, Copenhagen, Denmark, 2017) pp. 121–126.
[137] Q. Xie, X. Ma, Z. Dai, and E. Hovy, in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (volume 1: Long Papers)*, edited by R. Barzilay and M.-Y. Kan (Association for Computational Linguistics, Vancouver, Canada, 2017) pp. 950–962.
[138] M. Dehghani, S. Gouws, O. Vinyals, J. Uszkoreit, and Ł. Kaiser, in *International Conference on Learning Representations* (2018).
[139] G. Brunner, Y. Liu, D. Pascual, O. Richter, M. Ciaramita, and R. Wattenhofer, in *International Conference on Learning Representations* (2019).
[140] H. Chen and Y. Ji, in *Robust AI in Financial Services* (arXiv, 2019).
[141] K. Clark, U. Khandelwal, O. Levy, and C. D. Manning, in *Proceedings of the 2019 ACL Workshop Blackboxnlp: Analyzing and Interpreting Neural Networks for NLP*, edited by T. Linzen, G. Chrupała, Y. Belinkov, and D. Hupkes (Association for Computational Linguistics, Florence, Italy, 2019) pp. 276–286.
[142] A. Coenen, E. Reif, A. Yuan, B. Kim, A. Pearce, F. Viégas, and M. Wattenberg (2019), arXiv:1906.02715.
[143] S. Jain and B. C. Wallace, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (long and Short Papers)*, edited by J. Burstein, C. Doran, and T. Solorio (Association for Computational Linguistics, Minneapolis, Minnesota, 2019) pp. 3543–3556.
[144] S. Vashishth, S. Upadhyay, G. S. Tomar, and M. Faruqui, arXiv 10.48550/arXiv.1909.11218 (2019).
[145] J. Vig (2019), arXiv:1904.02679.
[146] S. Wiegreffe and Y. Pinter, in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, edited by K. Inui, J. Jiang, V. Ng, and X. Wan (Association for Computational Linguistics, Hong Kong, China, 2019) pp. 11–20.
[147] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, in *Proceedings of the 36th International Conference on Machine Learning* (PMLR, 2019) pp. 7354–7363.
[148] D. Pruthi, M. Gupta, B. Dhingra, G. Neubig, and Z. C. Lipton, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, edited by D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault (Association for Computational Linguistics, Online, 2020) pp. 4782–4793.
[149] B. Bai, J. Liang, G. Zhang, H. Li, K. Bai, and F. Wang, in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21 (Association for Computing Machinery, New York, NY, USA, 2021) pp. 25–34.
[150] A. Galassi, M. Lippi, and P. Torroni, IEEE Trans. Neural Netw. Learn. Syst. **32**, 4291 (2021).
[151] A. Bibal, R. Cardon, D. Alfter, R. Wilkens, X. Wang, T. François, and P. Watrin, in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (volume 1: Long Papers)*, edited by S. Muresan, P. Nakov, and A. Villavicencio (As-

sociation for Computational Linguistics, Dublin, Ireland, 2022) pp. 3889–3900.
[152] S. Abnar and W. Zuidema, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, edited by D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault (Association for Computational Linguistics, Online, 2020) pp. 4190–4197.
[153] T. Yuan, X. Li, H. Xiong, H. Cao, and D. Dou, in *eXplainable AI approaches for debugging and diagnosis.* (2021).
[154] L. Xu, X. Yan, W. Ding, and Z. Liu, J. Ambient Intell. Hum. Comput. **14**, 163 (2023).
[155] N. Metzger, C. Hahn, J. Siber, F. Schmitt, and B. Finkbeiner, in *International Conference on Learning Representations* (2023).
[156] B. Azarkhalili and M. W. Libbrecht, in *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (volume 1: Long Papers)*, edited by W. Che, J. Nabende, E. Shutova, and M. T. Pilehvar (Association for Computational Linguistics, Vienna, Austria, 2025) pp. 19954–19974.
[157] T. Hartke, Phys. Rev. Lett. **125**, 113601 (2020).
[158] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carrasquilla, Phys. Rev. Res. **2**, 23358 (2020).
[159] D. Luo, Z. Chen, K. Hu, Z. Zhao, V. M. Hur, and B. K. Clark, Phys. Rev. Res. **5**, 13216 (2023).
[160] Y.-H. Zhang, Phys. Rev. B **107**, 10.1103/PhysRevB.107.075147 (2023).
[161] E. Ibarra-García-Padilla, H. Lange, R. G. Melko, R. T. Scalettar, J. Carrasquilla, A. Bohrdt, and E. Khatami, Phys. Rev. Research **7**, 013122 (2025).
[162] H. Lange, G. Bornet, G. Emperauger, C. Chen, T. Lahaye, S. Kienle, A. Browaeys, and A. Bohrdt, Quantum **9**, 1675 (2025).
[163] D. Luo and B. K. Clark, Phys. Rev. Lett. **122**, 226401 (2019).
[164] A.-J. Liu and B. K. Clark, Phys. Rev. B **110**, 115137 (2024).
[165] L. Zhang and D. Luo (2025), arXiv:2509.09275 [cond-mat].

Supplemental Material - Interpretable Artificial Intelligence (AI) Analysis of Strongly Correlated Electrons

Changkai Zhang (张昌凯) and Jan von Delft
Arnold Sommerfeld Center for Theoretical Physics, Center for NanoScience, and Munich Center for Quantum Science and Technology, Ludwig-Maximilians-Universität München, 80333 Munich, Germany
(Dated: November 2, 2025)

In the supplemental material, we provide (S-I) an introduction for physicists to the transformer architecture; (S-II) detailed specifications of the snapshot dataset for the Hubbard model; (S-III) a performance benchmark of both architectures on an artificial derangement dataset; (S-IV) an analysis of orthogonality and attention maps; and (S-V) a performance preview of a 25-category omnimeter.

S-I. INTRODUCTION TO THE TRANSFORMER

In this section, we present an elementary introduction for physicists to the transformer architecture, grounded in the seminal work *Attention is All You Need* [6]. The transformer was originally developed for sequence transduction tasks in natural language processing (NLP). Subsequently, an encoder-only variant [8] has been proposed for generative or classification tasks, which we further develop into the *pro* architecture in the main text. Here, we focus on this particular instantiation of the transformer as applied to physical lattice models, wherein snapshots can be regarded as sequences in the *language* of the physical system. And classifying a given snapshot into one of the nine categories in phase space is akin to e.g. classifying a sentence into one of several sentiment classes in NLP.

Tokenization. — Tokens are the pre-defined elementary units of the input sequence. In NLP, tokens are typically words, whitespaces, punctuations, etc. For snapshots of a lattice system, tokens are the local states σ on each lattice site, e.g., empty, spin-up, spin-down, and double-occupied states for the Fermi-Hubbard model. The input sequence is then a one-dimensional array of tokens obtained by flattening the two-dimensional (2D) lattice snapshot in a row-major order. We assign `0`, `1`, `2`, and `3` to the four local states, respectively. Therefore, a snapshot of a lattice system with L sites is now *tokenized* into an input sequence $\vec{\sigma} \in \{0, 1, 2, 3\}^L$.

Input Embedding. — The tokenizer described above assigns a unique integer to each token (local state). However, these integers are purely nominal and do not encode any semantic information about the corresponding local state. A more informative strategy is to represent each token by a vector of *features*. Specifically, one could use an array $(\pi_c, n_c, s_z, S, \ldots)$ comprising, e.g., parity π_c, number of particles n_c, spin-z s_z, total spin S, etc., to represent each local state. In this example, the spin-up state would be encoded as $(-1, 1, +\frac{1}{2}, \frac{1}{2}, \ldots)$, and the other local states follow analogously. We refer to this hand-crafted representation as an *input encoding* of the sequence.

However, an input encoding requires manual identification of the relevant features for each local state and may thus be constrained by prior knowledge about the system. A more flexible approach is to allow the model to learn a suitable representation of each token directly from data. This is achieved via an *input embedding*, in which all features are learnable parameters. In practice, the embedding layer is essentially a lookup table that stores an embedding vector $\boldsymbol{e}(\sigma) \in \mathbb{R}^{d_{\text{model}}}$ for each local state σ. The dimension d_{model} (number of features) of the embedding vectors is a hyperparameter to be chosen when constructing the model. The components of $\boldsymbol{e}(\sigma)$ are denoted $e^{\mu}(\sigma)$, where $\mu = 1, 2, \ldots, d_{\text{model}}$ is the feature index.

Positional Encoding. — The embedding vector of a token depends only on the local state it represents and carries no information about its position in the sequence. We therefore need a separate mechanism to inject positional information. The transformer architecture contains neither recurrent nor convolutional structures — common devices in other architectures for capturing sequential order — and instead relies on a *positional encoding*. Analogous to the input encoding, the positional encoding is a fixed (non-learnable) map that converts each position i in the input sequence into a positional vector $\boldsymbol{\varrho}_i \in \mathbb{R}^{d_{\text{model}}}$. A common choice for the positional encoding is to use sine and cosine functions of different frequencies:

$$\varrho_i^{\mu} = \begin{cases} \sin(i/10000^{2\mu/d_{\text{model}}}), & \text{if } \mu \text{ is even,} \\ \cos(i/10000^{2\mu/d_{\text{model}}}), & \text{if } \mu \text{ is odd,} \end{cases} \tag{S1}$$

The motivation for this sinusoidal form is to enable the model to infer relative positions between tokens, since any $\boldsymbol{\varrho}_{i+k}$ can be expressed as a linear function of $\boldsymbol{\varrho}_i$. Similar to the input embedding, it is also possible to employ a learnable *positional embedding*. However, in practice, we do not observe a benefit from this upgrade, consistent with the findings in [6].

Input Codecs. — The input codecs consolidate the input embedding and positional encoding. The embedding vector $\boldsymbol{e}(\sigma_i) \equiv \boldsymbol{e}_i$ of the token at position i is scaled by w_e and added to the positional encoding $\boldsymbol{\varrho}_i$ to yield the final input representation $\Sigma_i^{\mu} = w_e e_i^{\mu} + \varrho_i^{\mu}$, i.e. $\boldsymbol{\Sigma}_i = w_e \boldsymbol{e}_i + \boldsymbol{\varrho}_i \in \mathbb{R}^{d_{\text{model}}}$. Consequently, for every snapshot, the tokenized input sequence $\vec{\sigma}$ of L tokens is transformed into a feature matrix

$$\begin{bmatrix}\sigma_1\\ \sigma_2\\ \vdots\\ \sigma_L\end{bmatrix} \to \begin{bmatrix}\Sigma_1^1 & \Sigma_1^2 & \cdots & \Sigma_1^{d_{\text{model}}}\\ \Sigma_2^1 & \Sigma_2^2 & \cdots & \Sigma_2^{d_{\text{model}}}\\ \vdots & \vdots & \ddots & \vdots\\ \Sigma_L^1 & \Sigma_L^2 & \cdots & \Sigma_L^{d_{\text{model}}}\end{bmatrix} \tag{S2}$$

whose rows enumerate positions and columns enumerate features. In [6], the weight is set to $w_e = \sqrt{d_{\text{model}}}$. In their NLP tasks, the vocabulary size (number of unique tokens) is approximately 37,000, while the sequence length is about 25,000. It is therefore natural to emphasize the input embedding relative to the positional encoding. In our physical applications, however, the vocabulary size is usually small (e.g., 4 for the Fermi-Hubbard model) whereas the sequence length can be comparatively large (e.g., 64 for an 8×8 lattice). Hence, we instead use $w_e = 1$ to place the input embedding and positional encoding on an equal footing.

Modular Design. — Contemporary AI systems commonly adopt a modular design, wherein the overall architecture comprises a sequence of *modules* (depicted as rectangular blocks in Fig. 2) with standardized inputs and/or outputs, enabling algorithms to be assembled in a building-block fashion. In the transformer architecture, all the constituent modules consume and/or emit data in the same format of feature matrix $\Sigma \in \mathbb{R}^{L\times d_{\text{model}}}$. This uniform interface greatly simplifies the construction of deep models via a straight-forward stacking of modules. Accordingly, it is natural to regard $\boldsymbol{\Sigma}_i$ as a *register* memory or a module *argument*, rather than a specific mathematical entity with fixed values, and one should understand its significance and the contents stored according to the context.

Dot-Product Attention. — The principal workhorse of the transformer is the attention mechanism, which enables the model to capture long-range (global) correlations across the input sequence. The attention module receives and converts $\boldsymbol{\Sigma}_i$ into three sets of vectors: the *queries* $\boldsymbol{\mathcal{Q}}_i \in \mathbb{R}^{d_k}$, the *keys* $\boldsymbol{\mathcal{K}}_i \in \mathbb{R}^{d_k}$, and the *values* $\boldsymbol{\mathcal{V}}_i \in \mathbb{R}^{d_v}$. For self-attention, one commonly takes $d_v = d_k$, and obtains the queries, keys, and values via linear projections of the input codecs:

$$\begin{aligned} Q_i^\nu &= \sum_{\mu=1}^{d_{\text{model}}} \Sigma_i^\mu W_Q^{\mu\nu}, \quad \text{or} \quad \boldsymbol{\mathcal{Q}}_i = \boldsymbol{\Sigma}_i W_Q \in \mathbb{R}^{d_k},\\ K_i^\nu &= \sum_{\mu=1}^{d_{\text{model}}} \Sigma_i^\mu W_K^{\mu\nu}, \quad \text{or} \quad \boldsymbol{\mathcal{K}}_i = \boldsymbol{\Sigma}_i W_K \in \mathbb{R}^{d_k},\\ V_i^\nu &= \sum_{\mu=1}^{d_{\text{model}}} \Sigma_i^\mu W_V^{\mu\nu}, \quad \text{or} \quad \boldsymbol{\mathcal{V}}_i = \boldsymbol{\Sigma}_i W_V \in \mathbb{R}^{d_v}, \end{aligned} \tag{S3}$$

where $W_Q, W_K \in \mathbb{R}^{d_{\text{model}}\times d_k}$, and $W_V \in \mathbb{R}^{d_{\text{model}}\times d_v}$ are learnable linear projection matrices. The same set of projection matrices are shared across all positions i.

The objective of the attention module is to compute *similarities* between queries and keys and to reweight the values accordingly. In the *scaled dot-product attention* [6], similarities are measured according to the dot products of queries with keys, so the *attention score* A_{ij} between the i-th query and the j-th key is given by

$$A_{ij} = \text{softmax}_j\left(\boldsymbol{\mathcal{Q}}_i\boldsymbol{\mathcal{K}}_j/\mathfrak{T}\right) = \frac{1}{Z_i}\exp(\boldsymbol{\mathcal{Q}}_i\boldsymbol{\mathcal{K}}_j/\mathfrak{T}), \tag{S4}$$

where

$$Z_i = \sum_{j=1}^{L}\exp(\boldsymbol{\mathcal{Q}}_i\boldsymbol{\mathcal{K}}_j/\mathfrak{T}) \tag{S5}$$

is the normalization factor, and $\mathfrak{T}$ the *model temperature* parameter (conceptually distinct from the actual physical temperature) that controls the distribution of attention scores. Following [6], we set $\mathfrak{T} = \sqrt{d_k}$. An inner product should be inferred in the expression

$$\boldsymbol{\mathcal{Q}}_i\boldsymbol{\mathcal{K}}_j = \sum_{\nu=1}^{d_k}\mathcal{Q}_i^\nu\mathcal{K}_j^\nu, \tag{S6}$$

and in Eq. S4, the subscript j in softmax_j indicates that the softmax operation is taken along the j index. The dot-product attention thus outputs

$$\text{attn}(\boldsymbol{\Sigma}_i \mid W_Q, W_K, W_V) = \sum_{j=1}^{L} A_{ij}\boldsymbol{\mathcal{V}}_j. \tag{S7}$$

Intuitively, the attention mechanism can be recognized as a *fuzzy* dictionary lookup. Rather than executing a *hard* retrieval that selects the value associated with a single, exactly matching key, the mechanism instead computes attention scores that quantify the *degree of similarity* or correspondence between the query and all available keys. These scores are then used to form a weighted combination of the associated values, thereby producing a context-dependent output representation.

Multi-Head Attention. — The above single attention can be extended to *multi-head attention* by partitioning the query, key, and value vectors into h parallel subspaces (heads) with per-head width d_k such that $d_{\text{model}} = h\cdot d_k$. The dot-product attention is then applied independently within each head. Concretely, the η-th head ($\eta = 1, 2, \ldots, h$) computes

$$\boldsymbol{\Sigma}_i^{(\eta)} = \text{attn}(\boldsymbol{\Sigma}_i \mid W_Q^\eta, W_K^\eta, W_V^\eta) \in \mathbb{R}^{d_v}. \tag{S8}$$

Note that each head has its own set of projection matrices W_Q^η, $W_K^\eta \in \mathbb{R}^{d_{\text{model}}\times d_k}$, and $W_V^\eta \in \mathbb{R}^{d_{\text{model}}\times d_v}$. The outputs of all heads are subsequently concatenated along the feature dimension to form a vector in $\mathbb{R}^{h\cdot d_v}$, before passing through a linear projection with a learnable matrix $W_O \in \mathbb{R}^{h\cdot d_v\times d_{\text{model}}}$. The final output of the multi-head attention layer is thus

$$\boldsymbol{\Sigma}_i^{\text{attn}} = \text{concat}(\boldsymbol{\Sigma}_i^{(1)}, \ldots, \boldsymbol{\Sigma}_i^{(h)})W_O \in \mathbb{R}^{d_{\text{model}}}. \tag{S9}$$

For a single-head attention, the output projection W_O is redundant, since it can be absorbed into W_V. We therefore omit this projection in the main text.

In practice, a convenient and computationally efficient implementation maintains three *shared* projection matrices W_Q, W_K, and W_V of shape $d_{\text{model}} \times d_{\text{model}}$. The overall queries, keys, and values are computed, and then partitioned into h heads:

$$\mathcal{Q} = \left[\mathcal{Q}^{(1)} \quad \mathcal{Q}^{(2)} \quad \mathcal{Q}^{(3)} \quad \cdots \quad \mathcal{Q}^{(h)} \right],$$
$$\mathcal{K} = \left[\mathcal{K}^{(1)} \quad \mathcal{K}^{(2)} \quad \mathcal{K}^{(3)} \quad \cdots \quad \mathcal{K}^{(h)} \right], \tag{S10}$$
$$\mathcal{V} = \left[\mathcal{V}^{(1)} \quad \mathcal{V}^{(2)} \quad \mathcal{V}^{(3)} \quad \cdots \quad \mathcal{V}^{(h)} \right].$$

Here, the i-th row of the matrices $\mathcal{Q}$, $\mathcal{K}$, and $\mathcal{V} \in \mathbb{R}^{L \times d_{\text{model}}}$ corresponds to the vectors $\boldsymbol{\mathcal{Q}}_i$, $\boldsymbol{\mathcal{K}}_i$, and $\boldsymbol{\mathcal{V}}_i$, respectively. Consequently, the output of the η-th head becomes

$$\mathcal{Z}^{(\eta)} = \text{softmax}\left[\mathcal{Q}^{(\eta)} \quad [\mathcal{K}^{\top}]^{(\eta)} \right] / \mathfrak{Z} \quad \mathcal{V}^{(\eta)}, \tag{S11}$$

where the softmax is applied row-wise. This vectorized implementation is more efficient in practice, as it leverages highly optimized parallel linear algebra routines. The final output of the multi-head attention block is thus

$$\mathcal{Z}^{\text{attn}} = \left[\mathcal{Z}^{(1)} \quad \mathcal{Z}^{(2)} \quad \mathcal{Z}^{(3)} \quad \cdots \quad \mathcal{Z}^{(h)} \right] W_O. \tag{S12}$$

Feed-Forward Network. — Following the multi-head attention block, a position-wise feed-forward network (FFN) is applied independently to each received $\boldsymbol{\mathcal{Z}}_i$. Again, the same FFN (i.e., the same parameters) is shared across all positions i. Conceptually, the FFN is a three-layer fully-connected perceptron comprising an input layer, a widened hidden layer, and an output layer. The input and output layers have the dimension d_{model}, matching the output of the multi-head attention, while the hidden layer has a larger width d_{hidden} to enhance the model's representational capacity. A schematic illustration is provided in Fig. S1.

Specifically, the FFN applies the following transformation to each position i in the sequence:

$$\text{FFN}(\boldsymbol{\mathcal{Z}}_i) = \text{ReLU}(\boldsymbol{\mathcal{Z}}_i W_1 + \boldsymbol{b}_1) W_2 + \boldsymbol{b}_2, \tag{S13}$$

where $W_1 \in \mathbb{R}^{d_{\text{model}} \times d_{\text{hidden}}}$ and $W_2 \in \mathbb{R}^{d_{\text{hidden}} \times d_{\text{model}}}$ are learnable projection matrices, and $\boldsymbol{b}_1 \in \mathbb{R}^{d_{\text{hidden}}}$, $\boldsymbol{b}_2 \in \mathbb{R}^{d_{\text{model}}}$ (learnable) bias vectors. The ReLU (*Rectified Linear Unit*) activation function, defined as $\text{ReLU}(x) = \max(0, x)$, supplies the crucial non-linearity in the FFN. Figure S2 depicts the overall shape of the ReLU function.

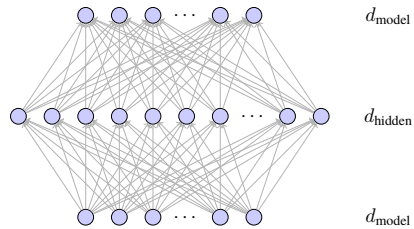


FIG. S1. Schematic diagram of a three-layer feed-forward network (FFN) used in the transformer architecture. The input and output layers have dimension d_{model}, while the hidden layer has dimension d_{hidden}. Each neuron in a given layer is connected to all neurons in the adjacent layers.

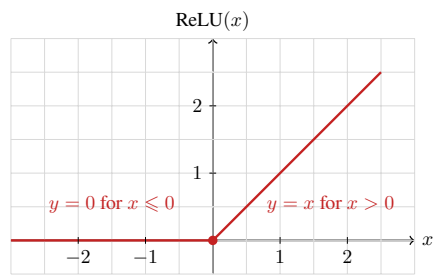


FIG. S2. Schematic diagram of the ReLU (Rectified Linear Unit) activation function. The function zeros out the negative inputs and increases linearly for positive inputs, providing essential non-linearity while maintaining computational simplicity.

Output Projection. — Following the final FFN of the transformer, vectors $\boldsymbol{\mathcal{Z}}_i$ are passed to an output projection module responsible for producing the model's logits. This output projection is usually implemented as a learnable linear transformation into the output space $\mathcal{G}$ of the downstream task (e.g. categories for classification or vocabulary in text generation). For each position i, the output logits $y_i \in \mathbb{R}^{|\mathcal{G}|}$ are computed as

$$\boldsymbol{y}_i = \boldsymbol{\mathcal{Z}}_i W_y + \boldsymbol{b}_y, \tag{S14}$$

where $W_y \in \mathbb{R}^{d_{\text{model}} \times |\mathcal{G}|}$ denotes the weight matrix and $\boldsymbol{b}_y \in \mathbb{R}^{|\mathcal{G}|}$ the bias. The logits are subsequently normalized via a softmax function to yield probabilities $\boldsymbol{p}_i = \text{softmax}(\boldsymbol{y}_i) \in \mathbb{R}^{|\mathcal{G}|}$. In generative scenarios, $\boldsymbol{p}_i$ guides the sampling of the output token at position i. For classification or regression tasks, one may aggregate the outputs across all positions (e.g., via averaging $\boldsymbol{p} = \text{avg}_i(\boldsymbol{p}_i)$) or employ a dedicated `CLS` token [8, 16] to derive a single, sequence-level prediction.

Residual Connection. — Deep neural networks are prone to vanishing gradients during training; the *residual connection* [126] is a standard remedy that markedly stabilizes the optimization process. The key idea is to introduce a shortcut path that

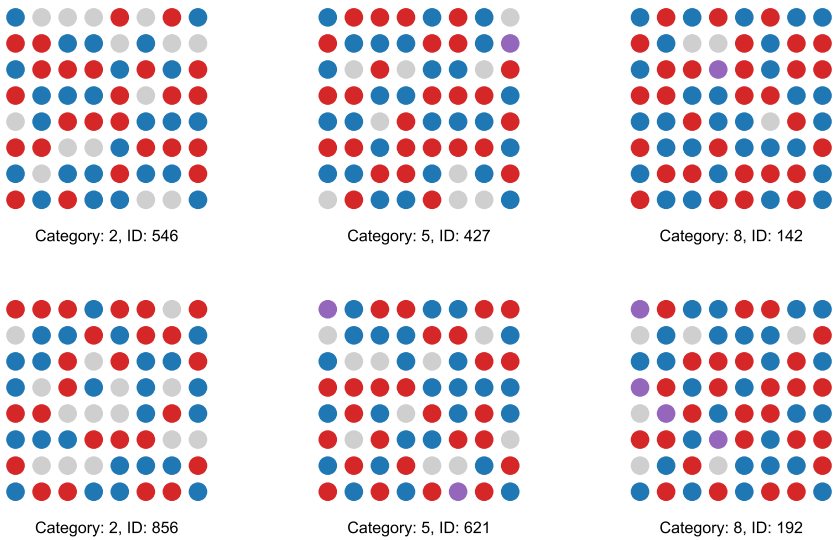


FIG. S3. Example snapshots from the 9-category XTRG dataset at the lowest temperature for the minimal Hubbard model on an 8×8 lattice. Each snapshot is represented as a 2D grid, where each site is color-coded according to its local state: empty (grey), spin-up (red), spin-down (blue), and doubly occupied (purple).

bypasses the block and adds the input directly to the block's output. Concretely, for an input vector $\boldsymbol{\Sigma}_i$, the output of a block with transformation function $\mathcal{F}(\boldsymbol{\Sigma}_i)$ is modified to

$$\boldsymbol{\Sigma}_i \leftarrow \boldsymbol{\Sigma}_i + \mathcal{F}(\boldsymbol{\Sigma}_i). \tag{S15}$$

In this formulation, even if the gradient through $\mathcal{F}$ becomes vanishingly small, the identity pathway preserves well-conditioned gradient flow, thereby facilitating effective backpropagation. In our architectural depiction (namely Fig. 2) in the main text, the `Residual` block and the `Add` in the `Add & Norm` block both represent this residual connection.

Layer Normalization. — *Layer normalization* (LayerNorm) [127] further stabilizes and accelerates training by normalizing the magnitude across the feature dimension per position. For an input vector $\boldsymbol{\Sigma}_i \in \mathbb{R}^{d_{\text{model}}}$, LayerNorm computes

$$\text{LayerNorm}(\boldsymbol{\Sigma}_i) = a \cdot \frac{\boldsymbol{\Sigma}_i - \text{mean}(\boldsymbol{\Sigma}_i)}{\text{std}(\boldsymbol{\Sigma}_i) + \epsilon} + b, \tag{S16}$$

where $\text{mean}(\boldsymbol{\Sigma}_i)$ and $\text{std}(\boldsymbol{\Sigma}_i)$ denote the mean and standard deviation of the elements of $\boldsymbol{\Sigma}_i$, respectively; a and b are learnable scale and shift parameters; and $\epsilon = 10^{-6}$ is a small constant that prevents division by zero. In our architectural diagram in the main text, the `Norm` in the `Add & Norm` block corresponds to this LayerNorm operation.

S-II. SPECIFICATIONS OF THE SNAPSHOT DATASET

Here, we report the temperature T, charge doping δ, and double occupancy $n_{\uparrow\downarrow}$ for the nine categories in the XTRG snapshot dataset of the minimal Hubbard model on an 8×8 lattice, as summarized in Table I. Several representative snapshots from each doping level at the lowest temperature are shown in Fig. S3.

Categories	T	δ	$n_{\uparrow\downarrow}$
Cat 0	1/4	0.2041	0.0123
Cat 1	1/16	0.2190	0.0151
Cat 2	1/256	0.2188	0.0156
Cat 3	1/4	0.1324	0.0149
Cat 4	1/16	0.1227	0.0186
Cat 5	1/256	0.1250	0.0182
Cat 6	1/4	0.0732	0.0179
Cat 7	1/16	0.0413	0.0228
Cat 8	1/256	0.0312	0.0220

TABLE I. Temperature T, charge doping δ, and double occupancy $n_{\uparrow\downarrow}$ of the nine categories in the XTRG snapshot dataset for the minimal Hubbard model.

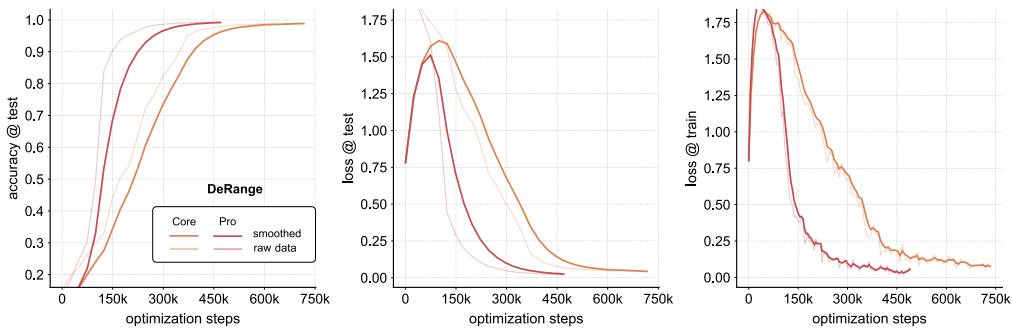


FIG. S4. Training profiles of the core and pro models for the derangement dataset. Metrics are displayed as raw data (thin lines with muted color) and with an exponential smoothing factor $\alpha = 0.4$ (thick lines with deep color). Both models achieve near-perfect training accuracy, demonstrating their capability to capture the imposed correlation structures, although the core model requires more optimization steps to converge. This indicates the efficacy of the semi-linear attention mechanism for modeling correlations in snapshot-type datasets.

S-III. BENCHMARK ON DERANGEMENTS

In this section, we benchmark our pro and core architectures on a synthetic dataset of derangements to demonstrate their ability to capture latent correlation structure. A derangement is a permutation in which no element remains in its original position. For example, given `[1, 2, 3]`, the derangements are `[2, 3, 1]` and `[3, 1, 2]`. We construct artificial 8×8 snapshots in which the left half (left four columns) is generated uniformly at random, while the right half (right four columns) is obtained by applying a (column-wise) derangement to the left half.

Categories	Derangements
Cat 0	Random
Cat 1	`[1,0,3,2]`
Cat 2	`[1,3,0,2]`
Cat 3	`[2,0,3,1]`
Cat 4	`[2,3,0,1]`
Cat 5	`[2,3,1,0]`
Cat 6	`[3,2,0,1]`

TABLE II. Derangements for the seven categories in the synthetic dataset. Numbers in the derangements denote the columns (not to be mistaken with tokens or local states). Category 0 contains snapshots generated completely randomly as a comparison baseline.

Table II enumerates the six derangements used in our dataset, together with category 0 comprising fully random snapshots as a comparative baseline. Each category contains 10,000 snapshots of size 8×8. Representative examples are shown in the upper panel of Fig. S7-(1-6). The derangement defining each category is indicated above the panel, and corresponding columns — i.e., columns that are identical by construction — are highlighted with matching background colors. In these synthetic snapshots, there is 100% correlation between corresponding columns across the left and right halves, and no correlations otherwise. The objective is to assess whether a trained model can assign snapshots to the correct category purely from these correlation patterns, i.e. whether it can correctly identify the permutation used to generate the right four columns from the left four random ones.

We train both the core and pro architectures on the derangement dataset without a locality bias; training profiles are summarized in Fig. S4. Both models attain near-perfect training accuracy, indicating successful identification of the imposed correlations, although the core model requires more optimization steps to converge. This observation further supports the efficacy of the semi-linear attention mechanism for modeling correlations in snapshot-type datasets.

Next, we examine the attention maps produced by the core model. Figs. S7-(1-6) display, for each derangement category, a snapshot together with the attention scores from the first attention layer. These visualizations reveal where the model *looks* when processing each position of the input. Each panel comprises an 8×8 array of subplots, each showing an 8×8 grid of cells. Within each subplot, the attention scores $\mathcal{A}_{ij}$ are encoded by a color scale; the query position i coincides with the subplot's location and is indicated by a red circle.

In general, we observe elevated attention at the same row of the query, and in most maps the dominant attention is devoted to the corresponding site (with an identical state by construction) in the opposite half of the snapshot, e.g. the attention maps at the 5th and 6th row, 1st column of Fig. S7-1 highlight the sites at the

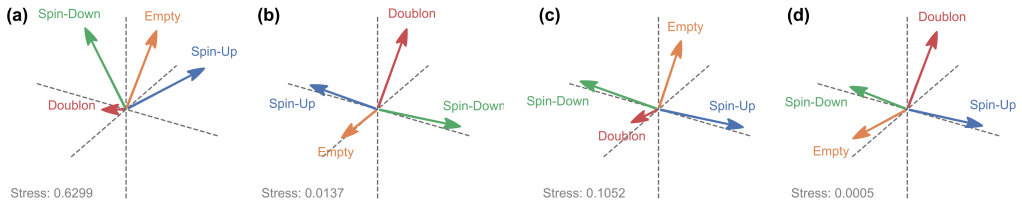


FIG. S5. Orthogonality relationships between the average embeddings under (a) mode 0 (no projection), (b) mode 1, (c) mode 2, and (d) mode 3 projection of the four local states (empty, spin-up, spin-down, and doubly occupied) from the core model for the XTRG Hubbard dataset. The angles between the original embeddings are close to 90°, a clear indication of the distinct local states; after the projections, the spin-up and spin-down states become nearly opposite in direction, reflecting the SU(2) rotational spin symmetry underlying the snapshots.

same row, 6th column. This behavior indicates that the model has internalized the per-row permutation structure and largely identified the strong cross-half correlations. The maps are not perfectly pristine — some spurious attention persists, e.g. the 2nd and 3rd rows in Fig. S7-1 — likely attributable to the correlations being sufficiently strong that high accuracy is achievable without completely disentangling all dependencies. Nevertheless, the attention visualizations collectively corroborate the model's capacity to recover the underlying correlation structure.

S-IV. ORTHOGONALITY AND ATTENTION MAPS

Orthogonality relations (angles) among the vectors $\boldsymbol{\Sigma}_i$ that represent tokens (local states) elucidate how the model internally encodes and discriminates between these states. Since the vectors $\boldsymbol{\Sigma}_i$ also depend on the position i, we define the *average embedding* $\boldsymbol{\Sigma}[\sigma]$ of token σ as

$$\boldsymbol{\Sigma}[\sigma] = \frac{1}{N_\sigma} \sum_{i \in \mathcal{I}_\sigma,\ \text{all } x} \boldsymbol{\Sigma}_i(x), \tag{S17}$$

where x denotes an input snapshot, $\boldsymbol{\Sigma}_i(x)$ the corresponding input embeddings, $\mathcal{I}_\sigma = \{i \mid \sigma_i = \sigma\}$ the set of positions at which token σ occurs, and $N_\sigma = |\mathcal{I}_\sigma|$ its multiplicity in the entire ensemble. The average embedding $\boldsymbol{\Sigma}[\sigma]$ is therefore the mean embedding vector of token σ aggregated over all of its occurrences in the dataset.

Also, for different propagation modes m in the attention stack, the input embeddings $\boldsymbol{\Sigma}_i$ are further projected according to mode-specific linear transformations $W_\mathcal{V}$ (see main text). For our core model with two attention layers (blocks), we identify four projectors $W^{(m)}$: $W^{(m=0)} = I$ (identity, no projection), $W^{(m=1)} = W_\mathcal{V}^{(\ell=1)}$ (layer 1 projection), $W^{(m=2)} = W_\mathcal{V}^{(\ell=2)}$ (layer 2 projection), and $W^{(m=3)} = W_\mathcal{V}^{(\ell=1)} W_\mathcal{V}^{(\ell=2)}$ (layer 1 & 2 projection). We then define the projected average embedding of token σ under mode m as

$$\boldsymbol{\Sigma}_m[\sigma] = \boldsymbol{\Sigma}[\sigma] W^{(m)}. \tag{S18}$$

In this regard, we can calculate the *overlap* (i.e the normalized inner product, and thus the cosine of their angle $\theta^{(m)}_{\sigma\sigma'}$) between the average embeddings of any pair of tokens σ and σ' for propagation mode m as

$$\cos\langle \boldsymbol{\Sigma}_m[\sigma], \boldsymbol{\Sigma}_m[\sigma'] \rangle = \frac{\boldsymbol{\Sigma}_m[\sigma] \cdot \boldsymbol{\Sigma}_m[\sigma']}{\|\boldsymbol{\Sigma}_m[\sigma]\| \cdot \|\boldsymbol{\Sigma}_m[\sigma']\|}. \tag{S19}$$

Detailed numerics are summarized in Table III.

To visualize these orthogonality relations for each mode m, we embed the four vectors $\boldsymbol{\Sigma}_m[\sigma]$, $\sigma = 0, 1, 2, 3$, into three-dimensional space via principal component analysis. Concretely, we construct the Gram matrix

$$G^{(m)}_{\sigma\sigma} = 1, \quad G^{(m)}_{\sigma\sigma'} = \cos\theta^{(m)}_{\sigma\sigma'} \quad \text{for} \quad \sigma \neq \sigma' \tag{S20}$$

and perform an eigen-decomposition $G^{(m)} = U\Lambda U^\top = XX^\top$ with $X = U\sqrt{\Lambda}$. The rows χ_r of X provide the coordinates of the visualization vectors. When rank $G^{(m)}$ exceeds three, we retain only the three largest eigenvalues and corresponding eigenvectors to obtain a three-dimensional approximation. The quality of this approximation is quantified by the *stress* metric

$$\text{stress} = \sqrt{\frac{\sum_{r<s} (G^{(m)}_{rs} - \chi_r^\top \chi_s)^2}{\sum_{r<s} (G^{(m)}_{rs})^2}}. \tag{S21}$$

Figure S5(a,b,c,d) visualizes the orthogonality relations for modes $m = 0, 1, 2, 3$, respectively. The angles between the original embeddings (mode 0) cluster near 90°, indicating that the model has learned to represent the four local states as nearly orthogonal vectors. The associated stress is large, consistent with the impossibility of embedding four almost mutually orthogonal vectors exactly in three dimensions.

After projection (i.e. modes 1, 2, and 3), the spin-up and spin-down embeddings become nearly antipodal, reflecting the un-

Token Pairs	Overlap	Angle (°)
(0, 1)	0.3121	71.82°
(0, 2)	0.2937	72.92°
(0, 3)	0.1858	79.29°
(1, 2)	0.1460	81.61°
(1, 3)	0.06737	86.14°
(2, 3)	0.1946	78.78°

(a) mode 0: original average embeddings

Token Pairs	Overlap	Angle (°)
(0, 1)	-0.2111	102.2°
(0, 2)	-0.1068	96.13°
(0, 3)	-0.5090	120.6°
(1, 2)	-0.9195	156.9°
(1, 3)	0.04032	87.69°
(2, 3)	0.1173	83.26°

(b) mode 1 projected average embeddings

Token Pairs	Overlap	Angle (°)
(0, 1)	0.1127	83.53°
(0, 2)	0.1275	82.67°
(0, 3)	0.2651	74.63°
(1, 2)	-0.7540	138.9°
(1, 3)	-0.2313	103.4°
(2, 3)	0.4466	63.48°

(c) mode 2 projected average embeddings

Token Pairs	Overlap	Angle (°)
(0, 1)	-0.1238	97.11°
(0, 2)	-0.1439	98.27°
(0, 3)	-0.6658	131.7°
(1, 2)	-0.9630	164.4°
(1, 3)	0.02786	88.40°
(2, 3)	0.1491	81.43°

(d) mode 3 projected average embeddings

TABLE III. Average inner product for each token-pair and corresponding angles (in degrees) for (a) mode 0: original embeddings (no projection), (b) mode 1, (c) mode 2, and (d) mode 3 projected average embeddings. All values are rounded to 4 significant digits.

derlying SU(2) spin-rotational symmetry of the snapshots. The low stress corroborates this symmetry-induced constraint, which effectively removes one independent basis state from the local Hilbert space. Moreover, the spinful states are broadly orthogonal to the plane spanned by the empty and doubly occupied states, capturing the distinction between sectors of different total spin.

These orthogonality relations collectively substantiate that the model faithfully captures the physical significance of the basis states in the local Hilbert space.

Beyond orthogonality, attention maps offer complementary insight into the model' s processing of snapshots. Figures S8-(1-3) present, for three representative snapshots from the XTRG Hubbard dataset, the attention rollout [152] of the core model. As before, each panel comprises an 8×8 array of subplots, each showing an 8×8 grid. Within each subplot, the rolled-out attention (with the identity component subtracted) $\mathcal{R}_{ij} - I_{ij}$ is encoded by a color scale; the query position i coincides with the subplot location and is marked with a red circle.

In contrast to the derangement benchmark, the attention maps for the Hubbard snapshots are substantially more challenging to comprehend. This is expected: correlations in the Hubbard data are far more intricate and less deterministic. A salient feature is produced by the locality bias, whereby attention concentrates on nearby sites. Another notable characteristic is that — unlike the derangement benchmark where attention typically condenses onto a few positions — the attention scores for the Hubbard snapshots are markedly more diffuse. This suggests that the Hubbard correlations can be high-order and spatially extended. Additional structures likely exist and remain to be elucidated through more refined analytical methods.

S-V. PREVIEW OF A 25-CATEGORY OMNIMETER

In this section, we present a technical preview of a 25-category *omnimeter* for the Hubbard model and demonstrate its advantage over the contemporary spin-correlation-based thermometer [85,106]. A comprehensive evaluation of this omnimeter and its applications will be detailed in a forthcoming technical report.

In the main text, we observed that the 9-category omnimeter can fail at doping levels absent from the training set. To address this limitation, we broaden the training coverage to five doping levels, $\delta \simeq 3\%, 7\%, 12\%, 17\%, 22\%$; and for each doping level, we consider five thermal exponents $n_T = -\log_2 T = 0, 2, 4, 6, 8$, yielding a total of 25 categories.

Current state-of-the-art thermometry for cold-atom Hubbard experiments relies on a direct comparison between measured spin correlations and calibrated values from numerical simulations [85, 106]. Concretely, one compares the spin correlations measured from a snapshot obtained from the quantum gas microscope and assigns a temperature based on the closest match to the calibrated correlations. However, this straight-forward approach becomes numerically unstable especially when the correlations saturate at low temperatures (see Table IV), leading to large fluctuations in the temperature estimates.

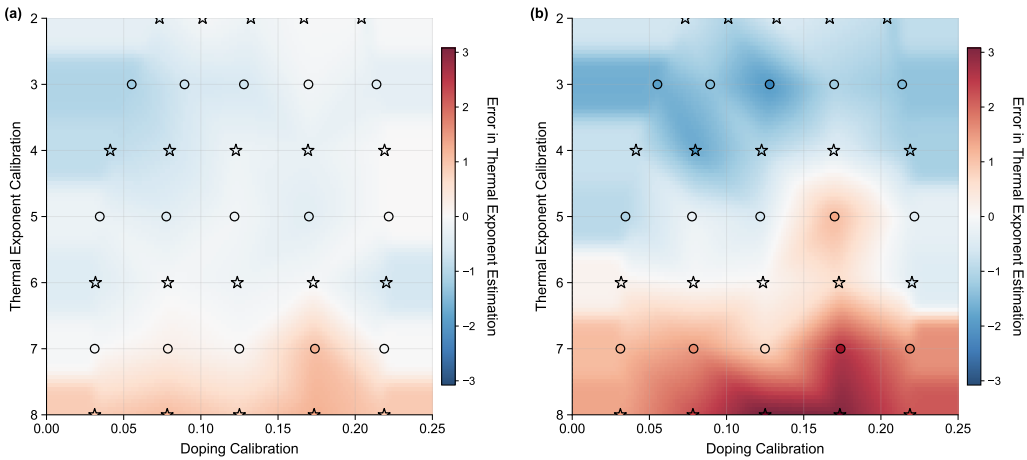


FIG. S6. Error of the thermal exponent estimate $n_T = -\log_2 T$ for (a) the 25-category omnimeter and (b) the spin-correlation-based thermometer, evaluated on a random 30-snapshot ensemble. Open circles/pentagrams mark the locations in phase space of the snapshot ensembles under evaluation, with pentagrams (circles) indicating data included (not included) in the training set. Color scales are obtained via interpolation and made equal in both panels for a direct comparison. The omnimeter consistently outperforms the spin-correlation thermometer in most regions of phase space, especially at lower temperatures where correlations saturate.

Therefore, instead of directly assigning the calibrated value, we adopt a probability-based formulation that delivers more stable estimates. Specifically, we postulate the probability (weight) of a snapshot x at thermal exponent n_T as

$$p(n_T|x) \propto 1/\|\Gamma^{zz}(x) - \Gamma^{zz}(n_T)\|, \tag{S22}$$

where $\Gamma^{zz}(x)$ and $\Gamma^{zz}(n_T)$ denote the nearest-neighbor spin-z correlations along the y (vertical) direction measured from snapshot x, and the calibrated values (via XTRG) at thermal exponent n_T, respectively. This empirical formula ensures that the the thermal exponent with the closer correlation value attains the higher probability, while still allocating non-zero weights to other categories to enhance stability. Other functional forms, e.g. exponential decay, can also be considered; however, alternative choices do not significantly affect the performance.

For an ensemble x of snapshots, the posterior $p(n_T|x)$ is obtained by averaging $p(n_T|x)$ over all x. The restriction to the y direction is required to match the 2D geometry to the tensor network structure (only neighboring sites along the y direction are guaranteed a bond directly connecting them) [79]. While one can construct composite estimators that fuse multiple correlation messengers, in practice these do not surpass the stability or accuracy of the single-messenger formulation.

We evaluate the performance of the 25-category omnimeter and the spin-correlation-based thermometer on random ensembles of 30 snapshots for each location in phase space; results are summarized in Fig. S6. As before, open circles/pentagrams mark the locations in phase space of the snapshot ensembles under evaluation, with pentagrams (circles) indicating data included (not included) in the training set. Color scales are obtained via interpolation and made equal for both panels to enable a direct, like-for-like comparison.

The AI omnimeter consistently outperforms the thermometer based on spin correlations across most of phase space, with a pronounced advantage at lower temperatures where spin correlations begin to saturate. To rationalize this behavior, Table IV reports the reference spin-z correlations Γ^{zz} together with their standard deviations std(Γ^{zz}) at the calibrated locations. At low temperatures (large n_T), the standard deviations exceed the separation between adjacent temperature categories, implying that nearest neighbor spin correlations alone cannot reliably discriminate fine temperature increments (e.g. in Table IV(a), the standard deviations std(Γ^{zz}) for $n_T = 6$ and 8 are around 0.018, whereas the difference between Γ^{zz} is only ≈ 0.001). By contrast, the omnimeter automatically exploits a broader spectrum of correlation features beyond Γ^{zz}, enabling substantially more accurate temperature estimation.

(a) $\mu = 1.2,\ \delta \approx 22\%$

n_T	Γ^{zz}	std(Γ^{zz})
0.0	-0.01081	0.02386
2.0	-0.03751	0.02319
4.0	-0.06626	0.02012
6.0	-0.07412	0.01853
8.0	-0.07559	0.01867

(b) $\mu = 1.4,\ \delta \approx 17\%$

n_T	Γ^{zz}	std(Γ^{zz})
0.0	-0.01126	0.02494
2.0	-0.04096	0.02541
4.0	-0.07556	0.02136
6.0	-0.08089	0.02037
8.0	-0.08156	0.02026

(c) $\mu = 1.6,\ \delta \approx 12\%$

n_T	Γ^{zz}	std(Γ^{zz})
0.0	-0.01281	0.02607
2.0	-0.04639	0.02623
4.0	-0.08709	0.02200
6.0	-0.09164	0.02141
8.0	-0.09061	0.02056

(d) $\mu = 1.8,\ \delta \approx 7\%$

n_T	Γ^{zz}	std(Γ^{zz})
0.0	-0.01377	0.02611
2.0	-0.05229	0.02630
4.0	-0.09902	0.02453
6.0	-0.1053	0.02413
8.0	-0.1052	0.02321

(e) $\mu = 2.0,\ \delta \approx 3\%$

n_T	Γ^{zz}	std(Γ^{zz})
0.0	-0.01381	0.02711
2.0	-0.05450	0.02621
4.0	-0.1078	0.02696
6.0	-0.1169	0.02705
8.0	-0.1192	0.02641

TABLE IV. The reference spin-z correlations Γ^{zz} and the corresponding standard deviations, std(Γ^{zz}), for different chemical potentials μ (and thus doping δ) across thermal exponents n_T. All correlation values and standard deviations are computed over all snapshots in each category, and rounded to 4 significant digits.

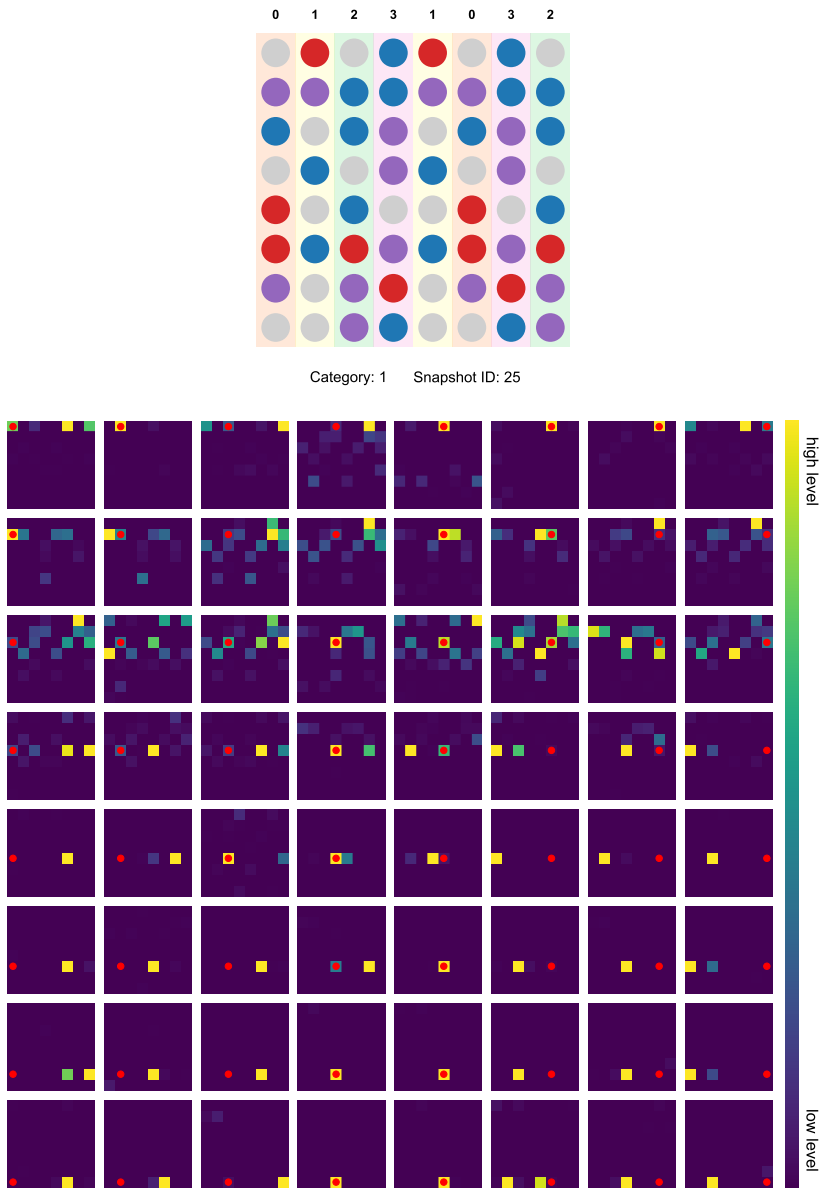


FIG. 7-1. Attention map (bottom) of the core model evaluated on a snapshot (top) from category 1 (`[1,0,3,2]`). The visualization consists of an 8×8 array of subplots, each displaying an 8×8 grid of cells. In each subplot, attention scores $\mathcal{A}_{ij}$ are encoded by a color scale; the query position i coincides with the subplot's position and is marked by a red circle. The prominent highlighting of the attention map indicates that the model has successfully captured the underlying correlations between corresponding columns in the left and right halves.

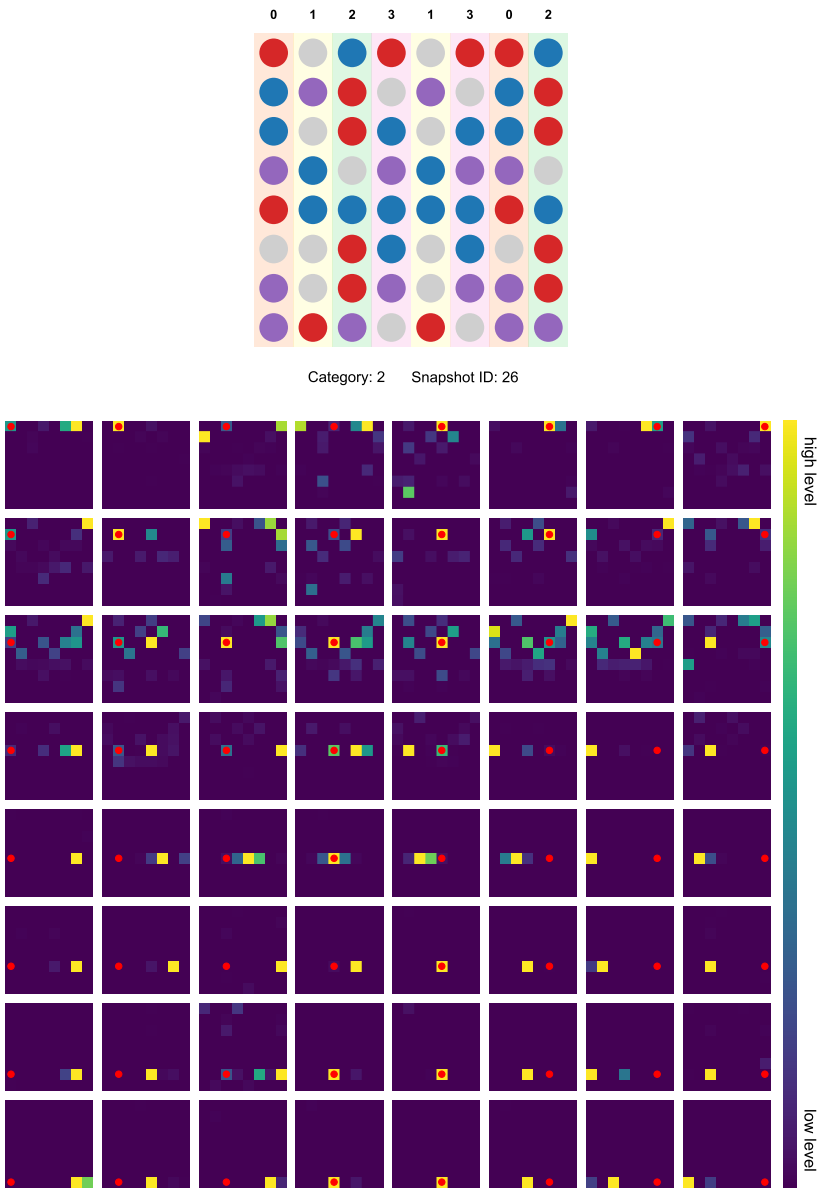


FIG. 7-2. Same as Fig. S7-1, now for a snapshot from category 2 (`[1,3,0,2]`). Same as Fig. S7-1, the visualization consists of an 8×8 array of subplots, each displaying an 8×8 grid of cells. In each subplot, attention scores $\mathcal{A}_{ij}$ are encoded by a color scale; the query position i coincides with the subplot's position and is marked by a red circle. The prominent highlighting of the attention map indicates that the model has successfully captured the underlying correlations between corresponding columns in the left and right halves.

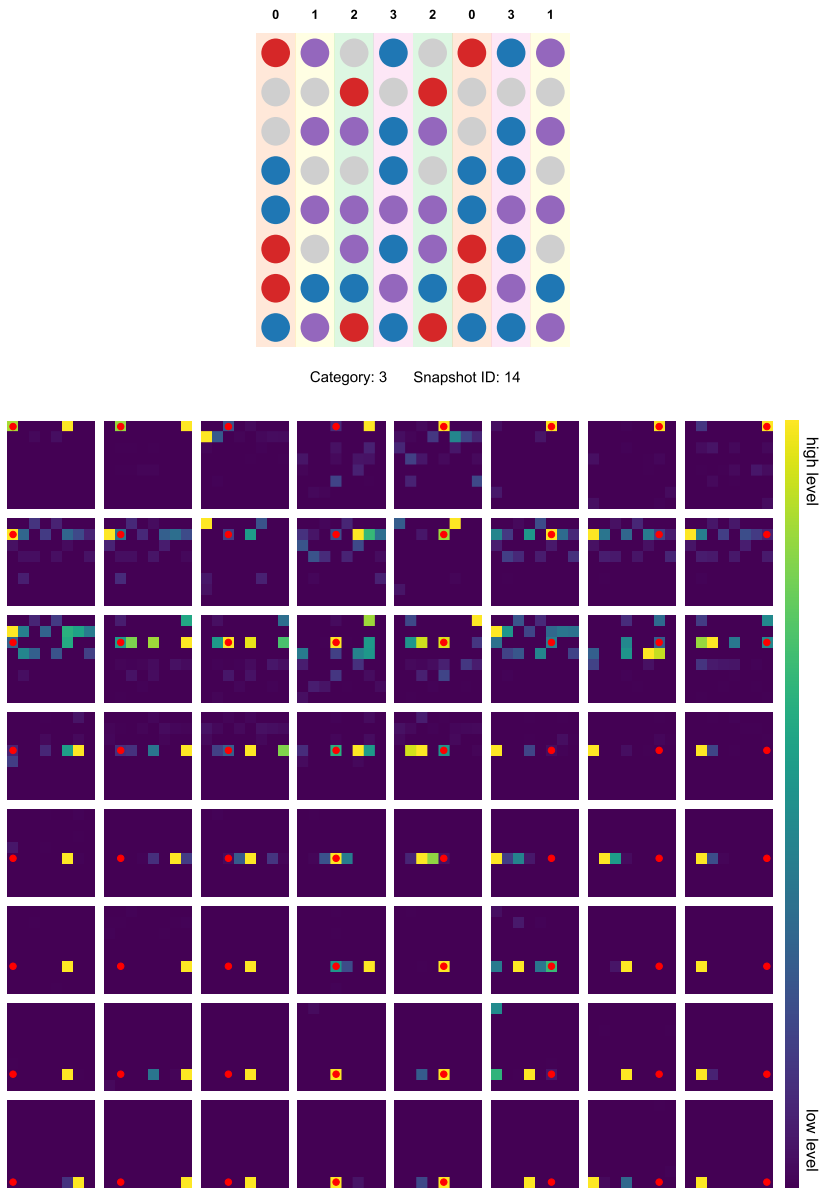


FIG. 7-3. Same as Fig. S7-1, now for a snapshot from category 3 (`[2,0,3,1]`). Same as Fig. S7-1, the visualization consists of an 8×8 array of subplots, each displaying an 8×8 grid of cells. In each subplot, attention scores $\mathcal{A}_{ij}$ are encoded by a color scale; the query position i coincides with the subplot's position and is marked by a red circle. The prominent highlighting of the attention map indicates that the model has successfully captured the underlying correlations between corresponding columns in the left and right halves.

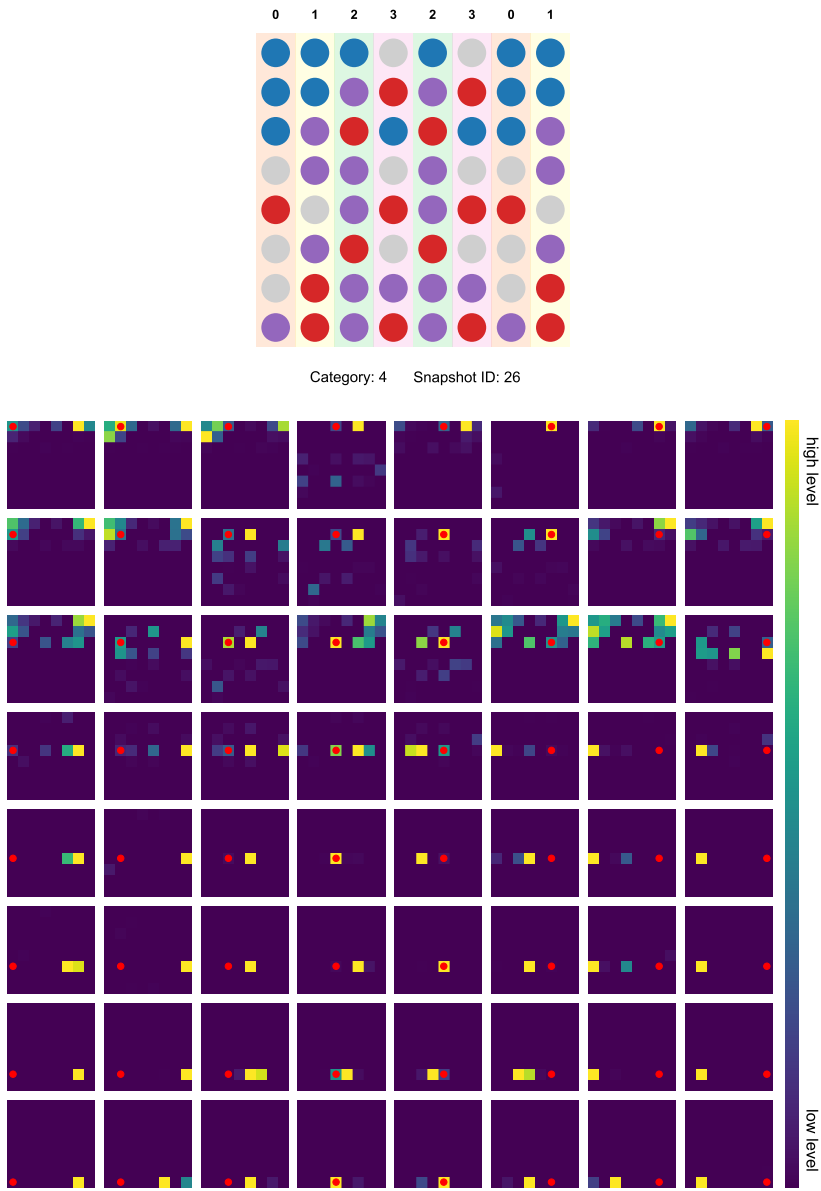


FIG. 7-4. Same as Fig. S7-1, now for a snapshot from category 4 (`[2,3,0,1]`). Same as Fig. S7-1, the visualization consists of an 8×8 array of subplots, each displaying an 8×8 grid of cells. In each subplot, attention scores $\mathcal{A}_{ij}$ are encoded by a color scale; the query position i coincides with the subplot's position and is marked by a red circle. The prominent highlighting of the attention map indicates that the model has successfully captured the underlying correlations between corresponding columns in the left and right halves.

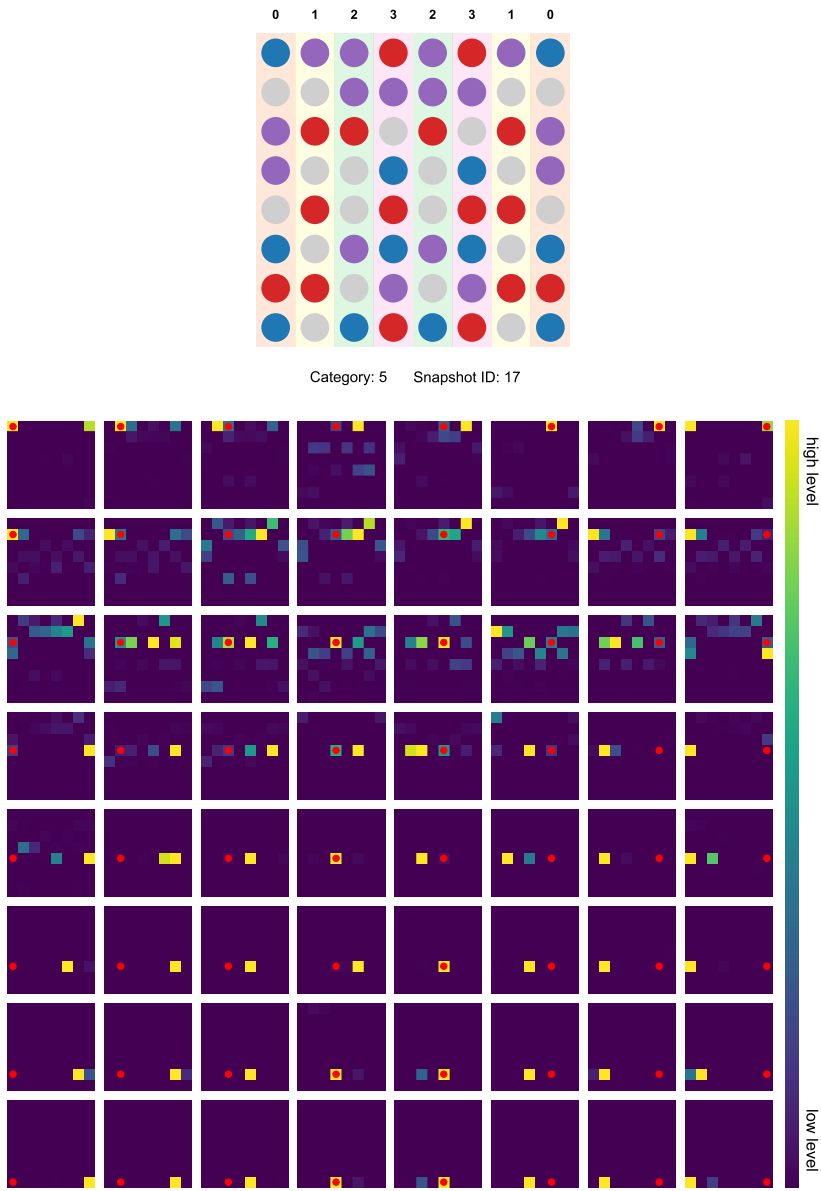


FIG. 7-5. Same as Fig. S7-1, now for a snapshot from category 5 (`[2,3,1,0]`). Same as Fig. S7-1, the visualization consists of an 8×8 array of subplots, each displaying an 8×8 grid of cells. In each subplot, attention scores $\mathcal{A}_{ij}$ are encoded by a color scale; the query position i coincides with the subplot's position and is marked by a red circle. The prominent highlighting of the attention map indicates that the model has successfully captured the underlying correlations between corresponding columns in the left and right halves.

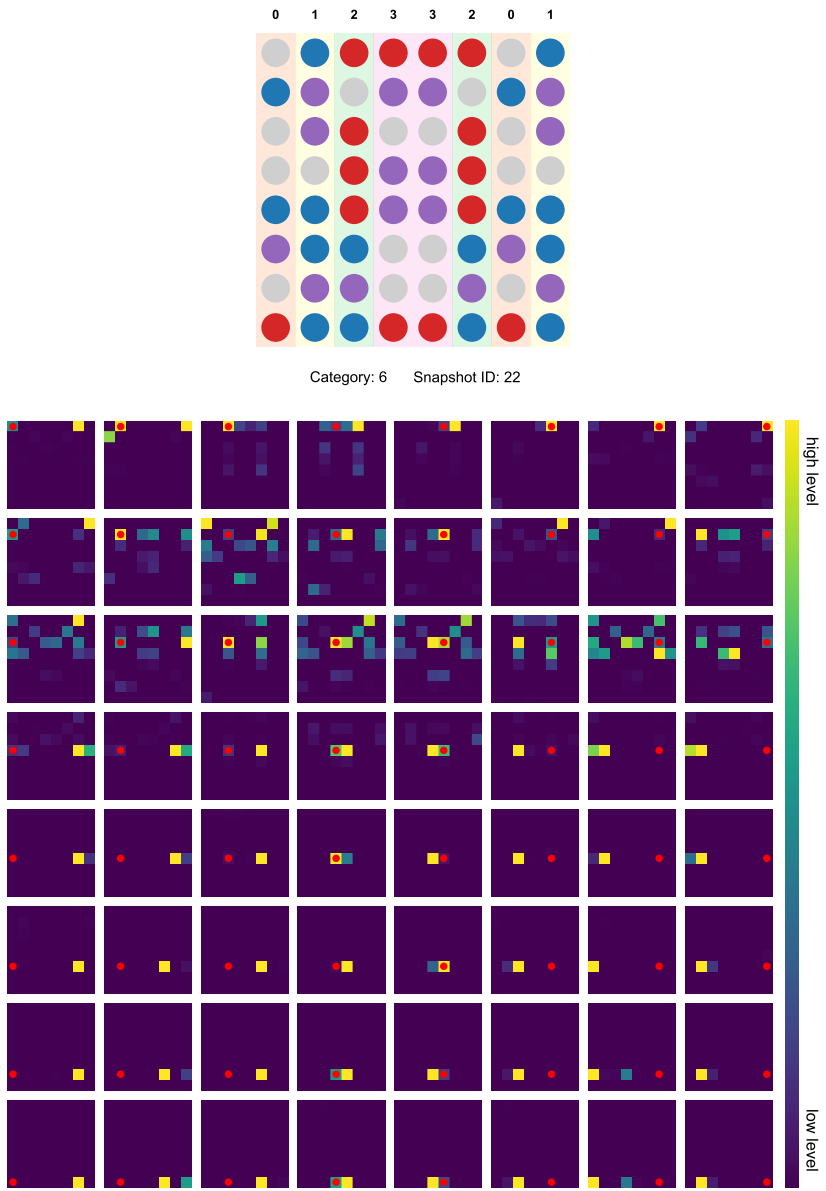


FIG. 7-6. Same as Fig. S7-1, now for a snapshot from category 6 (`[3,2,0,1]`). Same as Fig. S7-1, the visualization consists of an 8×8 array of subplots, each displaying an 8×8 grid of cells. In each subplot, attention scores $\mathcal{A}_{ij}$ are encoded by a color scale; the query position i coincides with the subplot's position and is marked by a red circle. The prominent highlighting of the attention map indicates that the model has successfully captured the underlying correlations between corresponding columns in the left and right halves.

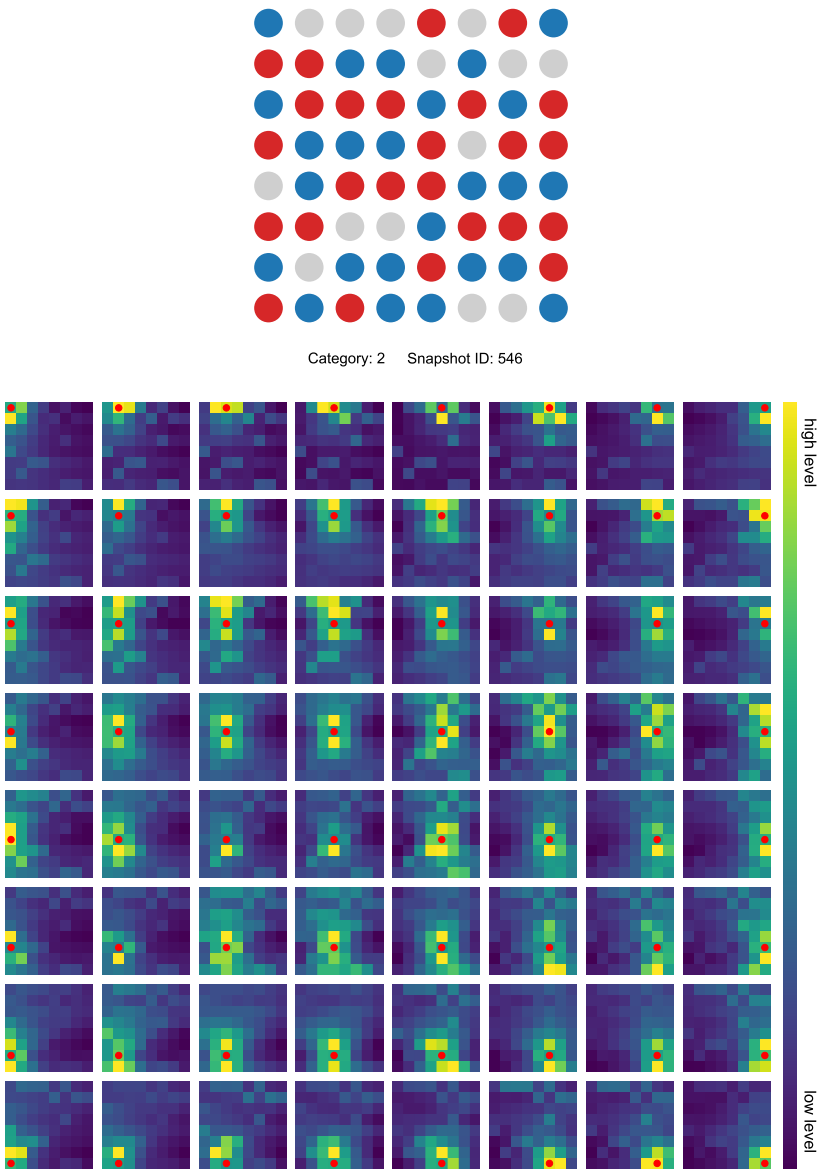


FIG. 8-1. Attention map (bottom) of the core model evaluated on a snapshot (top) at low temperature and *over-doped* region. The visualization consists of an 8×8 array of subplots, each displaying an 8×8 grid of cells. In each subplot, attention scores $\mathcal{A}_{ij}$ are encoded by a color scale; the query position i coincides with the subplot's position and is marked by a red circle.

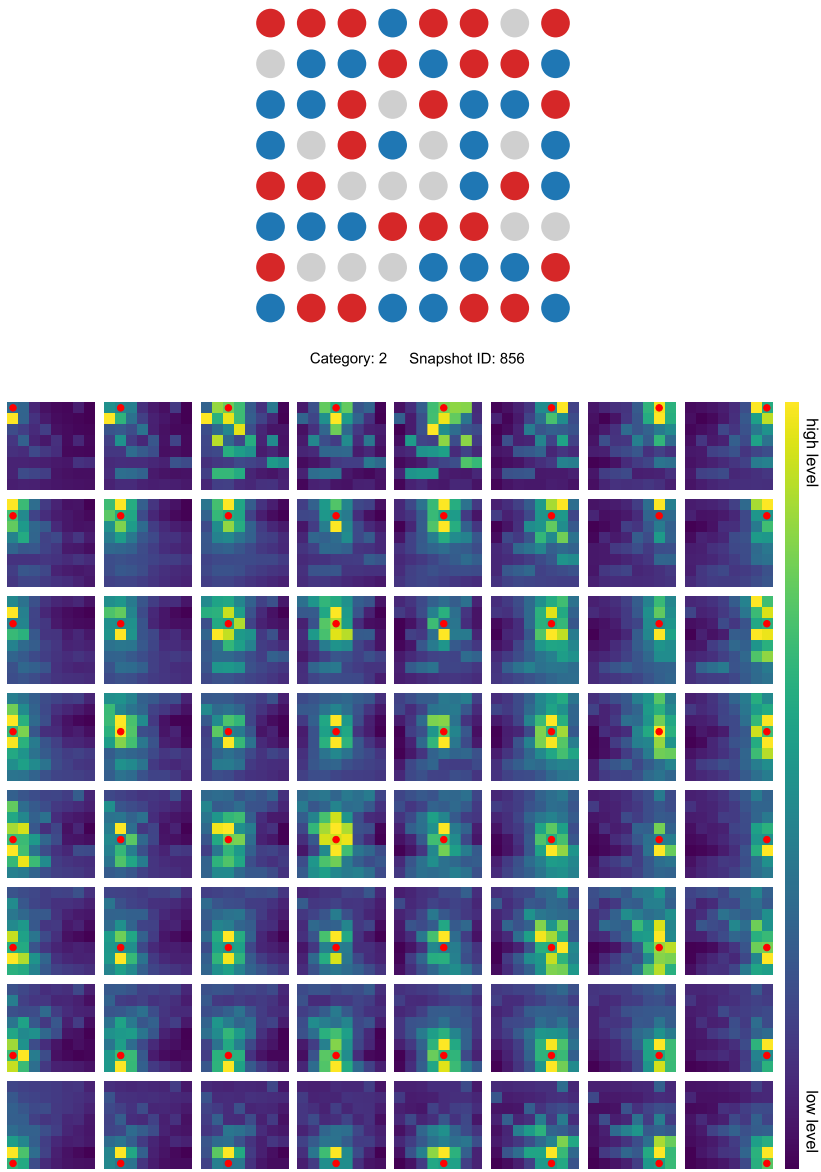


FIG. 8-2. Attention map (bottom) of the core model evaluated on a snapshot (top) at low temperature and *over-doped* region. Same as Fig. S8-1, the visualization consists of an 8×8 array of subplots, each displaying an 8×8 grid of cells. In each subplot, attention scores $\mathcal{A}_{ij}$ are encoded by a color scale; the query position i coincides with the subplot's position and is marked by a red circle.

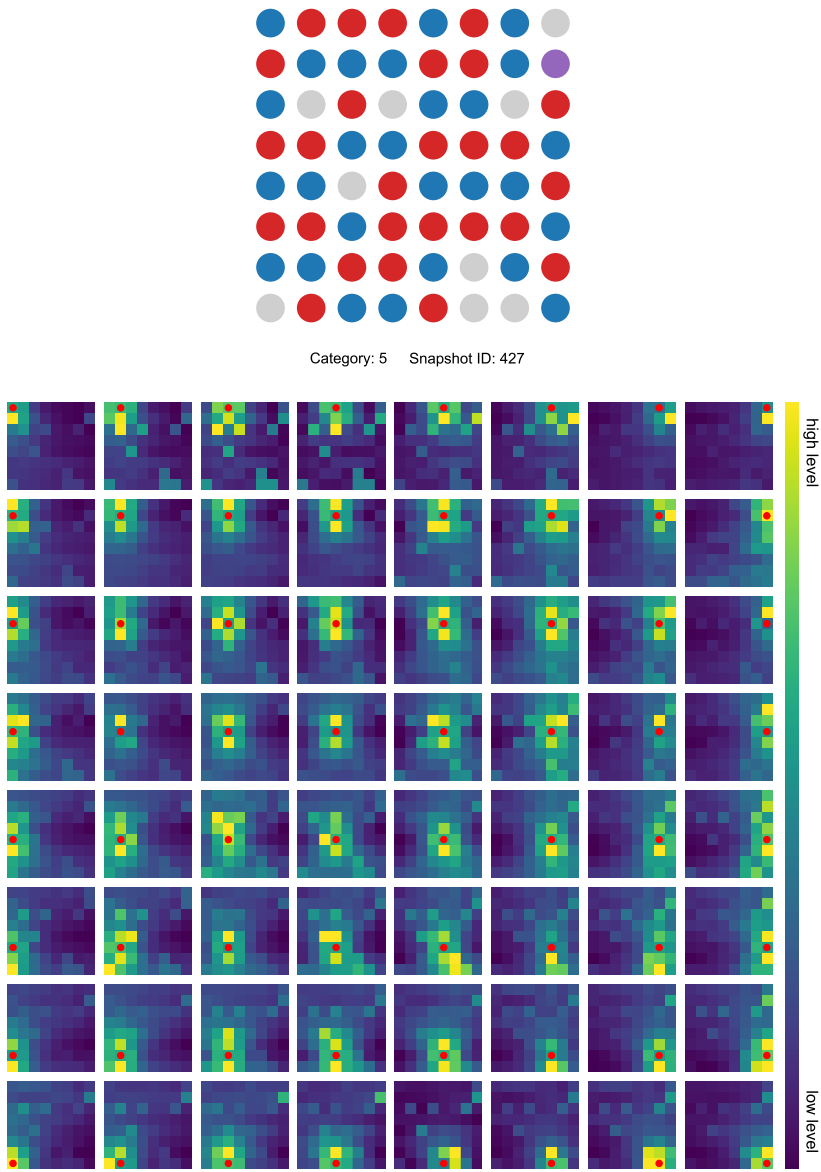


FIG. 8-3. Attention map (bottom) of the core model evaluated on a snapshot (top) at low temperature and *medium-doped* region. Same as Fig. S8-1, the visualization consists of an 8×8 array of subplots, each displaying an 8×8 grid of cells. In each subplot, attention scores $\mathcal{A}_{ij}$ are encoded by a color scale; the query position i coincides with the subplot's position and is marked by a red circle.

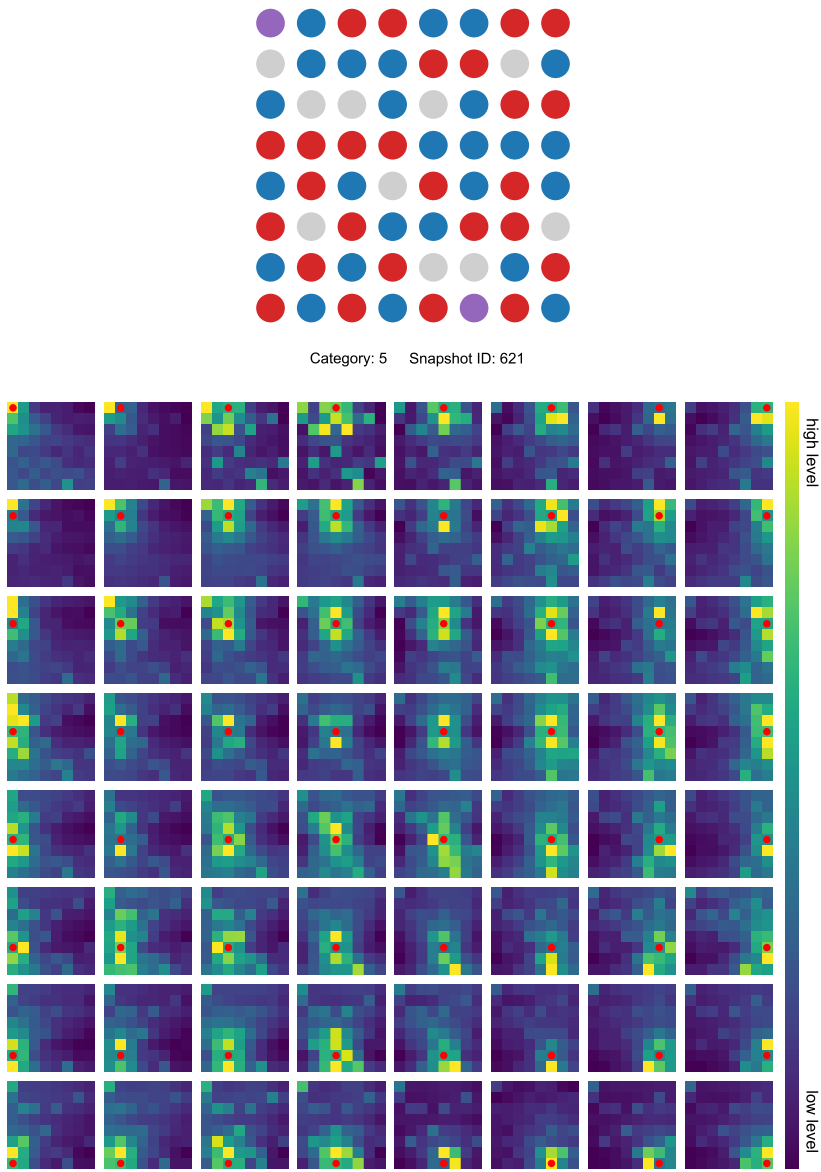


FIG. 8-4. Attention map (bottom) of the core model evaluated on a snapshot (top) at low temperature and *medium-doped* region. Same as Fig. S8-1, the visualization consists of an 8×8 array of subplots, each displaying an 8×8 grid of cells. In each subplot, attention scores $\mathcal{A}_{ij}$ are encoded by a color scale; the query position i coincides with the subplot's position and is marked by a red circle.

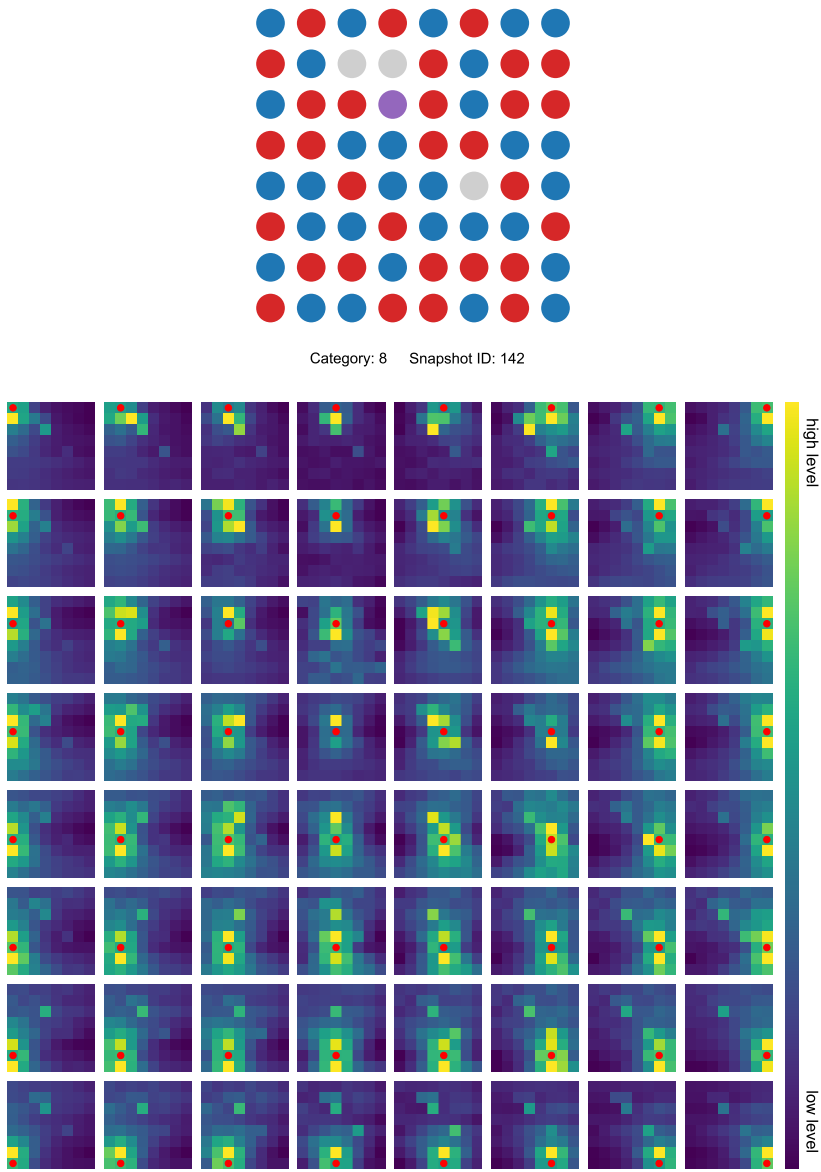


FIG. 8-5. Attention map (bottom) of the core model evaluated on a snapshot (top) at low temperature and *under-doped* region. Same as Fig. S8-1, the visualization consists of an 8×8 array of subplots, each displaying an 8×8 grid of cells. In each subplot, attention scores $\mathcal{A}_{ij}$ are encoded by a color scale; the query position i coincides with the subplot's position and is marked by a red circle.

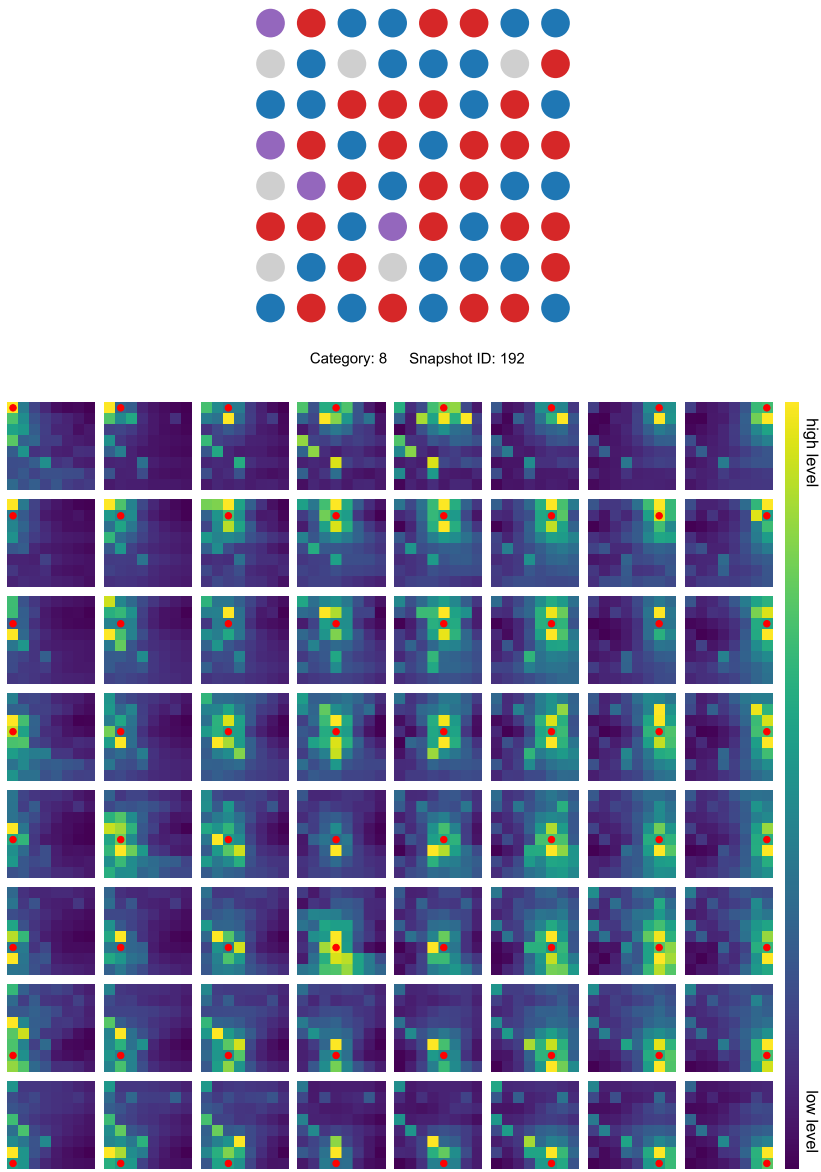


FIG. 8-6. Attention map (bottom) of the core model evaluated on a snapshot (top) at low temperature and *under-doped* region. Same as Fig. S8-1, the visualization consists of an 8×8 array of subplots, each displaying an 8×8 grid of cells. In each subplot, attention scores $\mathcal{A}_{ij}$ are encoded by a color scale; the query position i coincides with the subplot's position and is marked by a red circle.

Chapter 4
Conclusions and Outlook

Motivated by the open questions surrounding the exotic phenomena and strong correlations in cuprate superconductors, we develop and benchmark a comprehensive numerical toolbox for the 2D Hubbard model, spanning zero and finite temperatures and complemented by an AI snapshot analyzer. The framework integrates symmetry-aware ground-state iPEPS tensor networks with an accelerated finite-temperature XTRG into a coherent pipeline of many-body solvers, yielding controlled assessments of key order parameters across temperature and doping. Built atop this solver stack, we employ an interpretable, attention-based analyzer designed for system-level snapshots from both our simulations and quantum-gas microscopy; it learns the salient correlation patterns across temperature and doping and serves as a universal omnimeter, delivering highly accurate estimates of observables for ultracold atom quantum simulators. Taken together, the toolbox advances a coherent physical narrative of intertwined orders and furnishes a methodology that is portable to extended Hubbard, t-J, multi-band, and other well-defined quantum lattice models, while readily interfacing with snapshot-based experimental platforms.

Starting from the physical motivation and lattice modeling, we review how strong on-site repulsion gives rise to Mott physics and antiferromagnetism, thereby positioning the 2D Hubbard model on the square lattice as a minimal low-energy description relevant to high-T_c cuprates. We explain the reduction from multi-band formulations to an effective one-band Hamiltonian and emphasize the role of particle-hole asymmetry induced by NNN hopping t'. In conjunction with charge doping, this t' term serves as a tunable source of magnetic frustration that reshapes the competition between stripe order and d-wave pairing. We further survey zero- and finite-temperature numerical approaches at a conceptual level — QMC, DMRG, PEPS, and their finite-T counterparts — highlighting complementary strengths and limitations, before introducing the emergence of transformer-based AI as an analysis paradigm capable of reading snapshot data and recognizing correlation patterns without prior knowledge of any specific order parameters.

On the methodological side, we consolidate the formal machinery and computational infrastructure. We specify the Hubbard Hamiltonian, analyze particle-hole symmetry on the square lattice, and delineate notable derivative models. On the TN

side, we develop the tensorial building blocks for MPS, MPO, and PEPS, including fermionic and SU(N) symmetry implementation, canonical forms, PCA diagnostics, LRA techniques, environment contraction via CTMRG, and snapshot sampling for iPEPS. The algorithmic sections detail ground-state optimization (DMRG and iPEPS updates), the exponential-cooling philosophy of XTRG with variational MPO sum/product handling, and practical advances that improve accuracy without compromising efficiency. Afterwards, we set out design principles for AI systems in scientific workflows, elaborate encoder-only, decoder-only, and encoder-decoder variants for discrimination, autoregressive and context-conditioned generation, and examine ongoing debates concerning the interpretability and generalization of the attention mechanism.

Three capstone studies then operationalize this toolbox along a single throughline. A symmetry-preserved iPEPS analysis at $T=0$ performs a controlled comparison between stripe-admitting and stripe-forbidden sectors, demonstrating that increasing magnetic frustration — especially for positive t' — systematically impairs stripe formation and stabilizes a uniform state with a robust d-wave singlet pairing. A finite-temperature campaign using an enhanced 1s$^+$ XTRG scheme achieves rapid cooling, connecting thermal observations to ground-state benchmarks, revealing the strengthening of pairing tendencies at low T and the emergence of pseudogap behavior. Finally, an interpretable AI analyzer is trained to identify salient correlation patterns in snapshot data, facilitating a universal *omnimeter* for ultracold-atom quantum simulators. Collectively, these researches instantiate a unified, extensible workflow that proceeds from Hamiltonian specification through ground-state and thermal characterization, and to AI-assisted inference, in a manner readily transferable to other strongly correlated many-body lattice systems.

Implications of t' in the Hubbard Model

The significance of the NNN hopping terms (t') in the Hubbard model gained traction with the consensus of the multi-messenger handshake showing no substantial SC orders in the minimal Hubbard model. Therefore, anticipation is vested in the t' term for its improved real-world relevance and the explicit breaking of particle-hole symmetry. First-principles electronic structure calculations (DFT/LDA) [302, 303] determine the Cu-O antibonding-band dispersion and thereby differentiate electron-like from hole-like band topologies across cuprate families; systematic downfolding of this dispersion onto a minimal tight-binding representation fixes the sign and magnitude of t'. The resulting parametrization implies $t' < 0$ for systems whose low-energy physics is hole-doped and $t'>0$ for electron-doped compounds.

Yet large-scale numerical studies paint a different picture. Most state-of-the-art simulations concur that negative t' suppresses SC orders, whereas positive t' enhances them [48, 56, 158, 159, 372, 373] — a trend at odds with experiments, where hole-doped cuprates display higher T_c and more robust SC than their electron-doped counterparts. Beyond pairing, Section 3.1 shows that AFM order spans a broader doping range for $t'<0$ than for $t'>0$, with SC emerging only at overdoped regimes,

again reversing the experimental pattern in which electron-doped materials exhibit the wider AFM dome. Altogether, these observations imply that, to align with the phenomenology of order parameters, one should effectively assign $t' > 0$ ($t' < 0$) to hole-doped (electron-doped) systems.

This reasoning does not warrant abandoning the established first-principles determination of t' from band structure. Rather, it suggests that band-structure considerations are only weakly predictive of the mechanisms that trigger superconductivity. Additional factors — such as interference among hopping pathways mediated by oxygen orbitals or the magnetic frustration induced by t' — may exert a more pivotal influence on pair formation and the global phase diagram, while leaving the gross band structure largely intact. This perspective furnishes a guiding principle for future work regarding the decisive underlying mechanism of high-T_c superconductivity.

Multi-messenger Strategy in Solving Quantum Many-Body Systems

The intricate nature of strongly correlated quantum many-body systems often precludes a single numerical technique from delivering a comprehensive solution. Each computational approach carries inherent strengths and limitations, dictated by its foundational assumptions and algorithmic design. In the 2D Hubbard model, for instance, QMC methods excel around half-filling but grapple with severe sign problems upon charge doping; DMRG attains high accuracy for ground-state properties in quasi-1D geometries yet faces acute scaling challenges in genuine 2D; and PEPS offers a natural representation for 2D systems and efficiently captures area-law entanglement, but is constrained by the accuracy of environment contraction and compression schemes. A similar situation also arises for finite-temperature simulations: METTS, finite-T PEPS, XTRG, and tanTRG each provide distinct advantages, accompanied by equally distinct trade-offs.

Despite these well-known facts, the computational community often fails to respect the complementarity of methods, privileging a single approach and striving to push it beyond competing alternatives. Parallel patterns also occur in the AI community, where researchers and engineers have often been overly obsessed with achieving the so-called state-of-the-art (SOTA) performance on evaluation scores, while underweighting the systematic biases and limitations inherent to the selected architectures and training corpora.

Such tendencies risk overreliance on single-source/method/ansatz evidence and the concomitant loss of insights gleaned from alternative perspectives. Illustrative, and in hindsight cautionary, is the report of spurious SC orders in $t'/t < 0$ via DMRG on an overly narrow width-4 cylinder [156]; this claim was subsequently challenged and refuted by wider-cylinder DMRG studies [159, 372–375] and by iPEPS analyses [48]. A naive argument might suggest that, despite geometric mismatch, DMRG should recover the essential physics given sufficiently large bond dimension. In practice, however, the stringent entanglement constraints imposed by narrow cylinders can lead to even *qualitatively* incorrect conclusions relative to the 2D thermodynamic limit.

These experiences collectively motivate a *multi-messenger* strategy in which diverse numerical methods are deployed in concert to interrogate the same physical problem. By leveraging the complementary strengths of each approach and cross-validating their outcomes, one can arrive at a more reliable and nuanced understanding of complex quantum phenomena. This philosophy underpins our integrated workflow, which combines ground-state iPEPS, finite-T XTRG, and AI-assisted snapshot analysis to deliver a holistic and self-consistent perspective on the 2D Hubbard model.

Future of the Autoregressive NQS Ansatz

Neural Quantum States (NQS) [179, 376, 377] have demonstrated remarkable efficacy in tackling challenging quantum many-body problems. Among recent advances, the Neural Transformer Backflow (NTB) [378–380] ansatz and autoregressive NQS architectures [180, 182, 183, 381, 382] have achieved state-of-the-art performance on fermionic lattice systems. Notably, the NTB ansatz has shown strong competency in the notoriously difficult regimes of the Hubbard model [84]. However, certain qualitative features reported by NTB are not fully consistent with those obtained from TN simulations — plausibly due to autocorrelations inherent to Markov-chain Monte Carlo (MCMC) sampling.

Autoregressive neural networks circumvent this bottleneck by directly parameterizing the overlap between a quantum state and the many-body Fock basis. Concretely, a quantum state $|\psi\rangle$ may be expanded using the probability $p(\vec{\sigma})$ and phase $\phi(\vec{\sigma})$ as

$$|\psi\rangle = \sum_{\vec{\sigma}} \sqrt{p(\vec{\sigma})} e^{i\phi(\vec{\sigma})} |\vec{\sigma}\rangle, \tag{4.1}$$

where $\vec{\sigma} = \{\sigma_1, \sigma_2, \ldots, \sigma_L\}$ denotes a many-body configuration of a length-L lattice system. The autoregressive NQS factorizes the joint probability distribution into a product of conditional probabilities:

$$p(\vec{\sigma}) = \prod_{i=1}^{L} p(\sigma_i | \sigma_{<i}), \tag{4.2}$$

thereby yielding a normalized distribution amenable to unbiased ancestral sampling. A practical complication is the additional phase factor $\phi(\vec{\sigma})$, which increases architectural complexity and complicates training within the Variational Monte Carlo (VMC) framework. A viable remedy is to introduce a pre-training stage [183, 326] using a curated set of probable configurations, e.g., sourced from experiments or high-fidelity simulations. Empirically, such pre-training markedly improves optimization stability and convergence. Consequently, constructing a high-quality dataset for pre-training the autoregressive NQS is of central importance at the current stage.

A comprehensive TNS scan of the target Hamiltonian provides precisely such a high-quality dataset. More specifically, one can perform sequential sampling from reduced density matrices obtained via TNS, while logging the diagonal elements to supply ground-truth conditional probabilities $p(\sigma_j | \sigma_{<j})$. In parallel, approximate overlaps $\langle \sigma | \psi \rangle$ can be extracted by projecting the TN state onto basis configurations

$|\sigma\rangle$, which in turn constrains the phase $\phi(\sigma)$. With a suitable gauge choice, the continuous phase may effectively collapse into two discrete values, 0 or π (equivalently, signs ± 1). Consequently, highly precise overlap magnitudes are unnecessary for establishing reliable phase labels. These robust targets enable effective pre-training of the autoregressive NQS and are expected to substantially accelerate and stabilize subsequent VMC fine-tuning.

Optionally, one can further employ an encoder-decoder architecture [383] in which the encoder accepts system parameters (e.g., chemical potential, on-site interaction, NNN hopping, etc.), and the decoder performs autoregressive generation conditioned on the encoded context. This design aims to train a single, unified model capable of representing quantum states across a continuum of parameter settings. Unlike TN methods, which typically require re-optimization in their entirety when parameters change, the encoder-decoder NQS only necessitates updating the encoder (potentially with an additional LoRA). This approach promises a significant reduction in the computational effort required to chart complex many-body phase spaces.

References

[1] D. Vaknin, S. K. Sinha, D. E. Moncton, D. C. Johnston, J. M. Newsam, C. R. Safinya, and H. E. King, *Antiferromagnetism in La_2CuO_{4-y}*, Phys. Rev. Lett., **58**, 2802 (1987).

[2] R. J. Birgeneau, D. R. Gabbe, H. P. Jenssen, M. A. Kastner, P. J. Picone, T. R. Thurston, G. Shirane, Y. Endoh, M. Sato, K. Yamada, Y. Hidaka, M. Oda, Y. Enomoto, M. Suzuki, and T. Murakami, *Antiferromagnetic spin correlations in insulating, metallic, and superconducting $La_{2-x}Sr_xCuO_4$*, Phys. Rev. B, **38**, 6614 (1988).

[3] S. Sugai, M. Sato, T. Kobayashi, J. Akimitsu, T. Ito, H. Takagi, S. Uchida, S. Hosoya, T. Kajitani, and T. Fukuda, *High-energy spin excitations in the insulating phases of high-T_c superconducting cuprates and La_2NiO_4*, Phys. Rev. B, **42**, 1045 (1990).

[4] A. Damascelli, Z. Hussain, and Z.-X. Shen, *Angle-resolved photoemission studies of the cuprate superconductors*, Rev. Mod. Phys., **75**, 473 (2003).

[5] P. K. Mang, O. P. Vajk, A. Arvanitaki, J. W. Lynn, and M. Greven, *Spin Correlations and Magnetic Order in Nonsuperconducting $Nd_{2-x}Ce_xCuO_{4\pm\delta}$*, Phys. Rev. Lett., **93**, 27002 (2004).

[6] E. M. Motoyama, G. Yu, I. M. Vishik, O. P. Vajk, P. K. Mang, and M. Greven, *Spin correlations in the electron-doped high-transition-temperature superconductor $Nd_{2-x}Ce_xCuO_{4\pm\delta}$*, Nature, **445**, 186 (2007).

[7] J. Friedel, *Quasilowdimensionality in the weak coupling limit*, Physica C, **153–155**, 1610 (1988).

[8] H. Alloul, T. Ohno, and P. Mendels, *^{89}Y NMR evidence for a fermi-liquid behavior in $YBa_2Cu_3O_{6+x}$*, Phys. Rev. Lett., **63**, 1700 (1989).

[9] T. Timusk and B. Statt, *The pseudogap in high-temperature superconductors: an experimental survey*, Rep. Prog. Phys., **62**, 61 (1999).

[10] C. M. Varma, P. B. Littlewood, S. Schmitt-Rink, E. Abrahams, and A. E. Ruckenstein, *Phenomenology of the normal state of Cu-O high-temperature superconductors*, Phys. Rev. Lett., **63**, 1996 (1989).

[11] N. E. Hussey, *Phenomenology of the normal state in-plane transport properties of high-T_c cuprates*, J. Phys.: Condens. Matter, **20**, 123201 (2008).

[12] J. A. N. Bruin, H. Sakai, R. S. Perry, and A. P. Mackenzie, *Similarity of Scattering Rates in Metals Showing T-Linear Resistivity*, Science, **339**, 804 (2013).

[13] A. Legros, S. Benhabib, W. Tabis, F. Laliberté, M. Dion, M. Lizaire, B. Vignolle, D. Vignolles, H. Raffy, Z. Z. Li, P. Auban-Senzier, N. Doiron-Leyraud, P. Fournier, D. Colson, L. Taillefer, and C. Proust, *Universal T-linear resistivity and Planckian dissipation in overdoped cuprates*, Nat. Phys., **15**, 142 (2019).

[14] P. W. Phillips, N. E. Hussey, and P. Abbamonte, *Stranger than metals*, Science, **377**, eabh4273 (2022).

[15] J. G. Bednorz and K. A. Müller, *Possible high-T_c superconductivity in the Ba-La-Cu-O system*, Z. Phys. B: Condens. Matter, **64**, 189 (1986).

[16] M. K. Wu, J. R. Ashburn, C. J. Torng, P. H. Hor, R. L. Meng, L. Gao, Z. J. Huang, Y. Q. Wang, and C. W. Chu, *Superconductivity at 93 K in a new mixed-phase Y-Ba-Cu-O compound system at ambient pressure*, Phys. Rev. Lett., **58**, 908 (1987).

[17] H. Takagi, S. Uchida, and Y. Tokura, *Superconductivity produced by electron doping in CuO_2-layered compounds*, Phys. Rev. Lett., **62**, 1197 (1989).

[18] D. A. Wollman, D. J. Van Harlingen, W. C. Lee, D. M. Ginsberg, and A. J. Leggett, *Experimental determination of the superconducting pairing state in YBCO from the phase coherence of YBCO-Pb dc SQUIDs*, Phys. Rev. Lett., **71**, 2134 (1993).

[19] R. Prozorov, R. W. Giannetta, P. Fournier, and R. L. Greene, *Magnetic penetration depth in electron-doped cuprates - evidence for gap nodes*, Physica C, **341–348**, 1703 (2000).

[20] C. C. Tsuei and J. R. Kirtley, *Pairing symmetry in cuprate superconductors*, Rev. Mod. Phys., **72**, 969 (2000).

[21] J. Zaanen, G. A. Sawatzky, and J. W. Allen, *Band gaps and electronic structure of transition-metal compounds*, Phys. Rev. Lett., **55**, 418 (1985).

[22] S. Uchida, T. Ido, H. Takagi, T. Arima, Y. Tokura, and S. Tajima, *Optical spectra of $La_{2-x}Sr_xCuO_4$: Effect of carrier doping on the electronic structure of the CuO_2 plane*, Phys. Rev. B, **43**, 7942 (1991).

[23] B. O. Wells, Z. X. Shen, A. Matsuura, D. M. King, M. A. Kastner, M. Greven, and R. J. Birgeneau, *E versus* **k** *Relations and Many Body Effects in the Model Insulating Copper Oxide $Sr_2CuO_2Cl_2$*, Phys. Rev. Lett., **74**, 964 (1995).

[24] F. Ronning, C. Kim, D. L. Feng, D. S. Marshall, A. G. Loeser, L. L. Miller, J. N. Eckstein, I. Bozovic, and Z.-X. Shen, *Photoemission Evidence for a Remnant Fermi Surface and a d-wave-like Dispersion in Insulating $Ca_2CuO_2Cl_2$*, Science, **282**, 2067 (1998).

[25] D. N. Basov and T. Timusk, *Electrodynamics of high-T_c superconductors*, Rev. Mod. Phys., **77**, 721 (2005).

[26] J. Hubbard, *Electron correlations in narrow energy bands V. A perturbation expansion about the atomic limit*, Proc. R. Soc. London, Ser. A, **296**, 82 (1967).

[27] P. W. Anderson, *The Resonating Valence Bond State in La_2CuO_4 and Superconductivity*, Science, **235**, 1196 (1987).

[28] R. Blankenbecler, D. J. Scalapino, and R. L. Sugar, *Monte Carlo calculations of coupled boson-fermion systems. I*, Phys. Rev. D, **24**, 2278 (1981).

[29] G. Sugiyama and S. Koonin, *Auxiliary field Monte-Carlo for quantum many-body ground states*, Ann. Phys., **168**, 1 (1986).

[30] M. Qin, C.-M. Chung, H. Shi, E. Vitali, C. Hubig, U. Schollwöck, S. R. White, S. Zhang, and S. C. o. t. M.-E. Problem, *Absence of Superconductivity in the Pure Two-Dimensional Hubbard Model*, Phys. Rev. X, **10**, 031016 (2020).

[31] H. Xu, C.-M. Chung, M. Qin, U. Schollwöck, S. R. White, and S. Zhang, *Coexistence of superconductivity with partially filled stripes in the Hubbard model*, Science, **384**, eadh7691 (2024).

[32] M. Capone and G. Kotliar, *Competition between d-wave superconductivity and antiferromagnetism in the two-dimensional Hubbard model*, Phys. Rev. B, **74**, 054513 (2006).

[33] G. Knizia and G. K.-L. Chan, *Density Matrix Embedding: A Simple Alternative to Dynamical Mean-Field Theory*, Phys. Rev. Lett., **109**, 186404 (2012).

[34] B.-X. Zheng and G. K.-L. Chan, *Ground-state phase diagram of the square lattice Hubbard model from density matrix embedding theory*, Phys. Rev. B, **93**, 035126 (2016).

[35] T. I. Vanhala and P. Törmä, *Dynamical mean-field theory study of stripe order and d-wave superconductivity in the two-dimensional Hubbard model*, Phys. Rev. B, **97**, 075112 (2018).

[36] S. R. White, *Density matrix formulation for quantum renormalization groups*, Phys. Rev. Lett., **69**, 2863 (1992).

[37] S. R. White and D. J. Scalapino, *Density Matrix Renormalization Group Study of the Striped Phase in the 2D t-J Model*, Phys. Rev. Lett., **80**, 1272 (1998).

[38] U. Schollwöck, *The density-matrix renormalization group in the age of matrix product states*, Ann. Phys., **326**, 96 (2011).

[39] E. Stoudenmire and S. R. White, *Studying Two-Dimensional Systems with the Density Matrix Renormalization Group*, Annu. Rev. Condens. Matter Phys., **3**, 111 (2012).

[40] A. Gleis, J.-W. Li, and J. von Delft, *Controlled Bond Expansion for Density Matrix Renormalization Group Ground State Search at Single-Site Costs*, Phys. Rev. Lett., **130**, 246402 (2023).

[41] F. Verstraete and J. I. Cirac, *Renormalization algorithms for Quantum-Many Body Systems in two and higher dimensions*, Arxiv (2004).

[42] J. Jordan, R. Orús, G. Vidal, F. Verstraete, and J. I. Cirac, *Classical Simulation of Infinite-Size Quantum Lattice Systems in Two Spatial Dimensions*, Phys. Rev. Lett., **101**, 250602 (2008).

[43] H. C. Jiang, Z. Y. Weng, and T. Xiang, *Accurate Determination of Tensor Network State of Quantum Lattice Models in Two Dimensions*, Phys. Rev. Lett., **101**, 090603 (2008).

[44] T. Barthel, C. Pineda, and J. Eisert, *Contraction of fermionic operator circuits and the simulation of strongly correlated fermions*, Phys. Rev. A, **80**, 042333 (2009).

[45] C. V. Kraus, N. Schuch, F. Verstraete, and J. I. Cirac, *Fermionic projected entangled pair states*, Phys. Rev. A, **81**, 052338 (2010).

[46] P. Corboz, R. Orús, B. Bauer, and G. Vidal, *Simulation of strongly correlated fermions in two spatial dimensions with fermionic projected entangled-pair states*, Phys. Rev. B, **81**, 165104 (2010).

[47] P. Corboz, J. Jordan, and G. Vidal, *Simulation of fermionic lattice models in two dimensions with projected entangled-pair states: Next-nearest neighbor Hamiltonians*, Phys. Rev. B, **82**, 245119 (2010).

[48] C. Zhang, J.-W. Li, D. Nikolaidou, and J. von Delft, *Frustration-Induced Superconductivity in the t-t' Hubbard Model*, Phys. Rev. Lett., **134**, 116502 (2025).

[49] S. R. White, *Minimally Entangled Typical Quantum States at Finite Temperature*, Phys. Rev. Lett., **102**, 190601 (2009).

[50] E. M. Stoudenmire and S. R. White, *Minimally entangled typical thermal state algorithms*, New J. Phys., **12**, 55026 (2010).

[51] W. Li, S.-J. Ran, S.-S. Gong, Y. Zhao, B. Xi, F. Ye, and G. Su, *Linearized Tensor Renormalization Group Algorithm for the Calculation of Thermodynamic Properties of Quantum Lattice Models*, Phys. Rev. Lett., **106**, 127202 (2011).

[52] P. Czarnik, L. Cincio, and J. Dziarmaga, *Projected entangled pair states at finite temperature: Imaginary time evolution with ancillas*, Phys. Rev. B, **86**, 245101 (2012).

[53] P. Czarnik and J. Dziarmaga, *Variational approach to projected entangled pair states at finite temperature*, Phys. Rev. B, **92**, 035152 (2015).

[54] B.-B. Chen, L. Chen, Z. Chen, W. Li, and A. Weichselbaum, *Exponential Thermal Tensor Network Approach for Quantum Lattice Models*, Phys. Rev. X, **8**, 031082 (2018).

[55] H. Li, B.-B. Chen, Z. Chen, J. von Delft, A. Weichselbaum, and W. Li, *Thermal tensor renormalization group simulations of square-lattice quantum spin models*, Phys. Rev. B, **100**, 045110 (2019).

[56] Q. Li, Y. Gao, Y.-Y. He, Y. Qi, B.-B. Chen, and W. Li, *Tangent Space Approach for Thermal Tensor Network Simulations of the 2D Hubbard Model*, Phys. Rev. Lett., **130**, 226502 (2023).

[57] C. Zhang and J. von Delft, *Finite-Temperature Study of the Hubbard Model via Enhanced Exponential Tensor Renormalization Group*, (2025), doi:10.48550/arXiv.2510.25022.

[58] T. Nishino and K. Okunishi, *Corner Transfer Matrix Renormalization Group Method*, J. Phys. Soc. Jpn., **65**, 891 (1996).

[59] R. Orús and G. Vidal, *Simulation of two-dimensional quantum systems on an infinite lattice revisited: Corner transfer matrix for tensor contraction*, Phys. Rev. B, **80**, 094403 (2009).

[60] L. Chen, D.-W. Qu, H. Li, B.-B. Chen, S.-S. Gong, J. von Delft, A. Weichselbaum, and W. Li, *Two-temperature scales in the triangular-lattice Heisenberg antiferromagnet*, Phys. Rev. B, **99**, 140404 (2019).

[61] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, *Attention is All you Need*, in *Advances in Neural Information Processing Systems*, vol. 30, Curran Associates, Inc., doi:10.48550/arXiv.1706.03762.

[62] D. Bahdanau, K. Cho, and Y. Bengio, *Neural Machine Translation by Jointly Learning to Align and Translate*, in *International Conference on Learning Representations*, doi:10.48550/arXiv.1409.0473.

[63] T. Rocktäschel, E. Grefenstette, K. M. Hermann, T. Kočiský, and P. Blunsom, *Reasoning about Entailment with Neural Attention*, in *International Conference on Learning Representations*, doi:10.48550/arXiv.1509.06664.

[64] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, *Show, Attend and Tell: Neural Image Caption Generation with Visual Attention*, in *Proceedings of the 32nd International Conference on Machine Learning*, 2048–2057, PMLR, ISSN 1938-7228, doi:10.48550/arXiv.1502.03044.

[65] Y. Wang, M. Huang, X. Zhu, and L. Zhao, *Attention-based LSTM for Aspect-level Sentiment Classification*, in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (edited by J. Su, K. Duh,

and X. Carreras), 606–615, Association for Computational Linguistics, doi: 10.18653/v1/D16-1058.

[66] Y. Kim, C. Denton, L. Hoang, and A. M. Rush, *Structured Attention Networks*, in *International Conference on Learning Representations*, doi:10.48550/arXiv.1702.00887.

[67] L. Dong, S. Xu, and B. Xu, *Speech-Transformer: A No-Recurrence Sequence-to-Sequence Model for Speech Recognition*, in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5884–5888, ISSN 2379-190X, doi:10.1109/ICASSP.2018.8462506.

[68] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (long and Short Papers)* (edited by J. Burstein, C. Doran, and T. Solorio), 4171–4186, Association for Computational Linguistics, doi:10.18653/v1/N19-1423.

[69] Y. Ren, Y. Ruan, X. Tan, T. Qin, S. Zhao, Z. Zhao, and T.-Y. Liu, *FastSpeech: Fast, Robust and Controllable Text to Speech*, in *Advances in Neural Information Processing Systems*, vol. 32, Curran Associates, Inc., doi: 10.48550/arXiv.1905.09263.

[70] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, *Language Models are Few-Shot Learners*, in *Advances in Neural Information Processing Systems*, vol. 33, 1877–1901, Curran Associates, Inc., doi: 10.48550/arXiv.2005.14165.

[71] A. Gulati, J. Qin, C.-C. Chiu, N. Parmar, Y. Zhang, J. Yu, W. Han, S. Wang, Z. Zhang, Y. Wu, and R. Pang, *Conformer: Convolution-augmented Transformer for Speech Recognition*, in *Interspeech*, doi:10.21437/interspeech.2020-3015.

[72] S. Latif, A. Zaidi, H. Cuayahuitl, F. Shamshad, M. Shoukat, M. Usama, and J. Qadir, *Transformers in Speech Processing: A Survey*, in *Computer Science Review*, doi:10.48550/arXiv.2303.11607.

[73] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, *End-to-End Object Detection with Transformers*, in *Computer Vision – ECCV 2020* (edited by A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm), 213–229, Springer International Publishing, ISBN 978-3-030-58452-8, doi:10.1007/978-3-030-58452-8_13.

[74] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*, in *International Conference on Learning Representations*, arXiv, doi:10.48550/arXiv.2010.11929.

[75] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows*, in *IEEE International Conference on Computer Vision*, 9992–10002, arXiv, doi:10.1109/ICCV48922.2021.00986.

[76] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, *Transformers in Vision: A Survey*, ACM Comput, Surv,, **54**, 200:1 (2022).

[77] A. Jaegle, F. Gimeno, A. Brock, A. Zisserman, O. Vinyals, and J. Carreira, *Perceiver: General Perception with Iterative Attention*, in *International Conference on Machine Learning*, arXiv, doi:10.48550/arXiv.2103.03206.

[78] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, *Highly accurate protein structure prediction with AlphaFold*, Nature, **596**, 583 (2021).

[79] A. Rives, J. Meier, T. Sercu, S. Goyal, Z. Lin, J. Liu, D. Guo, M. Ott, C. L. Zitnick, J. Ma, and R. Fergus, *Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences*, Proc. Natl. Acad. Sci., **118**, e2016239118 (2021).

[80] L. Chen, K. Lu, A. Rajeswaran, K. Lee, A. Grover, M. Laskin, P. Abbeel, A. Srinivas, and I. Mordatch, *Decision Transformer: Reinforcement Learning via Sequence Modeling*, in *Neural Information Processing Systems*, 15084–15097, doi:10.48550/arXiv.2106.01345.

[81] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan, and L. Sun, *Transformers in Time Series: A Survey*, in *International Joint Conference on Artificial Intelligence*, arXiv, doi:10.24963/ijcai.2023/759.

[82] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, *Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting*, in *AAAI Conference on Artificial Intelligence*, arXiv, doi:10.1609/aaai.v35i12.17325.

[83] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, *Graph Attention Networks*, in *International Conference on Learning Representations*.

[84] Y. Gu, W. Li, H. Lin, B. Zhan, R. Li, Y. Huang, D. He, Y. Wu, T. Xiang, M. Qin, L. Wang, and D. Lv, *Solving the Hubbard model with Neural Quantum States*, (2025), doi:10.48550/arXiv.2507.02644.

[85] C. Zhang and J. von Delft, *Interpretable Artificial Intelligence (AI) Analysis of Strongly Correlated Electrons*, (2025), doi:10.48550/arXiv.2510.26864.

[86] A. Mazurenko, C. S. Chiu, G. Ji, M. F. Parsons, M. Kanász-Nagy, R. Schmidt, F. Grusdt, E. Demler, D. Greif, and M. Greiner, *A cold-atom Fermi–Hubbard antiferromagnet*, Nature, **545**, 462 (2017).

[87] C. S. Chiu, G. Ji, A. Mazurenko, D. Greif, and M. Greiner, *Quantum State Engineering of a Hubbard System with Ultracold Fermions*, Phys. Rev. Lett., **120**, 243201 (2018).

[88] J. Koepsell, J. Vijayan, P. Sompet, F. Grusdt, T. A. Hilker, E. Demler, G. Salomon, I. Bloch, and C. Gross, *Imaging magnetic polarons in the doped Fermi–Hubbard model*, Nature, **572**, 358 (2019).

[89] J. Koepsell, S. Hirthe, D. Bourgund, P. Sompet, J. Vijayan, G. Salomon, C. Gross, and I. Bloch, *Robust Bilayer Charge Pumping for Spin- and Density-Resolved Quantum Gas Microscopy*, Phys. Rev. Lett., **125**, 010403 (2020).

[90] T. Chalopin, P. Bojović, D. Bourgund, S. Wang, T. Franz, I. Bloch, and T. Hilker, *Optical Superlattice for Engineering Hubbard Couplings in Quantum Simulation*, Phys. Rev. Lett., **134**, 53402 (2025).

[91] M. Xu, L. H. Kendrick, A. Kale, Y. Gang, C. Feng, S. Zhang, A. W. Young, M. Lebrat, and M. Greiner, *A neutral-atom Hubbard quantum simulator in the cryogenic regime*, Nature, **642**, 909 (2025).

[92] J. Orenstein and A. J. Millis, *Advances in the Physics of High-Temperature Superconductivity*, Science, **288**, 468 (2000).

[93] B. Keimer, S. A. Kivelson, M. R. Norman, S. Uchida, and J. Zaanen, *From quantum matter to high-temperature superconductivity in copper oxides*, Nature, **518**, 179 (2015).

[94] D. Rybicki, M. Jurkutat, S. Reichardt, C. Kapusta, and J. Haase, *Perspective on the phase diagram of cuprate high-temperature superconductors*, Nat. Commun., **7**, 11413 (2016).

[95] C. Proust and L. Taillefer, *The Remarkable Underlying Ground States of Cuprate Superconductors*, Annu. Rev. Condens. Matter Phys., **10**, 409 (2019).

[96] G. Blumberg, P. Abbamonte, M. V. Klein, W. C. Lee, D. M. Ginsberg, L. L. Miller, and A. Zibold, *Resonant two-magnon Raman scattering in cuprate antiferromagnetic insulators*, Phys. Rev. B, **53**, R11930 (1996).

[97] R. Coldea, S. M. Hayden, G. Aeppli, T. G. Perring, C. D. Frost, T. E. Mason, S.-W. Cheong, and Z. Fisk, *Spin Waves and Electronic Interactions in La_2CuO_4*, Phys. Rev. Lett., **86**, 5377 (2001).

[98] O. Cyr-Choinière, R. Daou, F. Laliberté, C. Collignon, S. Badoux, D. LeBoeuf, J. Chang, B. J. Ramshaw, D. A. Bonn, W. N. Hardy, R. Liang, J.-Q. Yan, J.-G. Cheng, J.-S. Zhou, J. B. Goodenough, S. Pyon, T. Takayama, H. Takagi, N. Doiron-Leyraud, and L. Taillefer, *Pseudogap temperature T^* of cuprate superconductors from the Nernst effect*, Phys. Rev. B, **97**, 64502 (2018).

[99] M. Horio and A. Fujimori, *ARPES studies on new types of electron-doped cuprate superconductors*, J. Phys.: Condens. Matter, **30**, 503001 (2018).

[100] P. R. Mandal, T. Sarkar, and R. L. Greene, *Anomalous quantum criticality in the electron-doped cuprates*, Proc. Natl. Acad. Sci., **116**, 5991 (2019).

[101] Y. J. Uemura, G. M. Luke, B. J. Sternlieb, J. H. Brewer, J. F. Carolan, W. N. Hardy, R. Kadono, J. R. Kempton, R. F. Kiefl, S. R. Kreitzman, P. Mulhern, T. M. Riseman, D. Ll. Williams, B. X. Yang, S. Uchida, H. Takagi, J. Gopalakrishnan, A. W. Sleight, M. A. Subramanian, C. L. Chien, M. Z. Cieplak, G. Xiao, V. Y. Lee, B. W. Statt, C. E. Stronach, W. J. Kossler, and X. H. Yu, *Universal Correlations between T_c and n_s/m^* (Carrier Density over Effective Mass) in High-T_c Cuprate Superconductors*, Phys. Rev. Lett., **62**, 2317 (1989).

[102] Y. J. Uemura, *Classifying superconductors in a plot of T_c versus Fermi temperature T_F*, Physica C, **185–189**, 733 (1991).

[103] W. N. Hardy, D. A. Bonn, D. C. Morgan, R. Liang, and K. Zhang, *Precision measurements of the temperature dependence of λ in $YBa_2Cu_3O_{6.95}$: Strong evidence for nodes in the gap function*, Phys. Rev. Lett., **70**, 3999 (1993).

[104] Z.-X. Shen, D. S. Dessau, B. O. Wells, D. M. King, W. E. Spicer, A. J. Arko, D. Marshall, L. W. Lombardo, A. Kapitulnik, P. Dickinson, S. Doniach, J. DiCarlo, T. Loeser, and C. H. Park, *Anomalously large gap anisotropy in the a-b plane of $Bi_2Sr_2CaCu_2O_{8+\delta}$*, Phys. Rev. Lett., **70**, 1553 (1993).

[105] D. J. Van Harlingen, *Phase-sensitive tests of the symmetry of the pairing state in the high-temperature superconductors—Evidence for $d_{x^2-y^2}$ symmetry*, Rev. Mod. Phys., **67**, 515 (1995).

[106] J. E. Hoffman, E. W. Hudson, K. M. Lang, V. Madhavan, H. Eisaki, S. Uchida, and J. C. Davis, *A Four Unit Cell Periodic Pattern of Quasi-Particle States Surrounding Vortex Cores in $Bi_2Sr_2CaCu_2O_{8+\delta}$*, Science, **295**, 466 (2002).

[107] C. Proust, E. Boaknin, R. W. Hill, L. Taillefer, and A. P. Mackenzie, *Heat Transport in a Strongly Overdoped Cuprate: Fermi Liquid and a Pure d-wave BCS Superconductor*, Phys. Rev. Lett., **89**, 147003 (2002).

[108] M. Platé, J. D. F. Mottershead, I. S. Elfimov, D. C. Peets, R. Liang, D. A. Bonn, W. N. Hardy, S. Chiuzbaian, M. Falub, M. Shi, L. Patthey, and A. Damascelli, *Fermi Surface and Quasiparticle Excitations of Overdoped $Tl_2Ba_2CuO_{6+\delta}$*, Phys. Rev. Lett., **95**, 77001 (2005).

[109] B. Vignolle, A. Carrington, R. A. Cooper, M. M. J. French, A. P. Mackenzie, C. Jaudet, D. Vignolles, C. Proust, and N. E. Hussey, *Quantum oscillations in an overdoped high-T_c superconductor*, Nature, **455**, 952 (2008).

[110] N. Barišić, M. K. Chan, Y. Li, G. Yu, X. Zhao, M. Dressel, A. Smontara, and M. Greven, *Universal sheet resistance and revised phase diagram of the cuprate high-temperature superconductors*, Proc. Natl. Acad. Sci., **110**, 12235 (2013).

[111] C. Putzke, S. Benhabib, W. Tabis, J. Ayres, Z. Wang, L. Malone, S. Licciardello, J. Lu, T. Kondo, T. Takeuchi, N. E. Hussey, J. R. Cooper, and A. Carrington, *Reduced Hall carrier density in the overdoped strange metal regime of cuprate superconductors*, Nat. Phys., **17**, 826 (2021).

[112] J. M. Tranquada, B. J. Sternlieb, J. D. Axe, Y. Nakamura, and S. Uchida, *Evidence for stripe correlations of spins and holes in copper oxide superconductors*, Nature, **375**, 561 (1995).

[113] J. M. Tranquada, J. D. Axe, N. Ichikawa, Y. Nakamura, S. Uchida, and B. Nachumi, *Neutron-scattering study of stripe-phase order of holes and spins in $La_{1.48}Nd_{0.4}Sr_{0.12}CuO_4$*, Phys. Rev. B, **54**, 7489 (1996).

[114] K. Yamada, C. H. Lee, K. Kurahashi, J. Wada, S. Wakimoto, S. Ueki, H. Kimura, Y. Endoh, S. Hosoya, G. Shirane, R. J. Birgeneau, M. Greven, M. A. Kastner, and Y. J. Kim, *Doping dependence of the spatially modulated dynamical spin correlations and the superconducting-transition temperature in $La_{2-x}Sr_xCuO_4$*, Phys. Rev. B, **57**, 6165 (1998).

[115] A. W. Hunt, P. M. Singer, K. R. Thurber, and T. Imai, *^{63}Cu NQR Measurement of Stripe Order Parameter in $La_{2-x}Sr_xCuO_4$*, Phys. Rev. Lett., **82**, 4300 (1999).

[116] C. Howald, H. Eisaki, N. Kaneko, M. Greven, and A. Kapitulnik, *Periodic density-of-states modulations in superconducting $Bi_2Sr_2CaCu_2O_{8+\delta}$*, Phys. Rev. B, **67**, 014533 (2003).

[117] J. M. Tranquada, H. Woo, T. G. Perring, H. Goka, G. D. Gu, G. Xu, M. Fujita, and K. Yamada, *Quantum magnetic excitations from stripes in copper oxide superconductors*, Nature, **429**, 534 (2004).

[118] P. Abbamonte, A. Rusydi, S. Smadici, G. D. Gu, G. A. Sawatzky, and D. L. Feng, *Spatially modulated 'Mottness' in $La_{2-x}Ba_xCuO_4$*, Nat. Phys., **1**, 155 (2005).

[119] R. J. Birgeneau, C. Stock, J. M. Tranquada, and K. Yamada, *Magnetic Neutron Scattering in Hole-Doped Cuprate Superconductors*, J. Phys. Soc. Jpn., **75**, 111003 (2006).

[120] W. D. Wise, M. C. Boyer, K. Chatterjee, T. Kondo, T. Takeuchi, H. Ikuta, Y. Wang, and E. W. Hudson, *Charge-density-wave origin of cuprate checkerboard visualized by scanning tunnelling microscopy*, Nat. Phys., **4**, 696 (2008).

[121] T. Wu, H. Mayaffre, S. Krämer, M. Horvatić, C. Berthier, W. N. Hardy, R. Liang, D. A. Bonn, and M.-H. Julien, *Magnetic-field-induced charge-stripe order in the high-temperature superconductor $YBa_2Cu_3O_y$*, Nature, **477**, 191 (2011).

[122] G. Ghiringhelli, M. L. Tacon, M. Minola, S. Blanco-Canosa, C. Mazzoli, N. B. Brookes, G. M. D. Luca, A. Frano, D. G. Hawthorn, F. He, T. Loew, M. M. Sala, D. C. Peets, M. Salluzzo, E. Schierle, R. Sutarto, G. A. Sawatzky, E. Weschke, B. Keimer, and L. Braicovich, *Long-Range Incommensurate Charge Fluctuations in (Y,Nd)$Ba_2Cu_3O_{6+x}$*, Science, **337**, 821 (2012).

[123] E. H. d. S. Neto, P. Aynajian, A. Frano, R. Comin, E. Schierle, E. Weschke, A. Gyenis, J. Wen, J. Schneeloch, Z. Xu, S. Ono, G. Gu, M. L. Tacon, and A. Yazdani, *Ubiquitous Interplay Between Charge Ordering and High-Temperature Superconductivity in Cuprates*, Science, **343**, 393 (2014).

[124] R. Comin and A. Damascelli, *Resonant X-Ray Scattering Studies of Charge Order in Cuprates*, Annu. Rev. Condens. Matter Phys., **7**, 1 (2015).

[125] T. Wu, H. Mayaffre, S. Krämer, M. Horvatić, C. Berthier, W. Hardy, R. Liang, D. Bonn, and M.-H. Julien, *Incipient charge order observed by NMR in the normal state of $YBa_2Cu_3O_y$*, Nat. Commun., **6**, 6438 (2015).

[126] A. Mesaros, K. Fujita, S. D. Edkins, M. H. Hamidian, H. Eisaki, S.-i. Uchida, J. C. S. Davis, M. J. Lawler, and E.-A. Kim, *Commensurate $4a_0$-period charge density modulations throughout the $Bi_2Sr_2CaCu_2O_{8+x}$ pseudogap regime*, Proc. Natl. Acad. Sci. U.S.A., **113**, 12661 (2016).

[127] T. Imai, S. K. Takahashi, A. Arsenault, A. W. Acton, D. Lee, W. He, Y. S. Lee, and M. Fujita, *Revisiting ^{63}Cu NMR evidence for charge order in superconducting $La_{1.885}Sr_{0.115}CuO_4$*, Phys. Rev. B, **96**, 224508 (2017).

[128] H. Miao, J. Lorenzana, G. Seibold, Y. Y. Peng, A. Amorese, F. Yakhou-Harris, K. Kummer, N. B. Brookes, R. M. Konik, V. Thampy, G. D. Gu, G. Ghiringhelli, L. Braicovich, and M. P. M. Dean, *High-temperature charge density wave correlations in $La_{1.875}Ba_{0.125}CuO_4$ without spin–charge locking*, Proc. Natl. Acad. Sci. U.S.A., **114**, 12430 (2017).

[129] R. Daou, J. Chang, D. LeBoeuf, O. Cyr-Choinière, F. Laliberté, N. Doiron-Leyraud, B. J. Ramshaw, R. Liang, D. A. Bonn, W. N. Hardy, and L. Taillefer, *Broken rotational symmetry in the pseudogap phase of a high-T_c superconductor*, Nature, **463**, 519 (2010).

[130] J. Chang, N. Doiron-Leyraud, F. Laliberté, R. Daou, D. LeBoeuf, B. J. Ramshaw, R. Liang, D. A. Bonn, W. N. Hardy, C. Proust, I. Sheikin, K. Behnia, and L. Taillefer, *Nernst effect in the cuprate superconductor $YBa_2Cu_3O_{7-x}$: Broken rotational and translational symmetries*, Phys. Rev. B, **84**, 014507 (2011).

[131] J. D. Jorgensen, M. A. Beno, D. G. Hinks, L. Soderholm, K. J. Volin, R. L. Hitterman, J. D. Grace, I. K. Schuller, C. U. Segre, K. Zhang, and M. S.

Kleefisch, *Oxygen ordering and the orthorhombic-to-tetragonal phase transition in $YBa_2Cu_3O_{7-x}$*, Phys. Rev. B, **36**, 3608 (1987).

[132] J. D. Jorgensen, B. W. Veal, A. P. Paulikas, L. J. Nowicki, G. W. Crabtree, H. Claus, and W. K. Kwok, *Structural properties of oxygen-deficient $YBa_2Cu_3O_{7-x}$*, Phys. Rev. B, **41**, 1863 (1990).

[133] M. D. Ainslie, *Transport AC loss in high temperature superconducting coils* (2012).

[134] S. J. Hagen, T. W. Jing, Z. Z. Wang, J. Horvath, and N. P. Ong, *Out-of-plane conductivity in single-crystal $YBa_2Cu_3O_{7-x}$*, Phys. Rev. B, **37**, 7928 (1988).

[135] M. Takigawa, P. C. Hammel, R. H. Heffner, and Z. Fisk, *Spin susceptibility in superconducting $YBa_2Cu_3O_{7-x}$ from ^{63}Cu Knight shift*, Phys. Rev. B, **39**, 7371 (1989).

[136] X. G. Qiu, H. Koinuma, M. Iwasaki, T. Itoh, A. K. Sarin Kumar, M. Kawasaki, E. Saitoh, Y. Tokura, K. Takehana, G. Kido, and Y. Segawa, *Josephson plasma observed in underdoped (110) $YBa_2Cu_3O_{7-x}$ films*, Appl. Phys. Lett., **78**, 506 (2001).

[137] Z. A. Xu, J. Q. Shen, S. R. Zhao, Y. J. Zhang, and C. K. Ong, *Nernst effect and superconducting fluctuations in Zn-doped $YBa_2Cu_3O_{7-x}$*, Phys. Rev. B, **72**, 144527 (2005).

[138] P. Kuiper, G. Kruizinga, J. Ghijsen, M. Grioni, P. J. W. Weijs, F. M. F. De Groot, G. A. Sawatzky, H. Verweij, L. F. Feiner, and H. Petersen, *X-ray absorption study of the O 2p hole concentration dependence on O stoichiometry in $YBa_2Cu_3O_x$*, Phys. Rev. B, **38**, 6483 (1988).

[139] C. T. Chen, F. Sette, Y. Ma, M. S. Hybertsen, E. B. Stechel, W. M. C. Foulkes, M. Schulter, S.-W. Cheong, A. S. Cooper, L. W. Rupp, B. Batlogg, Y. L. Soo, Z. H. Ming, A. Krol, and Y. H. Kao, *Electronic States in $La_{2-x}Sr_xCuO_{4+\delta}$ Probed by Soft-X-Ray Absorption*, Phys. Rev. Lett., **66**, 104 (1991).

[140] D. S. Ellis, J. P. Hill, S. Wakimoto, R. J. Birgeneau, D. Casa, T. Gog, and Y.-J. Kim, *Charge-transfer exciton in La_2CuO_4 probed with resonant inelastic x-ray scattering*, Phys. Rev. B, **77**, 060501 (2008).

[141] V. J. Emery, *Theory of high-T_c superconductivity in oxides*, Phys. Rev. Lett., **58**, 2794 (1987).

[142] M. S. Hybertsen and M. Schlüter, *Models for the Electronic Structure of Cuprates*, in *New Horizons in Low-dimensional Electron Systems: A Festschrift in Honour of Professor H. Kamimura* (edited by H. Aoki, M. Tsukada, M. Schlüter, and F. Lévy). pages 229–245. Springer Netherlands, Dordrecht 1992, ISBN 978-94-011-3190-2. doi:10.1007/978-94-011-3190-2_16.

[143] F. C. Zhang and T. M. Rice, *Effective Hamiltonian for the superconducting Cu oxides*, Phys. Rev. B, **37**, 3759 (1988).

[144] L. H. Tjeng, B. Sinkovic, N. B. Brookes, J. B. Goedkoop, R. Hesper, E. Pellegrin, F. M. F. De Groot, S. Altieri, S. L. Hulbert, E. Shekel, and G. A. Sawatzky, Spin-*Resolved Photoemission on* Anti-*Ferromagnets: Direct Observation of Zhang-Rice Singlets in CuO*, Phys. Rev. Lett., **78**, 1126 (1997).

[145] N. B. Brookes, G. Ghiringhelli, O. Tjernberg, L. H. Tjeng, T. Mizokawa, T. W. Li, and A. A. Menovsky, *Detection of Zhang-Rice Singlets Using Spin-Polarized Photoemission*, Phys. Rev. Lett., **87**, 237003 (2001).

[146] D. J. Scalapino, *Monte Carlo calculations of coupled boson-fermion systems. II*, Phys. Rev. B, **24**, 4295 (1981).

[147] M. Troyer and U.-J. Wiese, *Computational Complexity and Fundamental Limitations to Fermionic Quantum Monte Carlo Simulations*, Phys. Rev. Lett., **94**, 170201 (2005).

[148] W. A. Al-Saidi, S. Zhang, and H. Krakauer, *Auxiliary-field quantum Monte Carlo calculations of molecular systems with a Gaussian basis*, J. Chem. Phys., **124**, 224101 (2006).

[149] S. Zhang, *Constrained Path Quantum Monte Carlo Method for Fermion Ground States*, Phys. Rev. Lett., **74**, 3652 (1995).

[150] S. Liang, *Approximate diagonalization using the density matrix renormalization-group method: A two-dimensional-systems perspective*, Phys. Rev. B, **49**, 9214 (1994).

[151] S. Jiang, *Density-Matrix Renormalization Group and Model Reduction Studies of Two-Dimensional Doped and Frustrated Systems*, Ph.D. thesis, UC Irvine (2024).

[152] M. B. Hastings, *An area law for one-dimensional quantum systems*, J. Stat. Mech: Theory Exp., **2007**, P08024 (2007).

[153] J. Eisert, M. Cramer, and M. B. Plenio, *Colloquium: Area laws for the entanglement entropy*, Rev. Mod. Phys., **82**, 277 (2010).

[154] I. Frérot, *Area law and its violation: A microscopic inspection into the structure of entanglement and fluctuations*, Phys. Rev. B, **92**, 115129 (2015).

[155] R. Orús, *A practical introduction to tensor networks: Matrix product states and projected entangled pair states*, Ann. Phys., **349**, 117 (2014).

[156] H.-C. Jiang and T. P. Devereaux, *Superconductivity in the doped Hubbard model and its interplay with next-nearest hopping t'*, Science, **365**, 1424 (2019).

[157] Y.-F. Jiang, J. Zaanen, T. P. Devereaux, and H.-C. Jiang, *Ground state phase diagram of the doped Hubbard model on the four-leg cylinder*, Phys. Rev. Res., **2**, 33073 (2020).

[158] C.-M. Chung, M. Qin, S. Zhang, U. Schollwöck, and S. R. White, *Plaquette versus ordinary d-wave pairing in the t'-Hubbard model on a width-4 cylinder*, Phys. Rev. B, **102**, 041106 (2020).

[159] Y.-F. Jiang, T. P. Devereaux, and H.-C. Jiang, *Ground-State Phase Diagram and Superconductivity of the Doped Hubbard Model on Six-Leg Square Cylinders*, Phys. Rev. B, **109**, 85121 (2024).

[160] F. Verstraete, V. Murg, and J. Cirac, *Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems*, Adv. Phys., **57**, 143 (2008).

[161] P. Corboz, S. R. White, G. Vidal, and M. Troyer, *Stripes in the two-dimensional t-J model with infinite projected entangled-pair states*, Phys. Rev. B, **84**, 041108 (2011).

[162] P. Corboz, T. M. Rice, and M. Troyer, *Competing States in the t-J Model: Uniform d-wave State versus Stripe State*, Phys. Rev. Lett., **113**, 046402 (2014).

[163] H. N. Phien, J. A. Bengua, H. D. Tuan, P. Corboz, and R. Orús, *Infinite projected entangled pair states algorithm improved: Fast full update and gauge fixing*, Phys. Rev. B, **92**, 035142 (2015).

[164] P. Corboz, *Variational optimization with infinite projected entangled-pair states*, Phys. Rev. B, **94**, 035133 (2016).

[165] H.-J. Liao, J.-G. Liu, L. Wang, and T. Xiang, *Differentiable Programming Tensor Networks*, Phys. Rev. X, **9**, 031041 (2019).

[166] W. Li, L. Ding, R. Yu, T. Roscilde, and S. Haas, *Scaling behavior of entanglement in two- and three-dimensional free-fermion systems*, Phys. Rev. B, **74**, 073103 (2006).

[167] E. Dagotto, *Correlated electrons in high-temperature superconductors*, Rev. Mod. Phys., **66**, 763 (1994).

[168] A. Weiße and H. Fehske, *Exact Diagonalization Techniques*, in *Computational Many-particle Physics* (edited by H. Fehske, R. Schneider, and A. Weiße). pages 529–544. Springer, Berlin, Heidelberg 2008, ISBN 978-3-540-74686-7. doi:10.1007/978-3-540-74686-7_18.

[169] G. Vidal, *Entanglement Renormalization*, Phys. Rev. Lett., **99**, 220405 (2007).

[170] G. Vidal, *Class of Quantum Many-Body States That Can Be Efficiently Simulated*, Phys. Rev. Lett., **101**, 110501 (2008).

[171] G. Evenbly and G. Vidal, *Tensor Network Renormalization*, Phys. Rev. Lett., **115**, 180405 (2015).

[172] G. Evenbly and G. Vidal, *Tensor Network Renormalization Yields the Multiscale Entanglement Renormalization Ansatz*, Phys. Rev. Lett., **115**, 200401 (2015).

[173] Z. Y. Xie, J. Chen, M. P. Qin, J. W. Zhu, L. P. Yang, and T. Xiang, *Coarse-graining renormalization by higher-order singular value decomposition*, Phys. Rev. B, **86**, 045139 (2012).

[174] N. V. Prokof'ev and B. V. Svistunov, *Polaron Problem by Diagrammatic Quantum Monte Carlo*, Phys. Rev. Lett., **81**, 2514 (1998).

[175] Y. Deng, E. Kozik, N. V. Prokof'ev, and B. V. Svistunov, *Emergent BCS regime of the two-dimensional fermionic Hubbard model: Ground-state phase diagram*, Epl-europhys Lett, **110**, 57001 (2015).

[176] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, *Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions*, Rev. Mod. Phys., **68**, 13 (1996).

[177] T. Maier, M. Jarrell, T. Pruschke, and M. H. Hettler, *Quantum cluster theories*, Rev. Mod. Phys., **77**, 1027 (2005).

[178] J. Otsuki, H. Hafermann, and A. I. Lichtenstein, *Superconductivity, antiferromagnetism, and phase separation in the two-dimensional Hubbard model: A dual-fermion approach*, Phys. Rev. B, **90**, 235132 (2014).

[179] G. Carleo and M. Troyer, *Solving the quantum many-body problem with artificial neural networks*, Science, **355**, 602 (2017).

[180] Y.-H. Zhang, *Transformer quantum state: A multipurpose model for quantum many-body problems*, Phys. Rev. B, **107** (2023).

[181] M. Hibat-Allah, E. Merali, G. Torlai, R. G. Melko, and J. Carrasquilla, *Recurrent neural network wave functions for Rydberg atom arrays on kagome lattice*, Commun. Phys., **8**, 308 (2025).

[182] E. Ibarra-García-Padilla, H. Lange, R. G. Melko, R. T. Scalettar, J. Carrasquilla, A. Bohrdt, and E. Khatami, *Autoregressive neural quantum states of Fermi Hubbard models*, Phys. Rev. Research, **7**, 013122 (2025).

[183] H. Lange, G. Bornet, G. Emperauger, C. Chen, T. Lahaye, S. Kienle, A. Browaeys, and A. Bohrdt, *Transformer neural networks and quantum simulators: a hybrid approach for simulating strongly correlated systems*, Quantum, **9**, 1675 (2025).

[184] J. Jaklič and P. Prelovšek, *Lanczos method for the calculation of finite-temperature quantities in correlated systems*, Phys. Rev. B, **49**, 5065 (1994).

[185] J. Jaklič and P. Prelovšek, *Thermodynamic Properties of the Planar t-J Model*, Phys. Rev. Lett., **77**, 892 (1996).

[186] J. Jaklič and P. Prelovšek, *Finite-temperature properties of doped antiferromagnets*, Adv. Phys., **49**, 1 (2000).

[187] K.-S. Chen, Z. Y. Meng, S.-X. Yang, T. Pruschke, J. Moreno, and M. Jarrell, *Evolution of the superconductivity dome in the two-dimensional Hubbard model*, Phys. Rev. B, **88**, 245110 (2013).

[188] Y.-Y. He, M. Qin, H. Shi, Z.-Y. Lu, and S. Zhang, *Finite-temperature auxiliary-field quantum Monte Carlo: Self-consistent constraint and systematic approach to low temperatures*, Phys. Rev. B, **99**, 45108 (2019).

[189] A. N. Rubtsov, V. V. Savkin, and A. I. Lichtenstein, *Continuous-time quantum Monte Carlo method for fermions*, Phys. Rev. B, **72**, 035122 (2005).

[190] E. Gull, A. J. Millis, A. I. Lichtenstein, A. N. Rubtsov, M. Troyer, and P. Werner, *Continuous-time Monte Carlo methods for quantum impurity models*, Rev. Mod. Phys., **83**, 349 (2011).

[191] M. Binder and T. Barthel, *Minimally entangled typical thermal states versus matrix product purifications for the simulation of equilibrium states and time evolution*, Phys. Rev. B, **92**, 125119 (2015).

[192] A. Wietek, R. Rossi, F. Šimkovic, M. Klett, P. Hansmann, M. Ferrero, E. M. Stoudenmire, T. Schäfer, and A. Georges, *Mott Insulating States with Competing Orders in the Triangular Lattice Hubbard Model*, Phys. Rev. X, **11**, 41013 (2021).

[193] A. Wietek, Y.-Y. He, S. R. White, A. Georges, and E. M. Stoudenmire, *Stripes, Antiferromagnetism, and the Pseudogap in the Doped Hubbard Model at Finite Temperature*, Phys. Rev. X, **11**, 31007 (2021).

[194] P. Czarnik and J. Dziarmaga, *Fermionic projected entangled pair states at finite temperature*, Phys. Rev. B, **90**, 035144 (2014).

[195] P. Czarnik, J. Dziarmaga, and P. Corboz, *Time evolution of an infinite projected entangled pair state: An efficient algorithm*, Phys. Rev. B, **99**, 35115 (2019).

[196] P. Czarnik and P. Corboz, *Finite correlation length scaling with infinite projected entangled pair states at finite temperature*, Phys. Rev. B, **99**, 245107 (2019).

[197] P. Schmoll, C. Balz, B. Lake, J. Eisert, and A. Kshetrimayum, *Finite temperature tensor network algorithm for frustrated two-dimensional quantum materials*, Phys. Rev. B, **109**, 235119 (2024).

[198] D.-W. Qu, Q. Li, S.-S. Gong, Y. Qi, W. Li, and G. Su, *Phase Diagram, d-Wave Superconductivity, and Pseudogap of the t-t'-J Model at Finite Temperature*, Phys. Rev. Lett., **133**, 256003 (2024).

[199] B.-B. Chen, C. Chen, Z. Chen, J. Cui, Y. Zhai, A. Weichselbaum, J. von Delft, Z. Y. Meng, and W. Li, *Quantum many-body simulations of the two-dimensional Fermi-Hubbard model in ultracold optical lattices*, Phys. Rev. B, **103**, L041107 (2021).

[200] X. Lin, B.-B. Chen, W. Li, Z. Y. Meng, and T. Shi, *Exciton Proliferation and Fate of the Topological Mott Insulator in a Twisted Bilayer Graphene Lattice Model*, Phys. Rev. Lett., **128**, 157201 (2022).

[201] S. Russell and P. Norvig, *Artificial Intelligence, Global Edition*, Prentice Hall series in artificial intelligence, fourth edition, global edition ed. Pearson Education, Boston, ISBN 978-1-292-40117-1.

[202] P. E. Hart, N. J. Nilsson, and B. Raphael, *A Formal Basis for the Heuristic Determination of Minimum Cost Paths*, IEEE Trans. Syst. Sci. Cybern., **4**, 100 (1968).

[203] R. E. Fikes and N. J. Nilsson, *Strips: A new approach to the application of theorem proving to problem solving*, Artif. Intell., **2**, 189 (1971).

[204] E. H. Shortliffe, *Mycin: A Knowledge-Based Computer Program Applied to Infectious Diseases*, Proc. Annu. Symp. Comput. Appl. Med. Care, 66–69 (1977).

[205] M. Minsky and S. A. Papert, *Perceptrons: An Introduction to Computational Geometry*. The MIT Press, ISBN 978-0-262-34393-0, doi:10.7551/mitpress/11301.001.0001.

[206] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning representations by back-propagating errors*, Nature, **323**, 533 (1986).

[207] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, ISBN 978-1-55860-479-7.

[208] L. Rabiner, *A tutorial on hidden Markov models and selected applications in speech recognition*, Proc. IEEE, **77**, 257 (1989).

[209] A. P. Dempster, N. M. Laird, and D. B. Rubin, *Maximum Likelihood from Incomplete Data via the EM Algorithm*, J. R. Stat. Soc., B (methodol.), **39**, 1 (1977).

[210] C. Cortes and V. Vapnik, *Support-vector networks*, Mach. Learn., **20**, 273 (1995).

[211] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press, ISBN 978-0-262-25693-3, doi:10.7551/mitpress/4175.001.0001.

[212] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, ISBN 978-0-262-25683-4, doi:10.7551/mitpress/3206.001.0001.

[213] S. F. Chen and J. Goodman, *An empirical study of smoothing techniques for language modeling*, Comput. Speech Lang., **13**, 359 (1999).

[214] A. L. Berger, V. J. D. Pietra, and S. A. D. Pietra, *A maximum entropy approach to natural language processing*, Comput, Linguist,, **22**, 39 (1996).

[215] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, *Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data*, in *Proceedings of the Eighteenth International Conference on Machine Learn-*

ing, ICML '01, 282–289, Morgan Kaufmann Publishers Inc., ISBN 978-1-55860-778-1.

[216] D. M. Blei, A. Y. Ng, and M. I. Jordan, *Latent dirichlet allocation*, J. Mach, Learn, Res,, **3**, 993 (2003).

[217] D. G. Lowe, *Distinctive Image Features from Scale-Invariant Keypoints*, Int. J. Comput. Vision, **60**, 91 (2004).

[218] N. Dalal and B. Triggs, *Histograms of oriented gradients for human detection*, in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, 886–893 vol. 1, ISSN 1063-6919, doi: 10.1109/CVPR.2005.177.

[219] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, *Gradient-based learning applied to document recognition*, Proc. IEEE, **86**, 2278 (1998).

[220] G. E. Hinton and R. R. Salakhutdinov, *Reducing the Dimensionality of Data with Neural Networks*, Science, **313**, 504 (2006).

[221] G. E. Hinton, S. Osindero, and Y.-W. Teh, *A Fast Learning Algorithm for Deep Belief Nets*, Neural Comput., **18**, 1527 (2006).

[222] Y. Bengio, R. Ducharme, P. Vincent, and C. Janvin, *A neural probabilistic language model*, J. Mach, Learn, Res,, **3**, 1137 (2003).

[223] T. Mikolov, W.-t. Yih, and G. Zweig, *Linguistic Regularities in Continuous Space Word Representations*, in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (edited by L. Vanderwende, H. Daumé III, and K. Kirchhoff), 746–751, Association for Computational Linguistics.

[224] T. Mikolov, K. Chen, G. Corrado, and J. Dean, *Efficient Estimation of Word Representations in Vector Space*, in *International Conference on Learning Representations*.

[225] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *ImageNet Classification with Deep Convolutional Neural Networks*, in *Advances in Neural Information Processing Systems*, vol. 25, Curran Associates, Inc.

[226] S. Hochreiter and J. Schmidhuber, *Long Short-Term Memory*, Neural Comput., **9**, 1735 (1997).

[227] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, *Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation*, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (edited by A. Moschitti, B. Pang, and W. Daelemans), 1724–1734, Association for Computational Linguistics, doi:10.3115/v1/D14-1179.

[228] I. Sutskever, O. Vinyals, and Q. V. Le, *Sequence to sequence learning with neural networks*, in *Proceedings of the 28th International Conference on Neu-*

ral Information Processing Systems - Volume 2, *NIPS'14*, vol. 2, 3104–3112, MIT Press.

[229] J. Pennington, R. Socher, and C. Manning, *GloVe: Global Vectors for Word Representation*, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (edited by A. Moschitti, B. Pang, and W. Daelemans), 1532–1543, Association for Computational Linguistics, doi:10.3115/v1/D14-1162.

[230] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, *Deep Contextualized Word Representations*, in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (long Papers)* (edited by M. Walker, H. Ji, and A. Stent), 2227–2237, Association for Computational Linguistics, doi:10.18653/v1/N18-1202.

[231] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. Bowman, *GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding*, in *Proceedings of the 2018 EMNLP Workshop Blackboxnlp: Analyzing and Interpreting Neural Networks for NLP* (edited by T. Linzen, G. Chrupała, and A. Alishahi), 353–355, Association for Computational Linguistics, doi:10.18653/v1/W18-5446.

[232] X. Du, M. Liu, K. Wang, H. Wang, J. Liu, Y. Chen, J. Feng, C. Sha, X. Peng, and Y. Lou, *Evaluating Large Language Models in Class-Level Code Generation*, in *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*, ICSE '24, 1–13, Association for Computing Machinery, ISBN 979-8-4007-0217-4, doi:10.1145/3597503.3639219.

[233] J. Wei, M. Bosma, V. Zhao, K. Guu, A. W. Yu, B. Lester, N. Du, A. M. Dai, and Q. V. Le, *Finetuned Language Models Are Zero-Shot Learners*, in *International Conference on Learning Representations Event*.

[234] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F. Kelton, L. Miller, M. Simens, A. Askell, P. Welinder, P. Christiano, J. Leike, and R. Lowe, *Training language models to follow instructions with human feedback*, in *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, 27730–27744, Curran Associates Inc., ISBN 978-1-7138-7108-8.

[235] P. F. Christiano, J. Leike, T. B. Brown, M. Martic, S. Legg, and D. Amodei, *Deep reinforcement learning from human preferences*, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, 4302–4310, Curran Associates Inc., ISBN 978-1-5108-6096-4.

[236] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. H. Chi, Q. V. Le, and D. Zhou, *Chain-of-thought prompting elicits reasoning in large language models*, in *Proceedings of the 36th International Conference on Neural*

Information Processing Systems, NIPS '22, 24824–24837, Curran Associates Inc., ISBN 978-1-7138-7108-8.

[237] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. Narasimhan, and Y. Cao, *ReAct: Synergizing Reasoning and Acting in Language Models*, in *International Conference on Learning Representations*, vol. abs/2210.03629, arXiv, doi: 10.48550/arXiv.2210.03629, 2210.03629.

[238] T. Schick, J. Dwivedi-Yu, R. Dessí, R. Raileanu, M. Lomeli, E. Hambro, L. Zettlemoyer, N. Cancedda, and T. Scialom, *Toolformer: language models can teach themselves to use tools*, in *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, 68539–68551, Curran Associates Inc.

[239] L. Gao, A. Madaan, S. Zhou, U. Alon, P. Liu, Y. Yang, J. Callan, and G. Neubig, *PAL: program-aided language models*, in *Proceedings of the 40th International Conference on Machine Learning*, *ICML'23*, vol. 202, 10764–10799, JMLR.org.

[240] K. Guu, K. Lee, Z. Tung, P. Pasupat, and M.-W. Chang, *REALM: retrieval-augmented language model pre-training*, in *Proceedings of the 37th International Conference on Machine Learning*, *ICML'20*, vol. 119, 3929–3938, JMLR.org.

[241] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, *Retrieval-augmented generation for knowledge-intensive NLP tasks*, in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, 9459–9474, Curran Associates Inc., ISBN 978-1-7138-2954-6.

[242] S. Borgeaud, A. Mensch, J. Hoffmann, T. Cai, E. Rutherford, K. Millican, G. van den Driessche, J.-B. Lespiau, B. Damoc, A. Clark, D. d. L. Casas, A. Guy, J. Menick, R. Ring, T. Hennigan, S. Huang, L. Maggiore, C. Jones, A. Cassirer, A. Brock, M. Paganini, G. Irving, O. Vinyals, S. Osindero, K. Simonyan, J. W. Rae, E. Elsen, and L. Sifre, *Improving language models by retrieving from trillions of tokens*, in *International Conference on Machine Learning*.

[243] K. He, X. Zhang, S. Ren, and J. Sun, *Deep Residual Learning for Image Recognition*, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778, ISSN 1063-6919, doi:10.1109/CVPR.2016.90.

[244] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative adversarial networks*, Commun, ACM, **63**, 139 (2020).

[245] J. Ho, A. Jain, and P. Abbeel, *Denoising diffusion probabilistic models*, in *Proceedings of the 34th International Conference on Neural Information Pro-*

cessing Systems, NIPS '20, 6840–6851, Curran Associates Inc., ISBN 978-1-7138-2954-6.

[246] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, *High-Resolution Image Synthesis with Latent Diffusion Models*, in *Computer Vision and Pattern Recognition*, arXiv, doi:10.1109/CVPR52688.2022.01042, 2112.10752.

[247] Y. Song, J. N. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, *Score-Based Generative Modeling through Stochastic Differential Equations*, in *International Conference on Learning Representations*, doi: 10.48550/arXiv.2011.13456.

[248] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, *LoRA: Low-Rank Adaptation of Large Language Models*, in *International Conference on Learning Representations*.

[249] M. Zhou, Y. Bai, Q. Yang, and T. Zhao, *StyleInject: Parameter Efficient Tuning of Text-to-Image Diffusion Models*, ACM Trans, Multimed. Comput, Commun, Appl,, **21**, 152:1 (2025).

[250] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, *Learning Transferable Visual Models From Natural Language Supervision*, in *International Conference on Machine Learning*.

[251] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, *Neural message passing for Quantum chemistry*, in *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, 1263–1272, JMLR.org.

[252] K. T. Schütt, P.-J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, and K.-R. Müller, *SchNet: a continuous-filter convolutional neural network for modeling quantum interactions*, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, 992–1002, Curran Associates Inc., ISBN 978-1-5108-6096-4.

[253] L. Zhang, D.-Y. Lin, H. Wang, R. Car, and W. E, *Active learning of uniformly accurate interatomic potentials for materials simulation*, Phys. Rev. Materials, **3**, 023804 (2019).

[254] B. Sanchez-Lengeling and A. Aspuru-Guzik, *Inverse molecular design using machine learning: Generative models for matter engineering*, Science, **361**, 360 (2018).

[255] Y. Yang, J. Sun, H. Li, and Z. Xu, *Deep ADMM-Net for Compressive Sensing MRI*, in *Advances in Neural Information Processing Systems*, vol. 29, Curran Associates, Inc.

[256] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, *Deep Convolutional Neural Network for Inverse Problems in Imaging*, IEEE Trans. Image Process.,

26, 4509 (2017).

[257] D. Ulyanov, A. Vedaldi, and V. Lempitsky, *Deep Image Prior*, Int. J. Comput. Vision, **128**, 1867 (2020).

[258] M. Paganini, L. De Oliveira, and B. Nachman, *CaloGAN: Simulating 3D high energy particle showers in multilayer electromagnetic calorimeters with generative adversarial networks*, Phys. Rev. D, **97**, 014021 (2018).

[259] C. Biscarat, S. Caillou, C. Rougier, J. Stark, and J. Zahreddine, *Towards a realistic track reconstruction algorithm based on graph neural networks for the HL-LHC*, EPJ Web Conf., **251**, 3047 (2021).

[260] D. George and E. A. Huerta, *Deep Learning for real-time gravitational wave detection and parameter estimation: Results with Advanced LIGO data*, Phys. Lett. B, **778**, 64 (2018).

[261] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, *Real-Time Gravitational Wave Science with Neural Posterior Estimation*, Phys. Rev. Lett., **127**, 241103 (2021).

[262] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, *Fourier Neural Operator for Parametric Partial Differential Equations*, in *International Conference on Learning Representations*, doi:10.48550/arXiv.2010.08895, 2010.08895.

[263] L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis, *Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators*, Nat. Mach. Intell., **3**, 218 (2021).

[264] M. Chen, J. Tworek, H. Jun, Q. Yuan, H. Pondé, J. Kaplan, H. Edwards, Y. Burda, N. Joseph, G. Brockman, A. Ray, R. Puri, G. Krueger, M. Petrov, H. Khlaaf, G. Sastry, P. Mishkin, B. Chan, S. Gray, N. Ryder, M. Pavlov, A. Power, L. Kaiser, M. Bavarian, C. Winter, P. Tillet, F. Such, D. Cummings, M. Plappert, F. Chantzis, E. Barnes, A. Herbert-Voss, W. H. Guss, A. Nichol, I. Babuschkin, S. Balaji, S. Jain, A. Carr, J. Leike, J. Achiam, V. Misra, E. Morikawa, A. Radford, M. Knight, M. Brundage, M. Murati, K. Mayer, P. Welinder, B. McGrew, D. Amodei, S. McCandlish, I. Sutskever, and W. Zaremba, *Evaluating Large Language Models Trained on Code*, (2021).

[265] L. Wang, C. Ma, X. Feng, Z. Zhang, H. Yang, J. Zhang, Z. Chen, J. Tang, X. Chen, Y. Lin, W. X. Zhao, Z. Wei, and J. Wen, *A survey on large language model based autonomous agents*, Front. Comput. Sci., **18**, 186345 (2024).

[266] B. P. MacLeod, F. G. L. Parlane, T. D. Morrissey, F. Häse, L. M. Roch, K. E. Dettelbach, R. Moreira, L. P. E. Yunker, M. B. Rooney, J. R. Deeth, V. Lai, G. J. Ng, H. Situ, R. H. Zhang, M. S. Elliott, T. H. Haley, D. J. Dvorak, A. Aspuru-Guzik, J. E. Hein, and C. P. Berlinguette, *Self-driving laboratory for accelerated discovery of thin-film materials*, Sci. Adv., **6**, eaaz8867 (2020).

[267] R. Balestriero, M. Ibrahim, V. Sobal, A. Morcos, S. Shekhar, T. Goldstein, F. Bordes, A. Bardes, G. Mialon, Y. Tian, A. Schwarzschild, A. G. Wilson, J. Geiping, Q. Garrido, P. Fernandez, A. Bar, H. Pirsiavash, Y. LeCun, and M. Goldblum, *A Cookbook of Self-Supervised Learning*, (2023), doi: 10.48550/arXiv.2304.12210.

[268] J. Gui, T. Chen, J. Zhang, Q. Cao, Z. Sun, H. Luo, and D. Tao, *A Survey on Self-Supervised Learning: Algorithms, Applications, and Future Trends*, IEEE Trans. Pattern Anal. Mach. Intell., **46**, 9052 (2024).

[269] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei, *Scaling Laws for Neural Language Models*, (2020), doi:10.48550/arXiv.2001.08361.

[270] J. Hoffmann, S. Borgeaud, A. Mensch, E. Buchatskaya, T. Cai, E. Rutherford, D. de Las Casas, L. A. Hendricks, J. Welbl, A. Clark, T. Hennigan, E. Noland, K. Millican, G. van den Driessche, B. Damoc, A. Guy, S. Osindero, K. Simonyan, E. Elsen, O. Vinyals, J. W. Rae, and L. Sifre, *Training compute-optimal large language models*, in *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, 30016–30030, Curran Associates Inc., ISBN 978-1-7138-7108-8.

[271] Y. Gal and Z. Ghahramani, *Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning*, in *Proceedings of the 33rd International Conference on Machine Learning*, 1050–1059, PMLR, ISSN 1938-7228.

[272] A. Kendall and Y. Gal, *What uncertainties do we need in Bayesian deep learning for computer vision?*, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, 5580–5590, Curran Associates Inc., ISBN 978-1-5108-6096-4.

[273] B. Recht, R. Roelofs, L. Schmidt, and V. Shankar, *Do ImageNet Classifiers Generalize to ImageNet?*, in *International Conference on Machine Learning*.

[274] A. Wang, Y. Pruksachatkun, N. Nangia, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman, *SuperGLUE: a stickier benchmark for general-purpose language understanding systems*, in *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 3266–3280, Curran Associates Inc.

[275] J. Carrasquilla, *Machine learning for quantum matter*, Adv. Phys.: X, **5**, 1797528 (2020).

[276] J. Carrasquilla and R. G. Melko, *Machine learning phases of matter*, Nat. Phys., **13**, 431 (2017).

[277] L. Wang, *Discovering phase transitions with unsupervised learning*, Phys. Rev. B, **94**, 195105 (2016).

[278] W. Hu, R. R. P. Singh, and R. T. Scalettar, *Discovering phases, phase transitions, and crossovers through unsupervised machine learning: A critical examination*, Phys. Rev. E, **95**, 062122 (2017).

[279] E. P. L. van Nieuwenburg, Y.-H. Liu, and S. D. Huber, *Learning phase transitions by confusion*, Nat. Phys., **13**, 435 (2017).

[280] J. F. Rodriguez-Nieva and M. S. Scheurer, *Identifying topological order through unsupervised machine learning*, Nat. Phys., **15**, 790 (2019).

[281] G. Carleo and M. Troyer, *Solving the quantum many-body problem with artificial neural networks*, Science, **355**, 602 (2017).

[282] E. Ibarra-García-Padilla, *Autoregressive neural quantum states of Fermi Hubbard models*, Phys. Rev. Res., **7** (2025).

[283] T. A. Kumar, L. Balents, T. H. Hsieh, and R. G. Melko, *Autoregressive Typical Thermal States*, (2025), doi:10.48550/arXiv.2508.13455.

[284] D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, Ab initio *solution of the many-electron Schrödinger equation with deep neural networks*, Phys. Rev. Research, **2**, 033429 (2020).

[285] J. Hermann, Z. Schätzle, and F. Noé, *Deep-neural-network solution of the electronic Schrödinger equation*, Nat. Chem., **12**, 891 (2020).

[286] Z. Chen, L. Newhouse, E. Chen, D. Luo, and M. Soljačić, *ANTN: bridging autoregressive neural networks and tensor networks for quantum many-body simulation*, in *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, 450–476, Curran Associates Inc.

[287] K. Temme, S. Bravyi, and J. M. Gambetta, *Error Mitigation for Short-Depth Quantum Circuits*, Phys. Rev. Lett., **119**, 180509 (2017).

[288] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, *Quantum error mitigation*, Rev. Mod. Phys., **95**, 045005 (2023).

[289] V. R. Pascuzzi, A. He, C. W. Bauer, W. A. De Jong, and B. Nachman, *Computationally efficient zero-noise extrapolation for quantum-gate-error mitigation*, Phys. Rev. A, **105**, 042406 (2022).

[290] A. Strikis, D. Qin, Y. Chen, S. C. Benjamin, and Y. Li, *Learning-Based Quantum Error Mitigation*, PRX Quantum, **2**, 040330 (2021).

[291] C. Kim, K. D. Park, and J.-K. Rhee, *Quantum Error Mitigation With Artificial Neural Network*, IEEE Access, **8**, 188853 (2020).

[292] D. Babukhin, *Echo-evolution data generation for quantum error mitigation via neural networks*, Quantum Inf. Process., **23**, 405 (2024).

[293] M. Liao, Y. Zhu, G. Chiribella, and Y. Yang, *Noise-agnostic quantum error mitigation with data augmented neural models*, npj Quantum Inf., **11**, 8 (2025).

[294] I. Cong, S. Choi, and M. D. Lukin, *Quantum convolutional neural networks*, Nat. Phys., **15**, 1273 (2019).

[295] P. Zapletal, N. A. McMahon, and M. J. Hartmann, *Error-tolerant quantum convolutional neural networks for symmetry-protected topological phases*, Phys. Rev. Res., **6**, 33111 (2024).

[296] J. Hubbard, *Electron correlations in narrow energy bands. II. The degenerate band case*, Proc. R. Soc. Lond., A. Math. Phys. Sci., **277**, 237 (1997).

[297] M. C. Gutzwiller, *Effect of Correlation on the Ferromagnetism of Transition Metals*, Phys. Rev. Lett., **10**, 159 (1963).

[298] J. Kanamori, *Electron Correlation and Ferromagnetism of Transition Metals*, Prog. Theor. Phys., **30**, 275 (1963).

[299] N. F. Mott, *The Basis of the Electron Theory of Metals, with Special Reference to the Transition Metals*, Proc. Phys. Soc. London, Sect. A, **62**, 416 (1949).

[300] D. B. McWhan, A. Menth, J. P. Remeika, W. F. Brinkman, and T. M. Rice, *Metal-Insulator Transitions in Pure and Doped V_2O_3*, Phys. Rev. B, **7**, 1920 (1973).

[301] M. Imada, A. Fujimori, and Y. Tokura, *Metal-insulator transitions*, Rev. Mod. Phys., **70**, 1039 (1998).

[302] M. Hirayama, Y. Yamaji, T. Misawa, and M. Imada, *Ab initio effective Hamiltonians for cuprate superconductors*, Phys. Rev. B, **98**, 134501 (2018).

[303] M. Hirayama, T. Misawa, T. Ohgoe, Y. Yamaji, and M. Imada, *Effective Hamiltonian for cuprate superconductors derived from multiscale ab initio scheme with level renormalization*, Phys. Rev. B, **99**, 245155 (2019).

[304] B. Ponsioen, S. S. Chung, and P. Corboz, *Period 4 stripe in the extended two-dimensional Hubbard model*, Phys. Rev. B, **100**, 195141 (2019).

[305] H. Xu, H. Shi, E. Vitali, M. Qin, and S. Zhang, *Stripes and Spin-Density Waves in the Doped Two-Dimensional Hubbard Model: Ground State Phase Diagram*, Phys. Rev. Res., **4**, 13239 (2022).

[306] J. R. Schrieffer and P. A. Wolff, *Relation between the Anderson and Kondo Hamiltonians*, Phys. Rev., **149**, 491 (1966).

[307] K. A. Chao, J. Spalek, and A. M. Oles, *Kinetic exchange interaction in a narrow S-band*, J. Phys. C: Solid State Phys., **10**, L271 (1977).

[308] J. Spałek, *t-J Model Then and Now: a Personal Perspective from the Pioneering Times*, Acta Phys. Pol. A, **111**, 409 (2007).

[309] I. L. Markov and Y. Shi, *Simulating Quantum Computation by Contracting Tensor Networks*, SIAM J. Comput., **38**, 963 (2008).

[310] D. G. a Smith and J. Gray, *opt_einsum - A Python package for optimizing contraction order for einsum-like expressions*, J. Open Source Softw., **3**, 753 (2018).

[311] J. Gray and S. Kourtis, *Hyper-optimized tensor network contraction*, Quantum, **5**, 410 (2021).

[312] A. Weichselbaum, *QSpace - An Open-Source Tensor Library for Abelian and Non-Abelian Symmetries*, SciPost Phys. Codebases, 40 (2024).

[313] A. Weichselbaum, *Open Source QSpace 4.0*, https://bitbucket.org/qspace4u.

[314] A. Weichselbaum, *Non-abelian symmetries in tensor networks: A quantum symmetry space approach*, Ann. Phys., **327**, 2972 (2012).

[315] B. Bruognolo, *Tensor network techniques for strongly correlated systems*, Ph.D. thesis (2017).

[316] A. Weichselbaum, *X-symbols for non-Abelian symmetries in tensor networks*, Phys. Rev. Res., **2**, 023385 (2020).

[317] A. Alex, M. Kalus, A. Huckleberry, and J. von Delft, *A numerical algorithm for the explicit calculation of SU(N) and SL(N,$\mathbb{C}$) Clebsch-Gordan coefficients*, Arxiv (2010).

[318] L. Devos and J. Haegeman, *TensorKit.jl: A Julia package for large-scale tensor computations, with a hint of category theory*, (2025), doi:10.48550/arXiv.2508.10076.

[319] R. Alkabetz and I. Arad, *Tensor networks contraction and the belief propagation algorithm*, Phys. Rev. Res., **3**, 23073 (2021).

[320] C. Guo, D. Poletti, and I. Arad, *Block belief propagation algorithm for two-dimensional tensor networks*, Phys. Rev. B, **108**, 125111 (2023).

[321] J. Tindall and M. Fishman, *Gauging tensor networks with belief propagation*, Scipost Phys., **15**, 222 (2023).

[322] Y. Wang, Y. E. Zhang, F. Pan, and P. Zhang, *Tensor Network Message Passing*, Phys. Rev. Lett., **132**, 117401 (2024).

[323] A. Gleis, J.-W. Li, and J. von Delft, *Projector formalism for kept and discarded spaces of matrix product states*, Phys. Rev. B, **106**, 195138 (2022).

[324] J.-W. Li, A. Gleis, and J. von Delft, *Time-Dependent Variational Principle with Controlled Bond Expansion for Matrix Product States*, Phys. Rev. Lett., **133**, 26401 (2024).

[325] C. Zhang, *Symmetric Infinite Projected Entangled-Pair State Study of Quantum Lattice Models*, Master's thesis, Ludwig-Maximilians-Universität München, (2021).

[326] M. S. Moss, S. Ebadi, T. T. Wang, G. Semeghini, A. Bohrdt, M. D. Lukin, and R. G. Melko, *Enhancing variational Monte Carlo simulations using a programmable quantum simulator*, Phys. Rev. A, **109**, 032410 (2024).

[327] H. N. Phien, I. P. McCulloch, and G. Vidal, *Fast convergence of imaginary time evolution tensor network algorithms by recycling the environment*, Phys. Rev. B, **91**, 115137 (2015).

[328] M. Lubasch, J. I. Cirac, and M.-C. Bañuls, *Algorithms for finite projected entangled pair states*, Phys. Rev. B, **90**, 064425 (2014).

[329] J. Naumann, E. L. Weerda, M. Rizzi, J. Eisert, and P. Schmoll, *An introduction to infinite projected entangled-pair state methods for variational ground state simulations using automatic differentiation*, Scipost Phys. Lect. Notes, 86 (2024).

[330] D. I. Nikolaidou, *Variational optimization of ground-state iPEPS via automatic differentiation*, Master's thesis, Ludwig-Maximilians-Universität München, (2024).

[331] E. L. Weerda and M. Rizzi, *Fractional quantum Hall states with variational projected entangled-pair states: A study of the bosonic Harper-Hofstadter model*, Phys. Rev. B, **109**, L241117 (2024).

[332] P. Czarnik, M. M. Rams, and J. Dziarmaga, *Variational tensor network renormalization in imaginary time: Benchmark results in the Hubbard model at finite temperature*, Phys. Rev. B, **94**, 235142 (2016).

[333] M. Suzuki, *Decomposition formulas of exponential operators and Lie exponentials with some applications to quantum mechanics and statistical physics*, J. Math. Phys., **26**, 601 (1985).

[334] G. Vidal, *Efficient Classical Simulation of Slightly Entangled Quantum Computations*, Phys. Rev. Lett., **91**, 147902 (2003).

[335] G. Vidal, *Classical Simulation of Infinite-Size Quantum Lattice Systems in One Spatial Dimension*, Phys. Rev. Lett., **98**, 070201 (2007).

[336] B. Lim, E. Bellec, M. Dupraz, S. Leake, A. Resta, A. Coati, M. Sprung, E. Almog, E. Rabkin, T. Schulli, and M.-I. Richard, *A convolutional neural network for defect classification in Bragg coherent X-ray diffraction*, npj Comput. Mater., **7**, 115 (2021).

[337] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, *SchNet – A deep learning architecture for molecules and materials*, J. Chem. Phys., **148**, 241722 (2018).

[338] H. Wang, L. Zhang, J. Han, and W. E, *DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics*, Comput. Phys. Commun., **228**, 178 (2018).

[339] P. Minch, R. Bhattarai, K. Choudhary, and T. D. Rhone, *Predicting magnetic properties of van der Waals magnets using graph neural networks*, Phys. Rev. Mater., **8**, 114002 (2024).

[340] T. Flöss, W. R. Coulton, A. J. Duivenvoorden, F. Villaescusa-Navarro, and B. D. Wandelt, *Denoising diffusion delensing: reconstructing the non-Gaussian CMB lensing potential with diffusion models*, Mon. Not. R. Astron. Soc., **533**, 423 (2024).

[341] R. Legin, M. Isi, K. W. K. Wong, Y. Hezaveh, and L. Perreault-Levasseur, *Gravitational-Wave Parameter Estimation in non-Gaussian noise using Score-Based Likelihood Characterization*, (2024), doi:10.48550/arXiv.2410.19956.

[342] O. Press and L. Wolf, *Using the Output Embedding to Improve Language Models*, in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers* (edited by M. Lapata, P. Blunsom, and A. Koller), 157–163, Association for Computational Linguistics.

[343] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, L. Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. R. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes, and J. Dean, *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*, (2016).

[344] A. Graves, *Generating Sequences With Recurrent Neural Networks*, (2014), doi:10.48550/arXiv.1308.0850.

[345] A. Fan, M. Lewis, and Y. Dauphin, *Hierarchical Neural Story Generation*, in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (volume 1: Long Papers)* (edited by I. Gurevych and Y. Miyao), 889–898, Association for Computational Linguistics, doi:10.18653/v1/P18-1082.

[346] A. Holtzman, J. Buys, L. Du, M. Forbes, and Y. Choi, *The Curious Case of Neural Text Degeneration*, in *International Conference on Learning Representations*.

[347] L. Dong, N. Yang, W. Wang, F. Wei, X. Liu, Y. Wang, J. Gao, M. Zhou, and H.-W. Hon, *Unified language model pre-training for natural language understanding and generation*, in *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 1170. pages 13063–13075. Curran Associates Inc., Red Hook, NY, USA 2019.

[348] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, *Language Models are Unsupervised Multitask Learners*.

[349] K. Clark, U. Khandelwal, O. Levy, and C. D. Manning, *What Does BERT Look at? An Analysis of BERT's Attention*, in *Proceedings of the 2019 ACL*

Workshop Blackboxnlp: Analyzing and Interpreting Neural Networks for NLP (edited by T. Linzen, G. Chrupała, Y. Belinkov, and D. Hupkes), 276–286, Association for Computational Linguistics, doi:10.18653/v1/W19-4828.

[350] M. Caron, H. Touvron, I. Misra, H. Jegou, J. Mairal, P. Bojanowski, and A. Joulin, *Emerging Properties in Self-Supervised Vision Transformers*, in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 9630–9640, ISSN 2380-7504, doi:10.1109/ICCV48922.2021.00951.

[351] J. Vig, A. Madani, L. R. Varshney, C. Xiong, R. Socher, and N. F. Rajani, *BERTology Meets Biology: Interpreting Attention in Protein Language Models*, (2020), doi:10.1101/2020.06.26.174417.

[352] S. Serrano and N. A. Smith, *Is Attention Interpretable?*, in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (edited by A. Korhonen, D. Traum, and L. Màrquez), 2931–2951, Association for Computational Linguistics, doi:10.18653/v1/P19-1282.

[353] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, and B. Kim, *Sanity Checks for Saliency Maps*, in *Advances in Neural Information Processing Systems*, vol. 31, Curran Associates, Inc.

[354] P. Michel, O. Levy, and G. Neubig, *Are Sixteen Heads Really Better than One?*, in *Advances in Neural Information Processing Systems*, vol. 32, Curran Associates, Inc.

[355] S. Jain and B. C. Wallace, *Attention is not Explanation*, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (long and Short Papers)* (edited by J. Burstein, C. Doran, and T. Solorio), 3543–3556, Association for Computational Linguistics, doi:10.18653/v1/N19-1357.

[356] D. Pruthi, M. Gupta, B. Dhingra, G. Neubig, and Z. C. Lipton, *Learning to Deceive with Attention-Based Explanations*, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (edited by D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault), 4782–4793, Association for Computational Linguistics, doi:10.18653/v1/2020.acl-main.432.

[357] S. Wiegreffe and Y. Pinter, *Attention is not not Explanation*, in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (edited by K. Inui, J. Jiang, V. Ng, and X. Wan), 11–20, Association for Computational Linguistics, doi:10.18653/v1/D19-1002.

[358] P. Madhyastha and R. Jain, *On Model Stability as a Function of Random Seed*, in *Proceedings of the 23rd Conference on Computational Natural Language Learning (conll)* (edited by M. Bansal and A. Villavicencio), 929–939, Association for Computational Linguistics, doi:10.18653/v1/K19-1087.

[359] A. Bibal, R. Cardon, D. Alfter, R. Wilkens, X. Wang, T. François, and P. Watrin, *Is Attention Explanation? An Introduction to the Debate*, in *Proceed-*

ings of the 60th Annual Meeting of the Association for Computational Linguistics (volume 1: Long Papers) (edited by S. Muresan, P. Nakov, and A. Villavicencio), 3889–3900, Association for Computational Linguistics, doi:10.18653/v1/2022.acl-long.269.

[360] J. Strout, Y. Zhang, and R. Mooney, *Do Human Rationales Improve Machine Explanations?*, in *Proceedings of the 2019 ACL Workshop Blackboxnlp: Analyzing and Interpreting Neural Networks for NLP* (edited by T. Linzen, G. Chrupała, Y. Belinkov, and D. Hupkes), 56–62, Association for Computational Linguistics, doi:10.18653/v1/W19-4807.

[361] R. Zhong, S. Shao, and K. McKeown, *Fine-grained Sentiment Analysis with Faithful Attention*, (2019), doi:10.48550/arXiv.1908.06870.

[362] E. Voita, D. Talbot, F. Moiseev, R. Sennrich, and I. Titov, *Analyzing Multi-Head Self-Attention: Specialized Heads Do the Heavy Lifting, the Rest Can Be Pruned*, in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (edited by A. Korhonen, D. Traum, and L. Màrquez), 5797–5808, Association for Computational Linguistics, doi: 10.18653/v1/P19-1580.

[363] G. Xiao, Y. Tian, B. Chen, S. Han, and M. Lewis, *Efficient Streaming Language Models with Attention Sinks*, in *The Twelfth International Conference on Learning Representations*.

[364] S. Abnar and W. Zuidema, *Quantifying Attention Flow in Transformers*, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (edited by D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault), 4190–4197, Association for Computational Linguistics, doi: 10.18653/v1/2020.acl-main.385.

[365] T. Yuan, X. Li, H. Xiong, H. Cao, and D. Dou, *Explaining Information Flow Inside Vision Transformers Using Markov Chain*, in *eXplainable AI approaches for debugging and diagnosis.*

[366] L. Xu, X. Yan, W. Ding, and Z. Liu, *Attribution rollout: a new way to interpret visual transformer*, J. Ambient Intell. Hum. Comput., **14**, 163 (2023).

[367] B. Azarkhalili and M. W. Libbrecht, *Generalized Attention Flow: Feature Attribution for Transformer Models via Maximum Flow*, in *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (volume 1: Long Papers)* (edited by W. Che, J. Nabende, E. Shutova, and M. T. Pilehvar), 19954–19974, Association for Computational Linguistics, ISBN 979-8-89176-251-0, doi:10.18653/v1/2025.acl-long.980.

[368] M. Sundararajan, A. Taly, and Q. Yan, *Axiomatic attribution for deep networks*, in *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, 3319–3328, JMLR.org.

[369] H. Chefer, S. Gur, and L. Wolf, *Transformer Interpretability Beyond Attention Visualization*, in *2021 IEEE/CVF Conference on Computer Vision*

and Pattern Recognition (CVPR), 782–791, ISSN 2575-7075, doi:10.1109/CVPR46437.2021.00084.

[370] V. Petsiuk, A. Das, and K. Saenko, *RISE: Randomized Input Sampling for Explanation of Black-box Models*, in *British Machine Vision Conference*, vol. abs/1806.07421.

[371] S. Hooker, D. Erhan, P.-J. Kindermans, and B. Kim, *A benchmark for interpretability methods in deep neural networks*, in *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 9737–9748, Curran Associates Inc.

[372] F. Chen, F. D. M. Haldane, and D. N. Sheng, *Global phase diagram of D-wave superconductivity in the square-lattice t-J model*, Proc. Natl. Acad. Sci., **122**, e2420963122 (2025).

[373] W. O. Wang and T. P. Devereaux, *Finite-temperature signatures of underlying superconductivity in the electron-doped Hubbard model*, (2025), doi:10.48550/arXiv.2510.16616.

[374] S. Gong, W. Zhu, and D. N. Sheng, *Robust d-wave Superconductivity in the Square-Lattice t-J Model*, Phys. Rev. Lett., **127**, 097003 (2021).

[375] S. Jiang, D. J. Scalapino, and S. R. White, *Ground-state phase diagram of the t-t'-J model*, Proc. Natl. Acad. Sci. U.S.A., **118**, e2109978118 (2021).

[376] H. Lange, A. Van de Walle, A. Abedinnia, and A. Bohrdt, *From architectures to applications: a review of neural quantum states*, Quantum Sci. Technol., **9**, 40501 (2024).

[377] Z. Denis and G. Carleo, *Accurate neural quantum states for interacting lattice bosons*, Quantum, **9**, 1772 (2025).

[378] D. Luo and B. K. Clark, *Backflow Transformations via Neural Networks for Quantum Many-Body Wave Functions*, Phys. Rev. Lett., **122**, 226401 (2019).

[379] A.-J. Liu and B. K. Clark, *Neural network backflow for ab initio quantum chemistry*, Phys. Rev. B, **110**, 115137 (2024).

[380] L. Zhang and D. Luo, *Neural Transformer Backflow for Solving Momentum-Resolved Ground States of Strongly Correlated Materials*, (2025), doi:10.48550/arXiv.2509.09275.

[381] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carrasquilla, *Recurrent neural network wave functions*, Phys. Rev. Res., **2**, 23358 (2020).

[382] D. Luo, Z. Chen, K. Hu, Z. Zhao, V. M. Hur, and B. K. Clark, *Gauge-invariant and anyonic-symmetric autoregressive neural network for quantum lattice models*, Phys. Rev. Res., **5**, 13216 (2023).

[383] D. Fitzek, Y. H. Teoh, H. P. Fung, G. A. Dagnew, E. Merali, M. S. Moss, B. MacLellan, and R. G. Melko, *RydbergGPT*, (2024), doi:10.48550/arXiv.2405.21052.