Machine Learning Enabled Large-scale Modeling of Functional Electron Materials

Sheng Zhang
Guangzhou, Guangdong, China
B.S. Physics, Sun Yat-sen University, 2016

A Dissertation presented to the Graduate Faculty
of the University of Virginia in Candidacy for the degree of
Doctor of Philosophy

Department of Physics
University of Virginia
November 2024

Professor Gia-Wei Chern
Research Advisor

Professor Israel Klich

Professor Kent Yagi

Professor Stephen Baek
GSAS Representative

Abstract

The emergence of complex, functional electron materials has generated significant interest due to their potential applications in various technologies, including quantum computing and energy storage. However, understanding and modeling these materials present a major challenge because of the complexity of their electronic structure and interactions. This dissertation explores the use of machine learning to enable large-scale modeling of functional electron materials, with a focus on the Falicov-Kimball model, itinerant electron magnets, and disordered spin systems. We leverage neural network architectures to create efficient energy models, thereby enabling simulations that are otherwise computationally prohibitive. Our results demonstrate that ML can capture complex phenomena, such as phase separation and spin dynamics, with high accuracy. This work not only provides insights into the behavior of functional electron materials but also establishes a framework for employing ML in computational physics to solve otherwise intractable problems.

Acknowledgments

I would like to express my deepest gratitude to my advisor, Professor Gia-Wei Chern, for his invaluable guidance, support, and inspiration throughout my PhD journey. His insights, understanding, and steadfast encouragement have been instrumental in shaping my research, guiding me through challenges, and fostering my growth as a physicist. He has always been there to provide insightful feedback and to push me to think beyond the obvious, and for that, I am profoundly thankful.

I also want to extend my heartfelt thanks to my research committee members, Professor Israel Klich and Professor Seunghun Lee, for their valuable advice and for challenging me to think critically and refine my ideas. Their constructive suggestions and expertise have significantly contributed to my research, providing new perspectives and enriching my work. I am also immensely grateful to my defense committee members, Professor Kent Yagi and Professor Stephen Baek, for their time, support, and thoughtful feedback, which made my defense a truly enriching and rewarding experience. Their questions and insights helped me to view my research in a broader context and added depth to my understanding.

To my friends and fellow graduate students, thank you for the camaraderie, the countless discussions, and the constant support that made the challenging times bearable and the successes even sweeter. The late-night study sessions, the brainstorming over coffee, and the shared laughter have made this journey unforgettable. Your en-

couragement and friendship have been a constant source of motivation and strength. I am also deeply grateful to Peter Cline and the physics department support staff for their invaluable assistance, which ensured that everything ran smoothly. Their dedication, efficiency, and willingness to help at all times have made a significant difference, and I truly appreciate their hard work behind the scenes.

I would also like to thank my collaborators, including Puhan Zhang, Preetha Saha, Zhijie Fan, Jing Luo, Chen Cheng, Zhongzheng Tian, Yi-Hsuan Liu, Jianhua Ma, Supriyo Ghosh, Niraj Aryal, Yunhao Fan, Menglin Shi, Xinlun Cheng, Phong Nguyen, Professor David Nichols, Professor Joe Poon, and Nan Jiang. Working alongside each of you has been an incredibly rewarding experience, and your expertise, dedication, and enthusiasm have greatly enriched my research. I am especially grateful to Nan Jiang, who is not only my collaborator but also my partner. Her support, understanding, and insightful contributions have been invaluable throughout this journey, and I could not have asked for a better partner both in life and in research.

Finally, I would like to thank my parents, Shaogang Zhang and Hongjuan Yuan, for their steadfast love, belief in me, and endless support. Their devotion, encouragement, and understanding have been my foundation, and I am forever grateful for everything they have done to help me pursue my dreams. Without their love and guidance, none of this would have been possible. Their belief in me has been a source of inspiration and has given me the strength to persevere through the toughest moments of my

academic journey.

List of Abbreviations

ML	Machine Learning
MC	Monte Carlo
FK	Falicov Kimball
CDW	Charge Density Wave
KMC	Kinetic Monte Carlo
NN	Neural Network
IR	Irreducible Representation
MSE	Mean Squared Error
ED	Exact Diagonalization
LSW	Lifshitz-Slyozov-Wagner
TDGL	Time-Dependent Ginzburg-Landau
DFT	Density Functional Theory
MD	Molecular Dynamics
BP	Behler-Parrinello
CNN	Convolutional Neural Network
LLG	Landau-Lifshitz-Gilbert
PES	Potential Energy Surface
RKKY	Ruderman-Kittel-Kasuya-Yosida
SkL	Skyrmion Lattice
KPM	Kernel Polynomial Method
MLP	Multilayer Perceptron
ReLU	Rectified Linear Unit
BZ	Brillouin Zone
SGD	Stochastic Gradient Descent
ERF	Effective Receptive Field
MLP	Multilayer Perceptron
PDE	Partial Differential Equation
ACSF	Atom-Centered Symmetry Function
DOS	Density of States

DFPT	Density Functional Perturbation Theory
AHC	Allen-Heine-Cardona
MOS-FET	Metal-Oxide Semiconductor Field-Effect Transistors
IMT	Insulator-to-Metal Transition
JT	Jahn-Telle
GR	General Relativity
EdGB	Einstein-dilation Gauss-Bonne
DINGO	Deep Inference for Gravitational-wave Obser- vations
MCMC	Markov Chain Monte Carlo
sGB	Scalar Gauss-Bonnet
BBH	Binary Black Hole
NSBH	Neutron Star Black Hole
SNR	Signal-to-Noise Ratio
PN	Post-Newtonian

Table of Contents

Abstract ii
Acknowledgments iii
List of Abbreviations vi
Table of Contents viii
List of Figures xi
List of Tables xxii

Chapter 1 Introduction 1

Chapter 2 Machine Learning in the Falicov-Kimball Model 7
2.1 Introduction 7
2.2 Kinetic Monte Carlo for the Falicov-Kimball Model 10
2.3 Machine Learning Architecture 12
2.3.1 Lattice Descriptor 13
2.3.2 Convolutional Neural Network 18
2.3.3 Results 21
2.4 Phase Separation in the Falicov-Kimball Model 23
2.5 Conclusion and Outlook 36

Chapter 3 Machine Learning for Itinerant Electron Magnets on Lattice 38
3.1 Introduction 38
3.2 s-d Model for Itinerant Magnets 42
3.3 Machine Learning Architecture 45
3.3.1 Convolutional Neural Network Model 46
3.3.2 Lattice Descriptor 56
3.3.3 Neural Network 63
3.4 Results 65
3.4.1 Dynamics for Convolutional Neural Network Model 67

3.4.2 Dynamics for Neural Network Model with Descriptor 79
3.5 Conclusion and Outlook 83

Chapter 4 Machine Learning for Metallic Spin Glass 85
4.1 Introduction 85
4.2 Adiabatic Dynamics of Itinerant Electron Magnets 89
4.3 Machine Learning Force-Field Model For Disordered Spins 92
4.3.1 Behler-Parrinello Machine Learning Framework 92
4.3.2 Magnetic Descriptor for Disordered Spins 97
4.4 Relaxation Dynamics of an Amorphous s-d Model 101
4.5 Conclusion and Outlook 114

Chapter 5 Machine Learning for Silicon Band Structures 116
5.1 Introduction 116
5.2 Stochastic Formalism and Machine Learning Models 121
5.2.1 *Ab Initio* Phonon Models 121
5.2.2 Stochastic Vibrational Average 123
5.2.3 Machine Learning Model 127
5.3 Descriptors for Frozen Phonon Configurations 129
5.4 Computational Details and Results 136
5.4.1 DFT Calculations 136
5.4.2 Neural Network Architecture and Model Training 139
5.5 Conclusion and Outlook 147

Chapter 6 Nonequilibrium Dynamics in Charge Density Wave Insulators 150
6.1 introduction 150
6.2 Gating Induced Charge Density Wave to Metal Transition 153
6.3 Nonequilibrium Green's Function 156
6.4 Results 157
6.5 Conclusion and Outlook 165

Chapter 7 Descriptors for Lattice Models 167
7.1 Introduction 167
7.2 Descriptor 170
7.2.1 Bispectrum Coefficients 175

7.2.2 Internal Symmetry . . . 183
7.2.3 Atom-Centered Symmetry Functions . . . 189
7.3 Example: Dynamics of Cooperative Jahn-Teller Coupling . . . 194
7.3.1 Machine Learning Force-Field Models . . . 201
7.3.2 Behler-Parrinello Machine-Learning Framework . . . 203
7.3.3 Descriptors for Lattice Distortions . . . 205
7.3.4 Implementation Details and Benchmarks . . . 210
7.3.5 Coarsening of Orbital Order in Cooperative Jahn-Teller Model 217
7.4 Summary and Discussion . . . 227

Appendix 232

Appendix A Details of the Descriptor for displacements on a diamond lattice 232

Appendix B Tests of Gravity with Gravitational-Wave through Machine Learning 238
B.1 introduction . . . 238
B.2 Einstein-dilation Gauss-Bonnet Gravity . . . 241
B.2.1 Theory . . . 241
B.2.2 Gravitational Waveforms . . . 243
B.3 Normalizing Flow . . . 244

List of Figures

2.1 Schematic diagram of NN energy model for kMC dynamics simulation of the FK system. A descriptor is used to construct effective coordinates $\{G_\ell\}$ from neighborhood f-electron configuration $\{n_j^f\}$ up to a cutoff $r_c = 10$. These feature variables $\{G_\ell\}$ are input to the NN which predicts the energy differences ΔE at the output. 13

2.2 The example of a block of the f-electron configuration $\{w_j\}$ that is used to generate block-diagonal representations of the D_4 point group of the square lattice. These eight sites are the 4^{th} nearest neighbor to the ion $w_0 = 1$, where w_j is either occupied (1) or unoccupied (0). . . 16

2.3 Five different nearest-neighbor configurations for updating an f-electron at the central site $n_0^f = 1$. The blue circles denote nearest neighbor sites that are occupied by an f-electron, while the yellow circles denote empty sites. 17

2.4 Correlation between the NN predictions and the exact solutions for energy differences $\Delta E_\rightarrow$, $\Delta E_\leftarrow$, $\Delta E_\uparrow$, and $\Delta E_\downarrow$, corresponding to the energy cost/gain of hopping to right, left, top, and bottom nearest neighbors of a randomly chosen f-electron. The blue dots show the training data prediction with MSE = 0.041 and the orange dots show the testing data prediction with MSE = 0.046. 20

2.5 (a) The characteristic length $\ell(t)$ of the f-electron configuration $\{n_j^f\}$ as a function of time for ED-kMC and ML-kMC simulations. (b) The time dependence of the characteristic linear size L of the Ising domains associated with the super-clusters. The panels on the right show the comparison of f-electron correlation function $C_{ij} = \langle n_i^f n_j^f \rangle$ after (c) $t = 50$ and (d) $t = 500$ from the same initial condition. 22

2.6 (a) Average size $\langle s \rangle$ of checkerboard cluster as a function of time obtained from ML-kMC simulations on a 150×150 lattice. The inset shows the time dependence of the characteristic length scale $\ell(t) = \ell_0 + \Delta\ell$, where $\ell = \langle s \rangle^{1/2}$. The dashed line indicates a power-law growth with exponent $\alpha \approx 0.35$. Here time is measured in terms of 100 MC steps. (b) A close-up view of f-electron configuration at $t = 800$ after quench. 24

2.7 Top: snapshots of f-electron configuration obtained from kMC simulations of phase separation in a 150×150 square-lattice FK model. The blue and red dots indicate f-electrons on the A- and B-sublattice, respectively. Bottom: configurations of Ising variables σ_i that characterize the Z_2-symmetry-breaking domains associated with super-clusters. 25

2.8 (a) Scaling plot of the time-dependent structure factor $S(\mathbf{k}, t)$ obtained from the Fourier transform of the Z_2 order parameter. The dashed line shows the k^{-3} Porod's law in 2D. (b) Characteristic length $L(t)$ of the super-clusters as a function of time for three different lattice sizes. The dashed line indicates the linear growth $\Delta L(t) \sim t$. 26

2.9 (a) Density plot of effective potential $V(\mathbf{r}_i) = -k_B T \log \nu_i$ for f-electrons created by a checkerboard cluster at the center. The potential field $V(\mathbf{r})$ exhibits the same staggering patten as that of the checkerboard cluster at the center. Ions in the neighborhood thus tend to reside on the same sublattice, leading to the growth of the super-cluster. The depth of the staggering potential, given by the averaged potential difference between the two sublattices ΔV_{AB} versus (b) the characteristic length L of super-clusters and (c) the Ising order parameter ϕ. 30

2.10 Panels (a) and (b) show two examples of f-electron configuration obtained by the block method described in the text. Yellow regions represent the same-shaped checkerboard cluster, randomly rotated and placed in each of the 9 blocks; the blue background represents the empty sites. Panels (c) and (d) are the corresponding Ising configuration σ_i according to the definition discussed in the main text. Red (blue) regions correspond to $\sigma = +1$ (-1). Panels (e) and (f) show the effective potential experienced by a test f-electron created by the collection of checkerboard clusters (shown in black). 32

2.11 (a) Strength of the staggered potential $\Delta V_{AB} = |V_A - V_B|$ versus the characteristic length L of Ising domains from 50 different f-electron configurations generated by the block method. Points of different color share the same block magnetization M, as discussed in the text. . . 33

3.1 Schematic diagram of the CNN-based ML model for spin-torque prediction of itinerant electron magnets. The spin configuration $\{\mathbf{S}_i\}$ on a lattice is first flattened into three arrays, corresponding to the three components of spins, which are input to a series of ResNet blocks. Details of the ResNet are presented in Fig. 3.2. The output of the ResNet blocks is processed through additional convolutional layers, resulting in three arrays that, once reshaped, correspond to the torques $\{\mathbf{T}_i\}$ driving the spin dynamics. 47

3.2 Diagram of a ResNet block. The input to a ResNet block goes through two pathways: the skip connection, where no operation is performed if input and output have the same number of channels, or a 1×1 convolution otherwise; and the main connection, where two 5×5 convolution-ReLU activation blocks are stacked. The outputs of both pathways are then added together. 49

3.3 Predicted spin torque components (T_x, T_y, T_z) versus ground truth components from the testing set. The red-dotted diagonal lines indicate perfect prediction. The top row shows the prediction results based on spin configurations obtained from LLG simulations of a perturbed SkL. Results from LLG simulations with random initial states are shown in the bottom row. 52

3.4 (a) Comparison of predicted field vector magnitude against the ground truth. The red line indicates perfect agreement between predictions and ground truth, while the outer red dotted line represents a deviation of 10^{-2} from the ground truth magnitude, and the inner one represents a deviation of 10^{-3}. The color indicates the logarithmic density. (b) Angular difference between the ground truth field vector and the predicted field vector. 55

3.5 The architecture for integrating a NN with LLLG dynamics simulation of the s-d model using the BP scheme. Under the adiabatic approximation, for a given local spin during an LLG step, we train a fully-connected NN, denoted as $y_\theta(\mathbf{x})$, to predict the local energy E_i based on the configuration $\{\mathcal{C}_i\}$ of the surrounding spins, resulting in $E_i = y_\theta(\mathbf{x}|\{\mathcal{C}_i\})$. By combining the BP scheme with Eq.(3.3) and Eq.(3.6), the local exchange field $\{\mathbf{H}_i\}$ is computed through automatic differentiation of the total energy $E = \sum_i E_i$. The input $\mathbf{x}$ for the NN includes descriptors such as the bond $\{b_{jk}\}$ and chirality $\{\chi_{jmn}\}$, which are selected to maintain the symmetry properties of the system. 57

3.6 Examples of different neighborhood spin configurations used to compute the symmetry-invariant descriptors. (a) The six bond variables $b_j = \mathbf{S}_i \cdot \mathbf{S}_j$, where $\{\mathbf{S}_j || \mathbf{r}_i - \mathbf{r}_j| = a\}$, form the basis of a six-dimensional reducible representation of the D_6 point group. (b) Similarly, the six bond variables $b_j = \mathbf{S}_i \cdot \mathbf{S}_j$, where $\{\mathbf{S}_j || \mathbf{r}_i - \mathbf{r}_j| = \sqrt{3}a\}$, form the basis of a reducible six-dimensional representation of D_6. (c) The twelve bond variables $b_j = \mathbf{S}_i \cdot \mathbf{S}_j$, where $\{\mathbf{S}_j || \mathbf{r}_i - \mathbf{r}_j| = \sqrt{7}a\}$, form the basis of a twelve-dimensional reducible representation. (d) The six scalar chirality variables $\chi_j = \mathbf{S}_i \cdot \mathbf{S}_m \times \mathbf{S}_n$, where $\{\mathbf{S}_{m,n} || \mathbf{r}_i - \mathbf{r}_{m,n}| = a\}$, form the basis of a six-dimensional reducible representation. 58

3.7 (a) Comparison between the exact local exchange force field $\mathbf{H}^{\perp}_{i,\text{exact}}$ and the NN prediction $\mathbf{H}^{\perp}_{i,\text{ML}}$ perpendicular to the same site spin $\mathbf{S}_i$ with the training datasets (blue) and the test datasets (orange). (b) Display examples of the intermediate configuration for the skyrmion phase from the ML-LLG simulations. (c) Demonstrate the structure factor of the final state of skyrmion phase in the first BZ. 65

3.8 Comparison of spin structure factors obtained by averaging 30 independent LLG simulations based on KPM (left) and the CNN model (right). The same set of random initial conditions on a 48×48 triangular lattice was used in both simulations. The red dashed lines indicate the first BZ of the momentum space. 68

3.9 Snapshot of the spin configuration at the end of the LLG simulation with random initial conditions on a 48×48 triangular lattice. 69

3.10 CNN-based LLG simulation on a 96×96 lattice demonstrating the restoration of a perturbed SkL. The CNN model was originally trained on a 48×48 lattice. The initial spin configuration is given by the SkL ansatz (3.21) with additional site-dependent random phases and amplitudes of S_z. 71

3.11 Evolution of the structure factor over time for a Skyrmion initial condition in a simulation extending far beyond the training duration. The dashed black line indicates the duration of the training simulation. Our ML-based LLG simulation maintains a stable structure factor for more than five times the duration of the training simulation. 73

3.12 Distribution of the equivariance error $err_{eq} := f_{\mathrm{CNN}}(\{R\mathbf{S}_i\}) - Rf_{\mathrm{CNN}}(\{\mathbf{S}_i\})$, where R is an arbitrary rotation. The overall prediction error ("predicted torque" minus "ground truth torque") on the test dataset is superimposed for comparison. The equivariance error is significantly smaller than the overall prediction error, indicating that the model effectively preserves the underlying symmetry of the physical system. 76

3.13 ERFs of the ML model. Since both the input and output tensors at each lattice site have three components, there are a total of nine ERFs corresponding to the nine partial derivatives of the three output torque components with respect to the three input spin components. The sum of the absolute values of these derivatives is presented in this figure, with darker pixels indicating smaller derivative values. The red line roughly traces the nonzero value regions of the ERF. 77

3.14 Comparison of the structure factor $\mathcal{S}(q)$ along the high-symmetry path $\Gamma(0,0) \to K(\frac{4\pi}{3},0) \to M(2\pi,0)$ obtained from KPM-LLG and ML-LLG simulations at different time steps from $t = 50$ to 4000. The blue lines represent the average value of $\mathcal{S}(q)$ from 50 individual KPM-LLG simulations on a 48×48 lattice, with the blue regions indicating the standard deviation. The red dots with error bars depict the average $\mathcal{S}(q)$ and its standard deviation from ML-LLG simulations. 80

3.15 Top panels: Density plots of the spin configuration $\{\mathbf{S}_i^z\}$ on a 150×150 lattice at different time steps from $t = 50$ to 4000. Bottom panels: Vector arrows representing the in-plane spin configuration $\{\mathbf{S}_i^{xy}\}$ at the top-right corner of the respective upper panels, with $\{\mathbf{S}_i^z\}$ represented by the arrow colors. 81

3.16 Left four panels (a): Structure factor density plots of the ML-LLG simulations on a 150×150 lattice in the first BZ from $t = 50$ to 4000. (b) Skyrmion structure factor peak order $\mathcal{M}$ as a function of time from $t = 10$ to 10^4. (c) Characteristic length of the ML-LLG simulations from $t = 10$ to 10^4. 82

4.1 Schematic diagram of ML force-field model for metallic spin glass. The input of the ML model is the magnetic configuration $\mathcal{C}_i$ centered at the i-th spin $\mathbf{S}_i$, while the output is the local energy ϵ_i associated with the center spin. The atomic and spin configuration within the neighborhood $\mathcal{C}_i$ is processed using a magnetic version of the ACSF method. The resultant feature variables $\{G_k\}$ are input to a feed-forward fully connected neural network with a single output that is the local energy ϵ_i. The total energy is obtained by applying the ML energy model to all spins in the system. The local effective field is computed using automatic differentiation. 93

4.2 Electron density of states of the disordered tight-binding model Eq. (4.17) in the absence of electron-spin coupling $J = 0$. The hopping constant $t_0 = 1$, which serves as the unit of the energy. The number of atoms is $N = 100$ in a cubical box of linear size $L = 5\ell$ with periodic boundary conditions. The electron filling fraction is set at $f = 0.5$. The zero of the energy, indicated by the dashed line, corresponds to the Fermi level. 103

4.3 Benchmark of ML prediction for the perpendicular local field. Panel (a) shows the predicted components of $\mathbf{H}_\perp$ versus the ground-truth values for the trained NN model based on a magnetic descriptor with all three symmetry functions G_2^m, G_3^m, and G_3^χ. The corresponding histogram of prediction error $\delta = H_\perp^{\mathrm{ML}} - H_\perp^{\mathrm{ED}}$ is shown in panel (b). The results for ML model with a descriptor that *excludes* the chirality feature variables G_3^χ are shown in panels (c) and (d). 107

4.4 Dynamical benchmark of the ML force-field model for the random s-d system. The spin-spin correlation functions $C(r_{ij}) = \langle \mathbf{S}_i \cdot \mathbf{S}_j \rangle$ at various time-steps after a thermal quench are obtained from LLG simulations of $N = 100$ spins using the ML force-field model and the ED method. The radius is measured in units of ℓ, and the shaded area indicates the forbidden region where $r < r_{\min}$. The simulation time is measured in units of $(\gamma t_0)^{-1}$, where γ is the gyromagnetic ratio and t_0 is the energy scale of electron hopping. 111

4.5 The correlation length ξ as a function of the the logarithm of time $\log(t)$ in a log-log plot; the inset shows the $\xi(t)$ in the linear scale. ML-LLG simulations are performed on a random s-d system of $N = 500$ atoms. The length ξ is computed from the resultant ensemble-averaged correlation functions after a thermal quench at $t = 0$. The simulation time is measured in units of $(\gamma t_0)^{-1}$, where γ is the gyromagnetic ratio and t_0 is the energy scale of electron hopping. The red line corresponds to a power law $\xi(t) \sim (\log t)^{1/\psi}$ with an exponent $\psi \approx 1.02$. 112

5.1 Machine learning model for band gap prediction for a given atomic displacement field $\mathbf{u}_i$ within the a box of linear size ℓ. The ML model is composed of two central components: the descriptor and the neural network. The input of the ML model is a cubic block of displacement vectors $\mathbf{u}_i$. The descriptor corresponds to a representation or feature variables $\boldsymbol{G} = (G_1, G_2, G_3, \cdots)$ of the displacement vectors that is invariant under symmetry operations of the point group of the lattice, which in the case of Si is the T_d group. The complex dependence of the band gap on the nuclei configurations is encoded in the NN which takes the symmetry-invariant feature variables $\boldsymbol{G}$ as well as the temperature T as the input, and the predicted gap energy at the output. 125

5.2 Group of atoms sharing same distance to the center of the box form an invariant block of the representation matrix of the symmetry group. Panels (a), (b), and (c) show such invariant groups of 4, 6, and 12 atoms, respectively, in a diamond lattice. 129

5.3 Ground state (a) electron and (b) phonon dispersion of Si crystal along high-symmetry directions of the lattice. 137

5.4 Histogram showing distribution of phonon-induced band gap correction at (a) $T = 0$K and (b) $T = 200$ K using 100 data points for Si-supercell of size $6 \times 6 \times 6$. 139
5.5 Scatter plot showing phonon-induced band gap correction (in meV units) predicted by ML model versus DFT calculation for $6 \times 6 \times 6$ Si-supercell at four different temperature values. 143
5.6 Histogram of the prediction error defined as $\delta = \Delta E_g^{\mathrm{ML}} - \Delta E_g^{\mathrm{DFT}}$ for temperatures at (a) $T = 0$ K, (b) 100 K, (c) 200 K, and (d) 300 K. . 144
5.7 Temperature dependence of the phonon-induced band-gap correction ΔE_g of Si. In total, 528 DFT calculations on 432-atom Si supercell were done. ML prediction was done on 1000 configurations at each temperature. Experimental data is extracted from Ref. [1]. The experimental data is shifted by DFT calculated zero-point band gap correction energy. 146

6.1 (a) Schematic diagram of the Gating-induced CDW to metal transition in the adiabatic Holstein model. (b) Energy diagram of the two electrodes and the center Holstein model. The CDW state is a band insulator with an energy gap $E_g = 2gQ_0$. Both the Fermi level ϵ_F at the bulk and the chemical potential μ_R of the right electrode are set at the middle of the energy gap, while the chemical potential of the left-electrode is lowered from the Fermi level by the gating voltage: $\mu_L = \epsilon_F - eV$. 155
6.2 NEGF-BD simulations of a driven Holstein model on a 40×30 square lattice. The driven voltage is applied along the longitudinal x direction. Panels (a)–(c) show the snapshots of the system local electron density $n_i = \langle \hat{c}_i^\dagger \hat{c}_i \rangle$ at various simulation times. The following parameters are used: electron-phonon coupling constant $g = 1.5t_{\mathrm{nn}}$, the effective spring constant $k_1 = 0.67g$, nearest-neighbor elastic coupling $\kappa = 0.12g$, damping constant $\lambda = 0.2g$, $\Gamma_{\mathrm{lead}} = 1.0$, $\Gamma_{\mathrm{bath}} = 0.001$, $k_B T = 0.1$, and the bias voltage $eV = 2.5t_{\mathrm{nn}}$. 158
6.3 Time dependence of (a) the transmission current I and (b) the CDW order parameter Φ at zero temperature. A constant voltage $eV = 2.5t_{\mathrm{nn}}$ is applied to the Holstein system during the interval $0 \leq t \leq t_{\mathrm{off}} = 400$. 159

6.4 (a) average position ξ of the CDW-metal domain-wall as a function of time for different gating voltages at zero temperature. (b) CDW-metal domain-wall velocity v as a function of the gate voltage eV. The velocity is computed from the average slope of the $\xi(t)$ curve in panel (a). 162
6.5 Average position ξ of the CDW-metal interface as a function of time for different temperatures with constant driven voltage $eV = 2.5t_{\rm nn}$. The inset shows the CDW-metal domain-wall velocity as a function of temperature. 164

7.1 Classical fields with different symmetry properties. An example of type-I case is the local JT distortion $\mathbf{Q}_j$ as shown in panel (a). The lattice rotation/reflection is accompanied by a simultaneous transformation of the JT phonons. Panel (b) shows the type-II case exemplified by local spins $\mathbf{S}_j$. The global rotation symmetry of spins is independent of the discrete point-group symmetry of the lattice. 172
7.2 Examples showing the atomic pair (jk) and triplet (jkl), which are related by the lattice rotation and reflection symmetries, in the neighborhood of the center site on a square lattice. 188
7.3 Schematic diagram of the vibronic modes for the MnO6 octahedron: (a) the symmetry-preserving breathing mode, (b) and (c) the symmetry-breaking JT modes. In terms of oxygen displacements, the coordinates of these normal modes are: $Q^{A_1} = (-X_1+X_2-X_3+X_4-X_5+X_6)/\sqrt{6}$, $Q^x = (-X_1+X_2+X_3-X_4)/2$, and $Q^z = (-X_1+X_2-X_3+X_4+2X_5-2X_6)/\sqrt{12}$, where X_i denotes the x coordinates of the i-th oxygen, and so on. 196
7.4 Schematic diagram of (a) C-type orbital order, and (b) the concomitant antiferro-distortive JT order. Also shown in panel (b) is the vector representation of the orbital/JT order. The arrows represent either the doublet vector $\boldsymbol{Q}^E$ or the expectation value of the orbital pseudospin $\langle\hat{\boldsymbol{\tau}}\rangle$. These two vectors are related to each other via Eq. (7.56) in the ground state. 198

7.5 Schematic diagram of the ML force-field model for the cooperative JT models. A lattice descriptor transforms the neighborhood distortion configuration $\mathcal{C}_i$ into effective coordinates $\{G_m\}$ which are then fed into a fully connected NN. The output node of the NN corresponds to the local site-energy ϵ_i. The combination of the descriptor and the NN provides an approximation for the universal function $\varepsilon(\cdot)$. The corresponding total potential energy E is obtained from the summation of the local energies. Automatic differentiation is employed to compute the derivatives $\partial E_{\rm ML}/\partial \boldsymbol{\mathcal{Q}}_i$ for the effective forces acting on the breathing and JT modes. 202

7.6 Benchmark of the ML force-field models for JT model with a filling fraction $f = 0.49$. Panels (a)–(c) on the left show the ML predicted forces $F^{\rm ML}$ versus the ground truth $F^{\rm ED}$ obtained from exact diagonalization method for the three vibronic modes of the octahedron. These forces are normalized by the electron-phonon coupling constant, which is set to $\lambda = 1.25$. The corresponding histograms of the prediction errors are shown in panels (d)–(f). Similar results are also obtained for the ML models for the half-filling case. 213

7.7 Comparison of time-dependent correlation functions $C^{xx}(r,t)$ defined in Eq. (7.68) obtained from Langevin simulations with the ML force-field model and the ED method. The correlation functions were obtained from 40 independent thermal quench simulations on a 30×30 lattice with a filling fraction $f = 0.49$. 214

7.8 Snapshots of local breathing mode Q^{A_1} and orbital Ising order parameter ϕ_i at various time steps after a thermal quench of a half-filled $f = 0.5$ JT model. An initially random configuration is suddenly quenched to a temperature $T = 0.001$ at time $t = 0$ ($n_{\rm step} = 0$). The ML-Langevin dynamics was used to simulate the relaxation of the system toward equilibrium. The red and blue regions correspond to orbital domains with order parameters $\phi = +1$ and -1, respectively. A time step $\Delta t = 0.05\tau_0$ is used in the simulations. 216

7.9 Snapshots of local breathing mode Q^{A_1} and orbital Ising order parameter ϕ_i at various time steps after a thermal quench for a JT model with $f = 0.49$ electron filling. 218

7.10 Localized checkerboard modulation of breathing mode Q^{A_1} and electron density. An example of such local structures is displayed in terms of (a) the amplitude of breathing mode as well as the $\boldsymbol{Q}^E$ vector and (b) the orbital Ising order parameter. 220
7.11 Structure factor $S(\mathbf{k}, t)$ of the JT Q^x distortion mode at different time steps after suddenly cooled to a temperature $T = 0.001$, starting from a random state. The system size is 100×100 with filling fraction $f = 0.49$, and averaged over 40 independent configurations. 223
7.12 The characteristic length $L(t)$ of orbital domains as a function of time, averaging over 40 independent ML-Langevin simulations. Panel (a) shows $L(t)$ curves for the case of half-filling and a filling fraction $f = 0.49$. Panel (b) shows the $L(t)$ curves at two different temperatures for the case of $f = 0.49$ electron filling. The two dashed lines in (a) and the two solid lines in (b) are fit using the Allen-Cahn growth law with a shifted time t_0 and an initial length L_0. The fitted expansion coefficients are $a = 8.1 \times 10^{-5}$ and $a = 9.1 \times 10^{-5}$ for temperature $T = 0.001$ and $T = 0.005$, respectively. 225

B.1 How normalizing flow works 244
B.2 Overall structure of the normalizing flow [2] from the base space (u) to the parameter space (θ), with optimal hyperparameter choices indicated. Red connections are invertible. The residual network is made up of n_{blocks} residual blocks, each with two fully-connected hidden layers of n_{hidden} units. Prior to each linear transformation [3], we inserted batch normalization layers to speed training [4] and Exponential Linear Units for nonlinearity [5]. Each block is also conditioned on the strain data s. 246
B.3 Marginalized one-dimensional posterior distributions over a subset of parameters, comparing the normalizing flow (red) and bilby dynesty (green). Contours represent 50% and 90% credible regions. Neural network posteriors are constructed from 10^7 samples with 931 effective importance samplings. 248

List of Tables

2.1 D_4 point group character table 16
2.2 The ML model structure 19

5.1 The architecture of neural network. GeLU activation function and drop-out rate of 0.3 was used in all layers except output layer. 140

Chapter 1

Introduction

Functional electron materials, such as high-temperature superconductors and materials exhibiting charge density waves, are renowned for their complex, emergent properties arising from strong electron correlations [6–25]. These materials exhibit a wide range of mesoscale patterns, including stripes, checkerboards, and spin textures, which emerge due to intricate electronic interactions. Understanding the origin, stability, and evolution of these patterns is essential, as they often underpin the unique electronic, magnetic, and transport properties of the materials, rendering them promising candidates for a variety of advanced technological applications, ranging from data storage to quantum information processing and energy-efficient devices.

The remarkable behaviors exhibited by functional electron materials are not solely determined by their atomic lattice structure but are profoundly influenced by electronic interactions that give rise to collective phenomena. For instance, the emergence of high-temperature superconductivity in certain cuprates and iron-based materials is

believed to be driven by unconventional electron pairing mechanisms, which remain the subject of intense theoretical investigation. Similarly, charge density waves and spin textures [26–34] often result from the competition between different interactions, such as Coulomb repulsion, lattice coupling, and magnetic exchange, driving the system into complex, spatially non-uniform states. These emergent properties are of significant interest not only because they represent novel phases of matter but also because they hold transformative potential for future technologies.

The theoretical modeling of such complex systems has traditionally relied on various numerical techniques, including Monte Carlo simulations, exact diagonalization, density functional theory (DFT), and etc [35–42]. These approaches, while highly successful for small systems, become computationally prohibitive as the system size increases, limiting our ability to explore emergent behavior in larger systems or to capture phenomena that manifest at realistic scales. Moreover, the non-equilibrium dynamics of these systems, which are often central to their functional properties, require an efficient and accurate means of modeling both local electronic interactions and large-scale structural evolution. The challenge lies in capturing the intricate interplay between local electron correlations and global structural order, which is especially important when attempting to understand the behavior of these materials under experimental conditions, such as temperature quenches or applied electromagnetic fields.

In recent years, machine learning (ML) has emerged as a powerful tool to address

these computational challenges [43–56]. Machine learning techniques, particularly deep learning, offer sophisticated methods for recognizing and exploiting complex patterns in data, which is crucial for modeling systems characterized by a high degree of freedom and strong correlations. By training neural networks on smaller, exactly solvable systems, we can construct predictive models that generalize effectively to larger, more complex configurations. This strategy, often termed multi-scale modeling, enables the simulation of the dynamics of functional electron materials at a fraction of the computational cost of conventional approaches. Neural networks (NN) are particularly well-suited for this task because they can effectively learn the underlying relationships between the microscopic states of a system and its emergent macroscopic properties, thereby obviating the need for computationally intensive calculations at every step.

A key advantage of machine learning in this context is its capacity to handle the high dimensionality inherent in many-body systems. Traditional methods are often hindered by the so-called "curse of dimensionality," in which the computational cost grows exponentially with the number of particles or degrees of freedom in the system. Machine learning models, however, can be trained to identify efficient representations of these high-dimensional systems, thereby reducing the effective dimensionality and making it possible to study larger and more complex configurations. Furthermore, ML models can be trained to predict energy landscapes, forces, and other relevant properties based on local configurations, which is particularly advantageous for inves-

tigating the emergent properties of these materials, especially due to the scalability of the system.

This dissertation presents a detailed investigation into the application of machine learning for large-scale modeling of functional electron materials. We begin with the Falicov-Kimball model [57], a canonical model that captures essential aspects of phase separation and charge ordering in correlated electron systems. This model serves as an ideal testbed because it incorporates many of the fundamental features of electron correlation while remaining tractable for detailed analysis. We demonstrate that neural networks can efficiently model the complex energy landscape of the Falicov-Kimball model, enabling large-scale kinetic Monte Carlo simulations that reveal the dynamics of phase separation and the development of charge density wave order. These simulations provide new insights into the mechanisms of phase separation, including the role of thermal fluctuations, domain growth, and the stabilization of long-range order.

We then extend our methodology to investigate the dynamics of itinerant electron magnets and disordered spin systems [19–25]. Itinerant electron magnets, in which magnetic moments arise from the collective behavior of delocalized electrons, pose distinct challenges compared to localized spin systems. The interplay between electron mobility and magnetic exchange interactions leads to rich phase diagrams and complex, non-trivial dynamical phenomena. By utilizing machine learning models, we simulate the time evolution of these systems, capturing intricate behaviors such

as spin-wave propagation, magnetic domain formation, and the effect of external perturbations on magnetic ordering. For disordered spin systems, where the presence of randomness significantly influences the physical properties, machine learning provides a powerful approach to explore the impact of disorder on both magnetic and electronic characteristics, offering fresh perspectives on spin-glass states and other disordered phases.

Our results show that machine learning can not only replicate the outcomes of traditional exact methods with remarkable accuracy but also extend the simulations to substantially larger system sizes [58, 59]. This capability is particularly crucial for the study of emergent phenomena, which often become prominent only at macroscopic length scales. For example, the formation of large-scale magnetic domains, the nucleation and growth of charge order, or the cooperative behavior in spin glasses necessitate simulations that include many thousands of particles—far beyond the scope of conventional computational approaches. By leveraging machine learning, we bridge the gap between small-scale theoretical models and the large-scale behavior observed in experiments, thereby enhancing our ability to understand and predict the physical properties of these materials.

In summary, the work presented here establishes a robust framework for employing machine learning to deepen our understanding of complex, correlated electron materials. This approach paves the way for studying emergent phenomena in systems that were previously intractable using conventional computational techniques,

thereby opening new opportunities for the design, discovery, and application of functional electron materials. By integrating the strengths of machine learning with traditional physics-based modeling, we can develop a more comprehensive understanding of these fascinating systems, ultimately contributing to advancements in energy storage, quantum computing, spintronics, and other cutting-edge fields of technology.

Chapter 2

Machine Learning in the Falicov-Kimball Model

2.1 Introduction

Complex mesoscopic textures are common in strongly correlated electron materials [6–18]. Notable examples include stripe and checkerboard patterns in high-T_c superconductors, as well as nanoscale mixtures of metallic and insulating domains in manganites. These mesoscopic textures are not only fundamentally important in correlated electron physics but are also central to the emergence of novel functionalities in these materials. The nanoscale patterns in correlated electron materials often result from phase-separation instabilities driven by electron correlation effects. Indeed, a common feature of lightly doped Mott insulators is their strong tendency toward phase separation, where doped holes are expelled from locally insulating antiferromagnetic domains [60–67]. Despite considerable efforts to understand the mechanisms of phase separation and the properties of mixed-phase states in strongly correlated materials, the non-equilibrium pattern-formation dynamics in these systems remains poorly

understood.

Intermediate states with complex structures are also observed in discontinuous phase transitions in many classical systems [68,69]. The kinetics of first-order transitions is a well-established subject with a long history [70–72]. These studies focus on the evolution of a system from an unstable or metastable state toward its equilibrium phase, often characterized by the emergence of complex spatial-temporal patterns. Several numerical techniques, ranging from kinetic Monte Carlo (KMC) and molecular dynamics (MD) simulations to phase-field modeling, have been developed for large-scale simulations of phase-separation phenomena [35–38]. Of particular interest in such studies are the dynamical universality class and the associated universal growth law [73,74]. The phase-ordering process is often modeled using partial differential equations for the order-parameter fields that describe the symmetry-breaking structure. However, most work in this field is based on empirical energy models, which often fail to capture the complex and long-ranged interactions of order-parameter fields, particularly in correlated electron systems.

Comprehensive modeling of correlation-driven phase separation requires simultaneous consideration of microscopic electronic processes and mesoscopic pattern formation dynamics. For instance, one could obtain the driving forces on the order-parameter fields by integrating out the electrons on-the-fly, meaning that the electronic structure problem must be solved at every time step of the macroscopic dynamics simulation. However, repeatedly solving the electronic problem using methods

ranging from exact diagonalization (ED) to more sophisticated many-body techniques can be prohibitively expensive for large-scale simulations. These computational challenges are partly why progress in understanding phase-ordering dynamics in correlated electron systems has been limited.

In this chapter, we take an important step toward the goal of multi-scale dynamical modeling of strongly correlated electron systems by utilizing machine learning (ML) techniques to develop an efficient yet accurate energy model. This approach has allowed us to achieve the first-ever large-scale simulation of phase separation phenomena in the Falicov-Kimball (FK) model [57], one of the canonical correlated electron systems. Originally introduced as a limiting case of the Hubbard model [75], the FK model was later independently proposed to describe semiconductor-metal transitions in rare-earth and transition-metal compounds [57]. The FK model describes interactions between conducting c-electrons and localized f-electrons through an on-site repulsive interaction. Its relative simplicity allows for numerically exact solutions, which serve as important benchmarks for sophisticated many-body methods [76]. The FK model itself has rich phase diagrams and is one of the most well-studied correlated electron systems that exhibit complex pattern formation and phase separation [77–85]. In particular, the FK model provides a proof of principle that stripe and checkerboard orders—prominent features in the phenomenology of high-T_c superconductivity—can arise purely from electronic correlation effects [78–81], such as the Kivelson-Emery scenario of phase separation.

2.2 Kinetic Monte Carlo for the Falicov-Kimball Model

In this work, we consider the spinless FK Hamiltonian on a square lattice [76, 81]:

$$\mathcal{H} = -t_{\text{nn}} \sum_{\langle ij \rangle} c_i^\dagger c_j + U \sum_i c_i^\dagger c_i n_i^f, \tag{2.1}$$

Here, $c_i^\dagger$ (c_i) is the creation (annihilation) operator for a c-electron at site i, $\langle ij \rangle$ denotes nearest-neighbor pairs on the lattice, n_i^f is the occupation number of the f-electron, t_{nn} is the nearest-neighbor hopping constant (serving as the energy unit), and $U > 0$ is the strength of the on-site repulsive interaction. Due to the quadratic nature of the c-electron Hamiltonian, the equilibrium phases of the FK model can, in principle, be solved exactly by combining the classical Monte Carlo (MC) method for f-electrons with ED for c-electrons [83–85]. Furthermore, within the framework of dynamical mean-field theory, the quantum impurity problem associated with the FK model can also be solved exactly [76, 82].

The equilibrium phases of the square-lattice FK model have been extensively studied over the years. At half-filling for both c and f electrons, the ground state exhibits a charge-density wave (CDW) order, with the f-electrons forming a checkerboard pattern [84, 85]. Away from half-filling, the model displays various stripes and incommensurate phases [80, 81]. With slight electron or hole doping, a phase-separated regime is stabilized [81–83], similar to the Hubbard model. Despite being a prominent model for electronic phase separation, the phase-ordering dynamics in FK systems has not been explored. Important questions, such as whether the system exhibits

dynamical scaling and the nature of the domain-growth law, remain open.

To address these issues, we develop a KMC method [37,86] to simulate the phase ordering process of the FK model under a temperature quench. While the c-electrons have well-defined dynamics in the FK Hamiltonian, the f-electrons are static discrete variables, with $n_i^f = 0$ or 1, similar to classical Ising spins. To provide dynamics for the f-electrons, we use a random-walk algorithm to model their diffusive motion. At each time step, an attempt is made to move a randomly chosen f-electron to one of its empty neighbors. Whether the move is accepted is determined by the standard Metropolis criterion [37]. We further assume that the equilibration of c-electrons is much faster compared to the random walks of f-electrons, analogous to the Born-Oppenheimer approximation in quantum MD [87]. Consequently, the motion of the heavier f-electrons depends on the free energy of the quasi-equilibrium c-electrons before and after the update. The acceptance probability for such a nearest-neighbor move is given by:

$$p_{i\to j} = \frac{1}{4}\min\left(1, e^{-\Delta E_{i\to j}/k_B T}\right), \tag{2.2}$$

where $\Delta E_{i\to j}$ represents the free-energy difference of c-electrons due to the hopping of an f-electron from site i to j. The probability that the f-electron stays in place is $p_{i\to i} = 1 - \sum_j p_{i\to j}$. It is worth noting that lattice gas systems combined with kMC simulations are widely used to describe phase separation in conventional alloys [88,89]. However, most works are based on empirical energy models, often formulated as

effective classical Ising Hamiltonians. In this work, we demonstrate that deep neural networks (NN), trained using exact solutions, can provide an accurate and efficient energy model for the f-electrons in the FK system.

The c-electron free energy can be calculated using either ED or more efficient techniques, such as the kernel polynomial method [90]. However, the quantum KMC simulation described above is very time-consuming for large systems, as one must solve the electron tight-binding problem four times at each time step to update just one f-electron.

2.3 Machine Learning Architecture

To address the computational bottleneck, we apply ML methods that have been successfully used to enhance the efficiency of quantum MD simulations [91–97]. Similar techniques have also been applied to enable large-scale quantum spin dynamics in double-exchange systems [98, 99]. The central idea of our approach, summarized in Fig. 2.1, is rooted in the principle of locality [100, 101], often referred to as the "nearsightedness" of electronic matter by W. Kohn. In our context, the locality principle implies that the energy change $\Delta E_{i\to j}$ depends only on the f-electron configuration within the vicinity of the local update.

Specifically, the energy change $\Delta E_{i\to j}$ of a local update is assumed to depend on the neighborhood configuration through a universal function:

$$\Delta E_{i\to j} = \varepsilon(\hat{\mathbf{n}}_{ij}, \mathcal{C}_i), \tag{2.3}$$

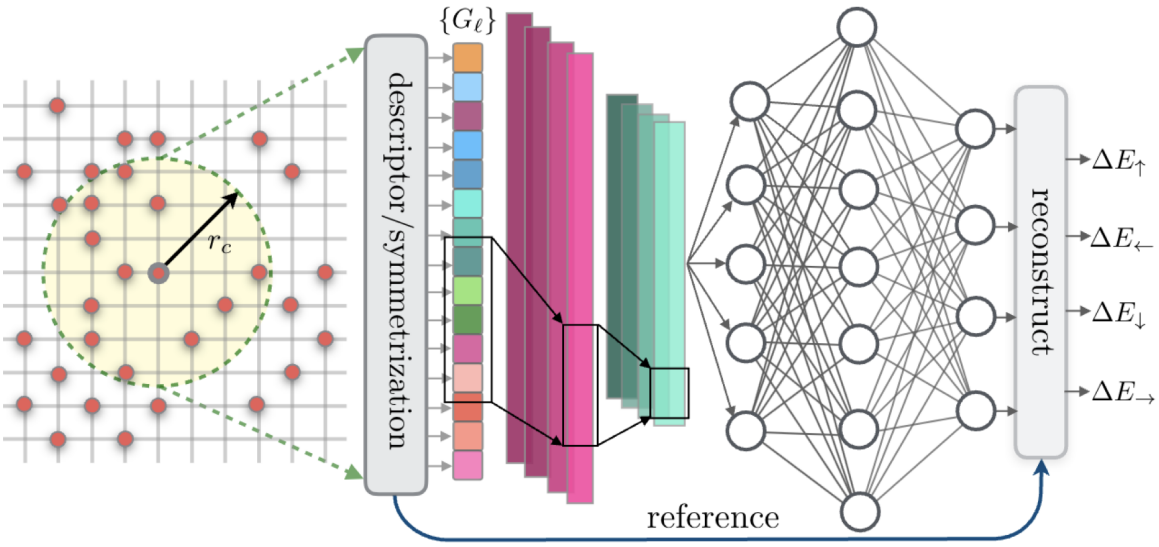

Figure 2.1: Schematic diagram of NN energy model for kMC dynamics simulation of the FK system. A descriptor is used to construct effective coordinates $\{G_\ell\}$ from neighborhood f-electron configuration $\{n_j^f\}$ up to a cutoff $r_c = 10$. These feature variables $\{G_\ell\}$ are input to the NN which predicts the energy differences ΔE at the output.

where $\hat{\mathbf{n}}_{ij} = \pm\hat{\mathbf{x}}, \pm\hat{\mathbf{y}}$ denotes the orientation of the $i \to j$ bond, and $\mathcal{C}_i = \left\{ n_j^f \,\middle|\, |\mathbf{r}_j - \mathbf{r}_i| \leq R_{\text{cutoff}} \right\}$ describes the neighborhood f-electron configuration up to a cutoff radius R_{cutoff}. The complex dependence of the energy function $\varepsilon(\cdot)$ on the local environment is encoded in a NN that is trained using exact solutions from small systems.

2.3.1 Lattice Descriptor

Next, the effective energy function $\varepsilon(\cdot)$ is expected to also preserve the site-symmetry of the lattice. A descriptor is used to construct effective coordinates $\{G_\ell\}$ from the neighborhood f-electron configuration $\{n_j^f\}$ up to a cutoff radius $R_{\text{cutoff}} = 10$. These feature variables $\{G_\ell\}$ are then fed into the NN, which predicts the energy differences $\Delta E_{i\to j}$ caused by the update at the output. As the energy prediction should preserve

the discrete lattice symmetry of the square-lattice FK model, the generalized coordinates $\{G_\ell\}$ are expected to be invariant under symmetry operations of the discrete point group, for example 90° rotation about the z-axis, or mirror reflection about xz, yz planes. To this end, we first use the fact that the f-electron occupation numbers $\{n_j^f\}$ within the cutoff form the basis of a high-dimensional representation of the point group, which in the case of the square lattice is the D_4 group. By decomposing this high-dimensional representation into the irreducible representations (IRs) of the symmetry group, invariants can be systematically constructed from the basis functions of the various IRs [102].

Since the distance is preserved by symmetry operations of the point group, the representation matrices of the neighborhood occupation are automatically block-diagonalized according to the distances to the central point. This significantly simplifies the task of finding the IRs. For example, Fig. 2.2 shows the case of the neighboring sites $\{w_j\}$ forming an 8-dimensional block, which can be readily decomposed as $8 = A_1 \oplus A_2 \oplus B_1 \oplus B_2 \oplus 2E$. The corresponding basis of the IRs are the linear

combination of $\{w_j\}$ as follows:

$$\begin{aligned}
f^{A_1} &= w_1 + w_2 + w_3 + w_4 + w_5 + w_6 + w_7 + w_8 \\
f^{A_2} &= w_1 - w_2 + w_3 - w_4 + w_5 - w_6 + w_7 - w_8 \\
f^{B_1} &= w_1 + w_2 - w_3 - w_4 + w_5 + w_6 - w_7 - w_8 \\
f^{B_2} &= w_1 - w_2 - w_3 + w_4 + w_5 - w_6 - w_7 + w_8 \\
\boldsymbol{f}^{E_1} &= (w_2 - w_3 - w_6 + w_7,\ w_1 + w_4 - w_5 - w_8) \\
\boldsymbol{f}^{E_2} &= (w_1 + w_2 - w_5 - w_6,\ w_3 + w_4 - w_7 - w_8)
\end{aligned} \tag{2.4}$$

By repeating the same procedures for each block, we arrange the resultant IR basis functions into a vector $\boldsymbol{f}_r^\Gamma = (f_{r,1}^\Gamma, f_{r,2}^\Gamma, \cdots, f_{r,D_\Gamma}^\Gamma)$ where Γ labels the IR, r enumerates the multiple occurrences of Γ in the decomposition of the f-electron configuration, and D_Γ is the dimension of the IR. The IR basis follows the D_4 point group's character table in Table 2.1. Given these basis functions, one can immediately obtain a set of invariants called the power spectrum $\{p_r^\Gamma\}$, which are the amplitude of each individual IR function, i.e. $p_r^\Gamma = \left|\boldsymbol{f}_r^\Gamma\right|^2$. However, feature variables based only on the power spectrum are incomplete in the sense that the relative phases between different IRs are ignored. For example, the relative "angle" between two IRs of the same type: $\cos\theta = (\boldsymbol{f}_{r_1}^\Gamma \cdot \boldsymbol{f}_{r_2}^\Gamma)/|\boldsymbol{f}_{r_1}^\Gamma||\boldsymbol{f}_{r_2}^\Gamma|$ is also an invariant of the symmetry group. Without such phase information, the NN model might suffer from additional error due to spurious symmetry, namely two IRs can freely rotate independently of each other.

A systematic approach to include all relevant invariants, including both amplitudes

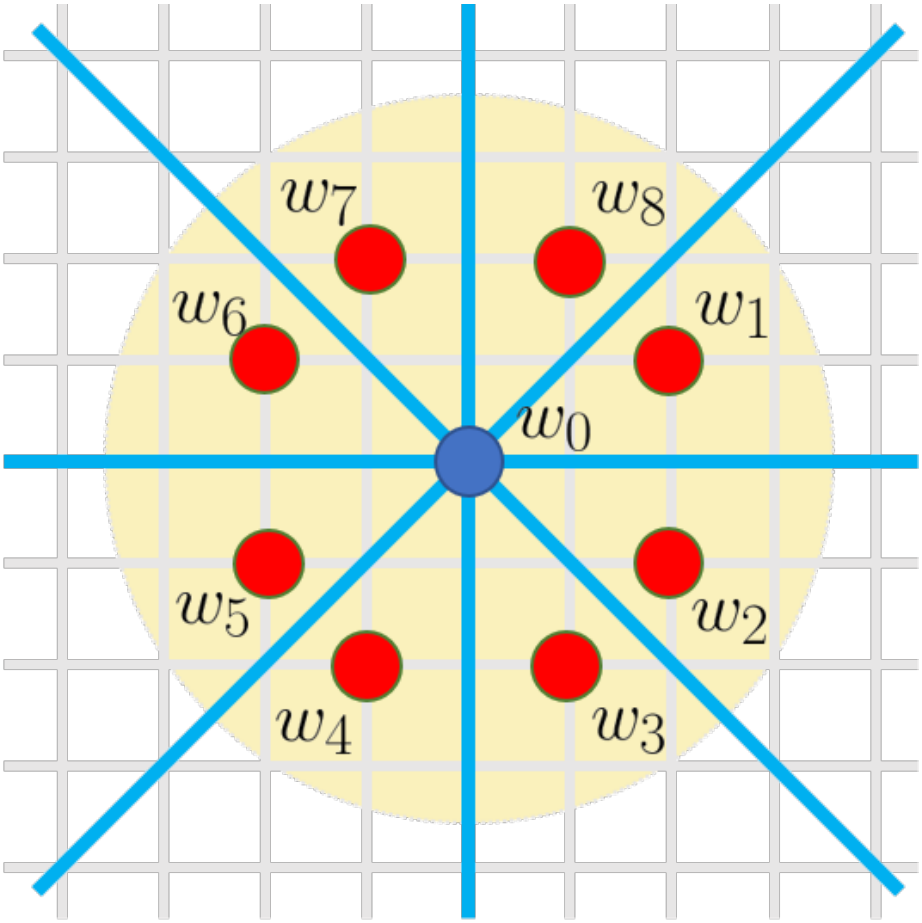

Figure 2.2: The example of a block of the f-electron configuration $\{w_j\}$ that is used to generate block-diagonal representations of the D_4 point group of the square lattice. These eight sites are the 4^{th} nearest neighbor to the ion $w_0 = 1$, where w_j is either occupied (1) or unoccupied (0).

	E	$2C_4(z)$	$C_2(z)$	$2C_2'$	$2C_2''$
A_1	1	1	1	1	1
A_2	1	1	1	-1	-1
B_1	1	-1	1	1	-1
B_2	1	-1	1	-1	1
E	2	0	-2	0	0

Table 2.1: D_4 point group character table

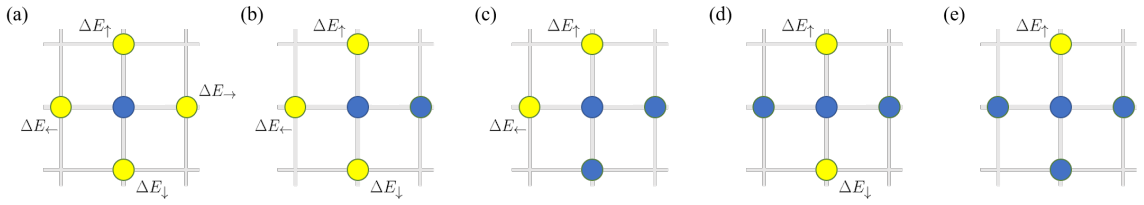

Figure 2.3: Five different nearest-neighbor configurations for updating an f-electron at the central site $n_0^f = 1$. The blue circles denote nearest neighbor sites that are occupied by an f-electron, while the yellow circles denote empty sites.

and relative phases, is the bispectrum method [103, 104]. In this work, we develop a descriptor similar to the one used in Ref. [105], that is modified from the bispectrum method. We introduce a set of reference basis functions $\boldsymbol{f}_{\text{ref}}^{\Gamma}$ for each IR of the point group. These reference basis are computed by averaging large blocks of f-electrons, such that they are less sensitive to small changes in the neighborhood f-electron configurations. We then define the relative "phase" of a IR as the projection of its basis functions onto the reference basis: $\eta_r^{\Gamma} \equiv \boldsymbol{f}_r^{\Gamma} \cdot \boldsymbol{f}_{\text{ref}}^{\Gamma}/|\boldsymbol{f}_r^{\Gamma}||\boldsymbol{f}_{\text{ref}}^{\Gamma}|$. The effective coordinates are then obtained from the power spectrum and the relative phases: $\{G_\ell\} = \{p_r^{\Gamma}\ ,\ \eta_r^{\Gamma}\}$.

It is worth noting that the energy cost of hopping to a neighbor that is already occupied by another f-electron is infinite. Numerically, it is very difficult to include this special situation among finite predictions. However, for such infinite energy cost, the corresponding probability is zero, which means we can preclude such transition probability in our consideration for such situations. Consequently, it is easier in practice to implement several different NN models, one for each of the five different nearest-neighbor f-electron configurations shown in Fig. 2.3. Importantly, depending

on the nearest-neighbor configuration, different point-group symmetry has to be used for computing the generalized coordinates $\{G_\ell\}$. Our discussion of the descriptor above is mainly focused on the case of four empty neighbors shown in Fig. 2.3(a). For configurations shown in Fig. 2.3(b), (c), and (e), the symmetry group is reduced to C_2, while the situation shown in Fig. 2.3(d) is described by point group D_2.

2.3.2 Convolutional Neural Network

Inorder to predict the energy cost, we have built an 8-layer NN model on PyTorch [106] for the large-scale ML-kMC simulations of the FK model. Details of the NN model are summarized in Table 2.2. We have included convolutional layers in our NN to extract characteristic patterns of the input f-electron configurations [107]. Notably, we have verified that NN models with additional convolutional layers outperform those based only on fully connected layers.

The NN model was trained by datasets generated from kMC simulations on a 30×30 square lattice using ED method. The following parameters are used: nearest-neighbor hopping $t_{\rm nn} = 1$, on-site repulsive energy $U = 2\,t_{\rm nn}$, temperature $T = 0.05$, f-electron density $\rho_f = 0.187$, and c-electron density $\rho_c = 0.55$. As discussed in the next section, these parameters were chosen in order to realize a low-temperature phase consisting of a mixture of checkerboard patterns and uniform domains that are free of f-electrons.

The training datasets come from both random configurations and ED-kMC simulations for a thermal quench to $T = 0.05$. On average, the NN models for the five

Layer	Network
Input Layer	[316,1][a]
Convolutional Layer 1	conv(5,1,0,16)[b] maxpool(3,3)[c] act[d] = ReLU
Convolutional Layer 2	conv(5,1,1,32) maxpool(3,3) act = ReLU flatten[e] $\rightarrow$ [1088]
Hidden Layer 3	FC(1088,256)[f] act = ReLU
Hidden Layer 4	FC(256,128) act = ReLU
Hidden Layer 5	FC(128,64) act = ReLU
Hidden Layer 6	FC(64,32) act = ReLU
Output Layer	FC(32,1)

[a]The shape of the input data [one-dimensional dataset length, No. channels]
[b]One-dimensional convolutional filter with arguments (filter size, stride, padding, No. filters).
[c]Max-pooling layer with arguments (pool size, stride length).
[d]The activation function.
[e]Flatten the multi-channel output of the previous layer to the one-dimensional neurons.
[f]Fully connected layer with arguments (input size, output size).

Table 2.2: The ML model structure

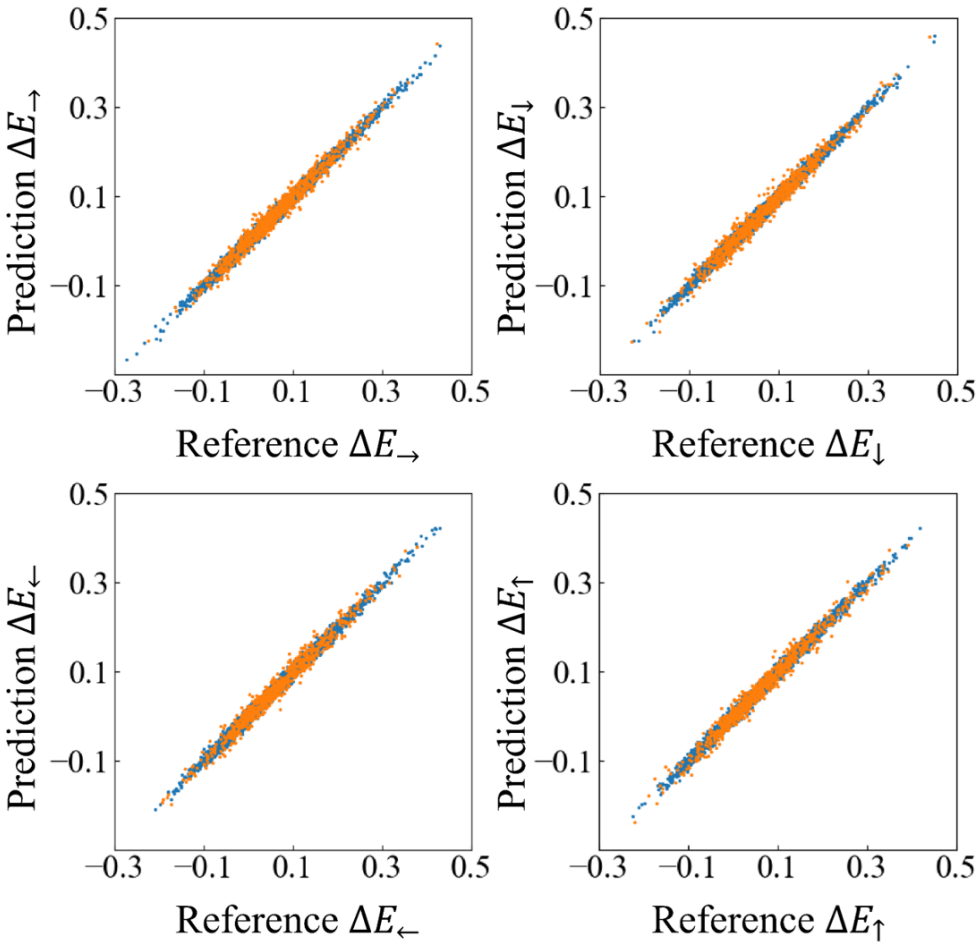

Figure 2.4: Correlation between the NN predictions and the exact solutions for energy differences $\Delta E_{\rightarrow}$, $\Delta E_{\leftarrow}$, $\Delta E_{\uparrow}$, and $\Delta E_{\downarrow}$, corresponding to the energy cost/gain of hopping to right, left, top, and bottom nearest neighbors of a randomly chosen f-electron. The blue dots show the training data prediction with MSE $= 0.041$ and the orange dots show the testing data prediction with MSE $= 0.046$.

different nearest-neighbor configurations shown in Fig. 2.3 are each trained by at least 10^6 datasets. The Adam optimizer [108] with a learning rate of 0.001 is used to minimize the loss function, which is defined as the mean square error (MSE) of the energy prediction. We have obtained nice overall agreement between the predicted values and exact calculations, as demonstrated by the validation results shown in Fig. 2.4.

2.3.3 Results

Next, we integrate our NN with the kMC to further benchmark the performance of the ML methods compared with the ED kMC results. We first calculate 100 independent simulations using the ED-kMC with parameters above. Another 100 independent simulations with the same parameters are then carried out by ML-kMC using the trained NN model. Fig. 2.5(a) and (b) show the characteristic lengths of the checkerboard clusters and their super-clusters (or effective Ising domains defined in the main text) obtained from ED and ML-kMC simulations. These two characteristic lengths, ℓ and L, are obtained from the time-dependent structure factor of f-electron configurations and Ising configurations, respectively. In both cases, reasonable agreements between the two methods are obtained. Comparisons of the f-electron correlation functions $C_{ij} = \langle n_i^f n_j^f \rangle$ obtained from ED and ML-kMC simulations are shown in Fig. 2.5(c) and (d) for 50 and 500 time-steps, respectively, after a thermal quench. The reasonable agreements of the correlation functions indicate the ML-model can successfully capture the relaxation dynamics of the FK model.

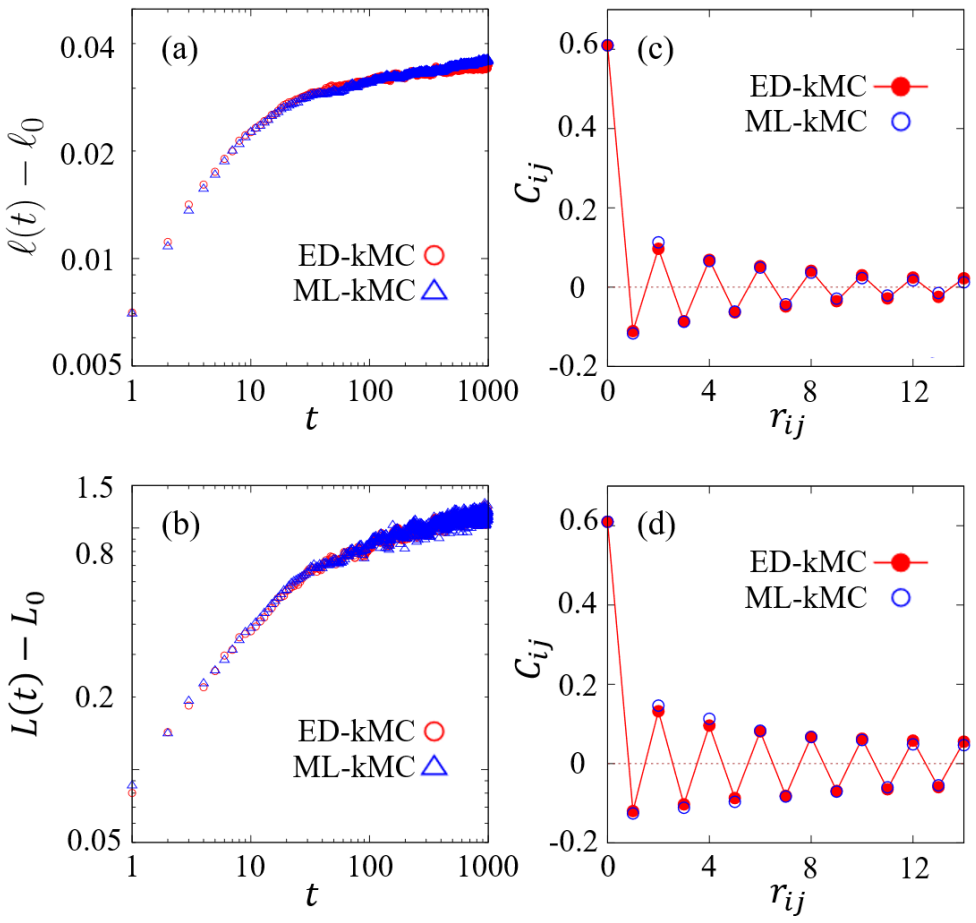

Figure 2.5: (a) The characteristic length $\ell(t)$ of the f-electron configuration $\{n_j^f\}$ as a function of time for ED-kMC and ML-kMC simulations. (b) The time dependence of the characteristic linear size L of the Ising domains associated with the super-clusters. The panels on the right show the comparison of f-electron correlation function $C_{ij} = \langle n_i^f n_j^f \rangle$ after (c) $t = 50$ and (d) $t = 500$ from the same initial condition.

2.4 Phase Separation in the Falicov-Kimball Model

We employ the benchmarked NN energy model to perform large-scale ML-KMC simulations of the FK model with up to 10^5 lattice sites. Our objective is to investigate the growth dynamics of checkerboard clusters after a temperature quench. Specifically, we consider a slightly doped c-electron system with a filling fraction of $\rho_c = 0.55$ and a low f-electron density of $\rho_f = 0.187$. The repulsive interaction is set at $U = 2\,t_{\rm nn}$. In the low-temperature phase, the system exhibits phase separation characterized by a mixture of checkerboard CDW ordering of f-electrons and a metallic region devoid of f-electrons [82]. Some stripe-like orders are also observed. In our simulations, the system is initially prepared in a random configuration, and the temperature is abruptly reduced to $T = 0.05\,t_{\rm nn}$ at time $t = 0$. A snapshot of the f-electron configuration at a later time after the quench, as shown in Fig. 2.6(b), clearly displays multiple checkerboard clusters along with some diagonal stripe patterns of the f-electrons.

Figure 2.6(a) shows the growth of the average checkerboard cluster size $\langle s \rangle$ over time, indicating the aggregation of f-electrons to form CDW order during relaxation. Since the number of f-electrons is conserved, the growth of checkerboard domains resembles the phase separation of a conserved order parameter, which is expected to follow a $t^{1/3}$ power-law growth as predicted by the Lifshitz-Slyozov-Wagner (LSW) theory [109, 110] or the model-B dynamical model [73, 111]. As shown in the inset of Fig. 2.6(a), the characteristic length scale of checkerboard clusters follows a power-law $\Delta\ell \sim t^{\alpha}$ in the early stages of phase separation, with an exponent $\alpha \approx 0.35$ that

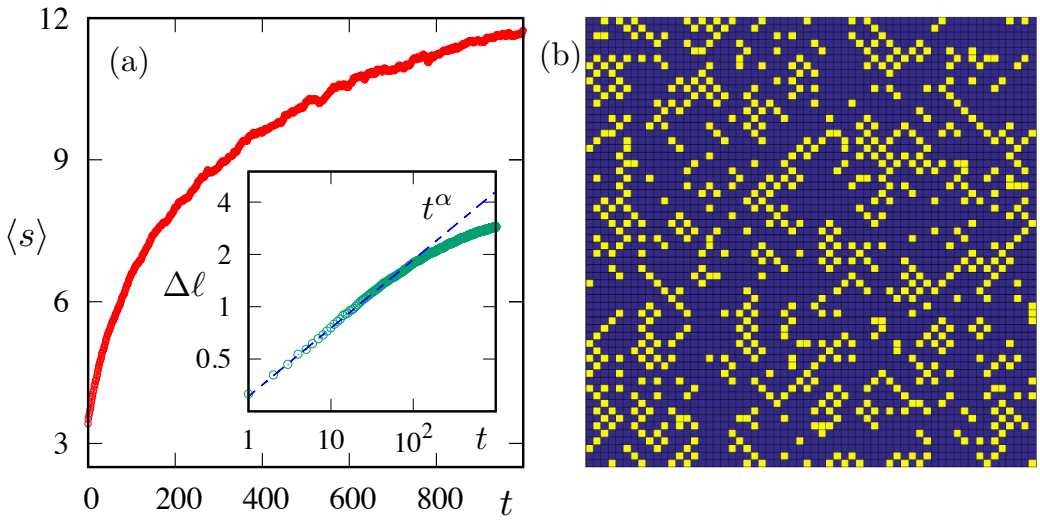

Figure 2.6: (a) Average size $\langle s \rangle$ of checkerboard cluster as a function of time obtained from ML-kMC simulations on a 150×150 lattice. The inset shows the time dependence of the characteristic length scale $\ell(t) = \ell_0 + \Delta\ell$, where $\ell = \langle s \rangle^{1/2}$. The dashed line indicates a power-law growth with exponent $\alpha \approx 0.35$. Here time is measured in terms of 100 MC steps. (b) A close-up view of f-electron configuration at $t = 800$ after quench.

is slightly higher than the LSW prediction. This discrepancy may arise from the fact that LSW scaling is generally expected to hold during the coarsening of very large domains at late times [112].

However, this power-law growth is only sustained for a short duration, and the domain growth significantly slows down at later stages. This stagnation in domain growth cannot be attributed to finite-size effects, as the average cluster size remains much smaller than the system size at late times. Instead, the freezing of the checkerboard clusters is associated with a hidden sublattice symmetry breaking in the phase separation process, which we discuss in detail below. To illustrate this effect, we use different colors to label the f-electrons on the two sublattices of the square lattice. As shown in the top row of Fig. 2.7, when the checkerboard clusters are small, f-

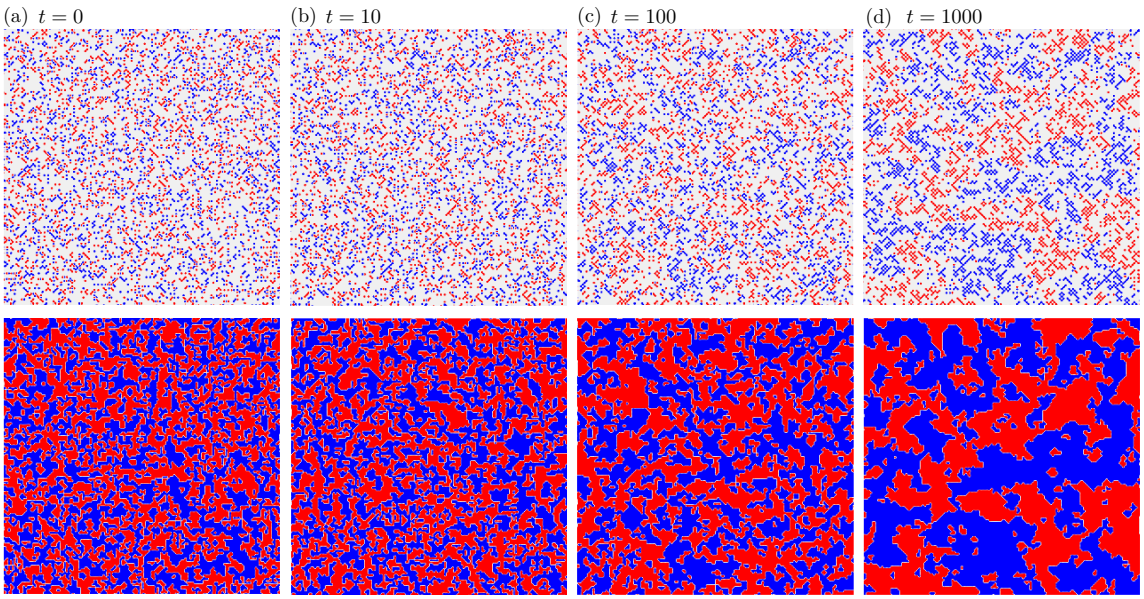

Figure 2.7: Top: snapshots of f-electron configuration obtained from kMC simulations of phase separation in a 150×150 square-lattice FK model. The blue and red dots indicate f-electrons on the A- and B-sublattice, respectively. Bottom: configurations of Ising variables σ_i that characterize the Z_2-symmetry-breaking domains associated with super-clusters.

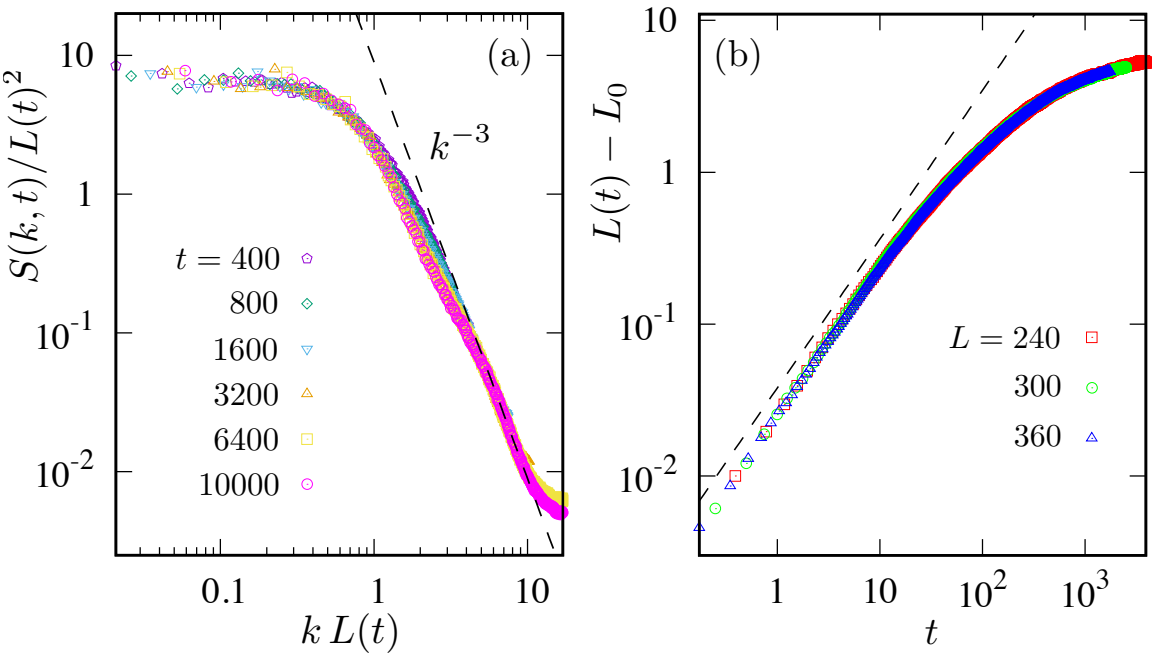

Figure 2.8: (a) Scaling plot of the time-dependent structure factor $S(\mathbf{k}, t)$ obtained from the Fourier transform of the Z_2 order parameter. The dashed line shows the k^{-3} Porod's law in 2D. (b) Characteristic length $L(t)$ of the super-clusters as a function of time for three different lattice sizes. The dashed line indicates the linear growth $\Delta L(t) \sim t$.

electrons on the same sublattice tend to cluster together, forming super-clusters of checkerboards. Importantly, the formation of such super-structures also breaks the Z_2 sublattice symmetry.

It is important to note that two distinct symmetry-breaking processes are occurring simultaneously during the temperature quench: the conventional CDW-metal phase separation at smaller length scales and the coarsening of super-clusters or effective Ising domains at larger scales. The formation of super-clusters is not an inevitable outcome of phase separation involving CDW order. Since the f-electrons in a CDW cluster can occupy either the A or B sublattice, each cluster can be charac-

terized by a Z_2 variable, referred to as its polarity. In a typical phase-separated state, one would expect a mixture of CDW clusters with opposite polarities. However, our simulation results suggest that the energy of these mixed-phase states can be further reduced through the alignment of CDW clusters with the same polarity, resulting in the formation of super-clusters. As discussed below, this alignment is driven by a nonlocal interaction mediated by the c-electrons.

To describe the larger-scale Z_2 symmetry breaking associated with super-cluster coarsening, we introduce an Ising variable σ_i at each lattice site, such that $\sigma_i = +1$ (-1) if the f-electron closest to site i belongs to the A (B) sublattice. Based on this definition, the bottom row of Fig. 2.7 shows the Ising configurations $\{\sigma_i\}$ corresponding to the respective f-electron distributions $\{n_i^f\}$ shown on the top. In the language of Ising spins, the clustering of checkerboards into super-clusters corresponds to the growth of Ising ferromagnetic domains. The order parameter ϕ describing this Z_2 symmetry breaking is then given by the magnetization density of Ising spins, i.e., $\phi = \langle \sigma_i \rangle$. It is noteworthy that ϕ is not conserved in the KMC dynamics of f-electrons. Phenomenologically, such a non-conserved field is governed by the time-dependent Ginzburg-Landau (TDGL) equation, or model-A dynamics [73]. The resulting domain coarsening is characterized by the Allen-Cahn power law [70, 72]

$$L \sim t^{1/2}. \tag{2.5}$$

However, as we will show, the coarsening of super-clusters in our system does not

follow the expected power law due to an unusual self-confinement effect among the f-electrons.

To characterize the growth of Ising domains associated with the super-clusters, we compute the structure factor of the Ising spins, $S(\mathbf{k},t) = \left|\frac{1}{N}\sum_i \sigma_i(t)\exp(i\mathbf{k}\cdot\mathbf{r}_i)\right|^2$. Ferromagnetic ordering implies that $S(\mathbf{k},t)$ exhibits a growing peak at $\mathbf{k}=0$. The inverse of the width of this peak can be used as a measure of the characteristic length scale of the super-clusters: $L^{-1}(t) \sim \Delta k = \sum_{\mathbf{k}} S(\mathbf{k},t)|\mathbf{k}|/\sum_{\mathbf{k}} S(\mathbf{k},t)$. Using this characteristic length as a scale factor, Fig. 2.8(d) shows the scaled time-dependent structure factor versus the dimensionless momentum $|\mathbf{k}|L(t)$. As seen in the figure, the data points at different times collapse roughly onto the same curve, indicating that the coarsening of the Ising domains exhibits dynamical scaling,

$$S(\mathbf{k},t) = L^2(t)\mathcal{G}[|\mathbf{k}|L(t)], \tag{2.6}$$

where $\mathcal{G}(x)$ is a universal scaling function. The $1/k^3$ power-law tail at large momenta, consistent with the 2D Porod's law [71], results from the sharp interfaces between the two different types of Ising domains, or super-clusters with opposite polarities.

The characteristic length $L(t)$ extracted from the structure factor is shown in Fig. 2.8(b) as a function of time for three different lattice sizes. Interestingly, except for a short initial period (up to $t \sim 10$), the growth of this length scale does not follow the expected power law, especially at late times. Moreover, even the initial seemingly power-law growth is inconsistent with the $\alpha = 1/2$ Allen-Cahn law. Instead, L seems

to increase linearly with time in this initial regime. To understand this anomalous behavior, we note that the TDGL equation, or the Allen-Cahn theory, describes an interface-controlled domain growth in which the interfacial velocity is proportional to the curvature of the domain interface [113]. In our case, since the Z_2 order parameter is defined by whether the aggregating f-electrons are on the A- or B-sublattice, the resulting domain growth need not rely on the expansion of an existing boundary. Instead, a super-cluster can quickly increase its size as f-electrons in its neighborhood move from one sublattice to another via only a nearest-neighbor hopping. Due to the collective movement of f-electrons, the growth of super-clusters exhibits an avalanche-like behavior similar to the Barkhausen effect in magnetic domain growth. A faster linear growth of the super-clusters thus arises from such avalanche dynamics at the early stage. As discussed below, this collective behavior is induced by an effective non-local interaction between f-electrons.

Although the repulsion U between the two types of electrons is local in the FK model, the heavier f-electrons experience an effective long-range interaction mediated by the itinerant c-electrons. In particular, due to this non-local interaction, the presence of a checkerboard cluster creates a staggered potential in its neighborhood that takes alternating values on neighboring sites of the bipartite lattice. This effective potential is illustrated in Fig. 2.9(a), where a test f-electron is placed in the neighborhood of a checkerboard cluster at the center. Exact MC simulation was used to obtain the frequency ν_i with which the test particle stays at site i, from which the potential

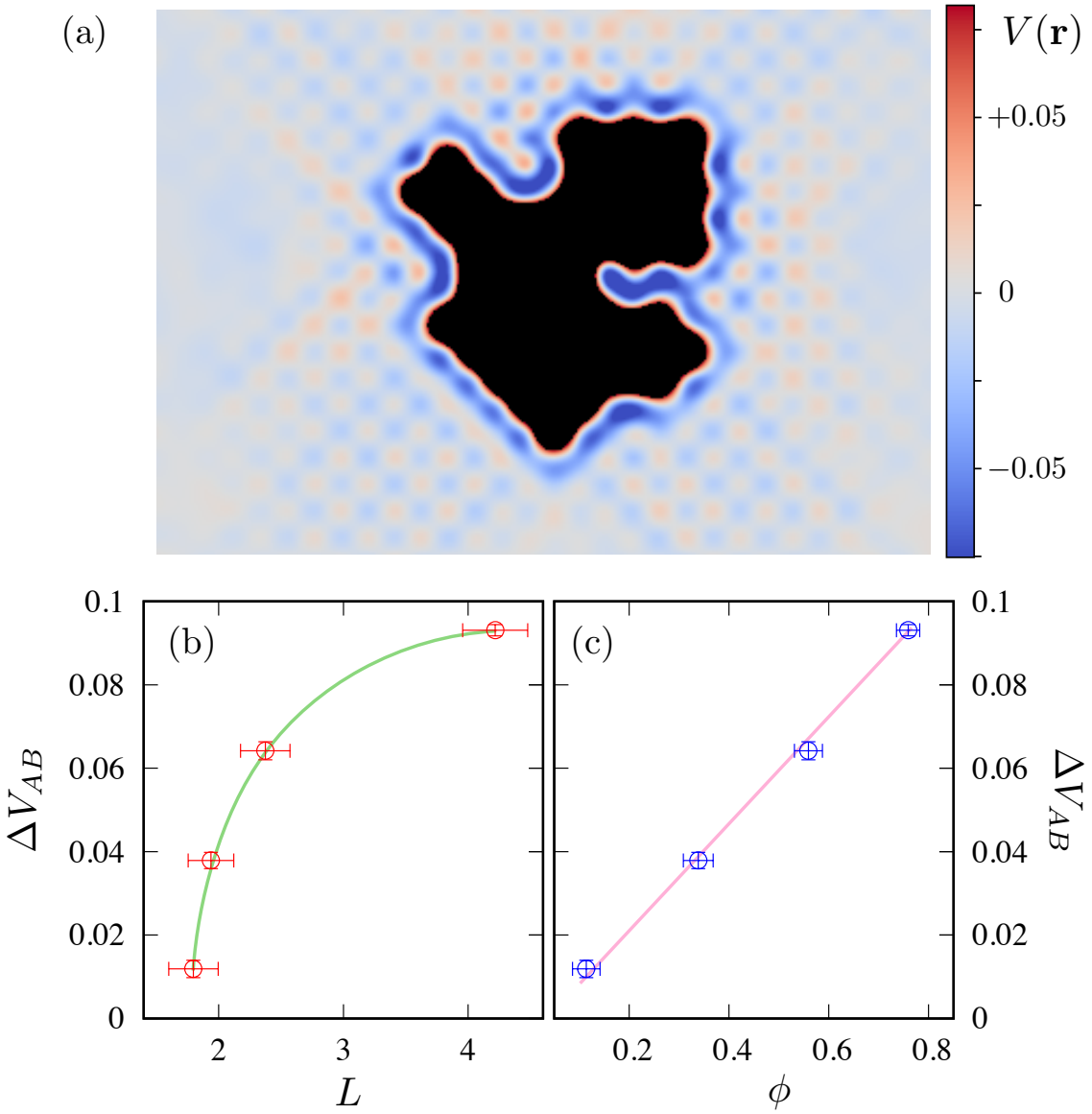

Figure 2.9: (a) Density plot of effective potential $V(\mathbf{r}_i) = -k_B T \log \nu_i$ for f-electrons created by a checkerboard cluster at the center. The potential field $V(\mathbf{r})$ exhibits the same staggering patten as that of the checkerboard cluster at the center. Ions in the neighborhood thus tend to reside on the same sublattice, leading to the growth of the super-cluster. The depth of the staggering potential, given by the averaged potential difference between the two sublattices ΔV_{AB} versus (b) the characteristic length L of super-clusters and (c) the Ising order parameter ϕ.

is computed as $V_i = -k_B T \log \nu_i$. As shown in Fig. 2.9(a), the effective potential field $V(\mathbf{r}_i)$ exhibits the same staggered pattern whose polarity is determined by that of the center checkerboard cluster. Consequently, f-electrons in the neighborhood of this checkerboard cluster tend to be trapped in the same sublattice. Importantly, this staggered potential also causes existing checkerboards in the neighborhood to change their polarity, thereby leading to the formation of a super-cluster and subsequent growth that is not captured by the interface-controlled mechanism.

In the late stage of the phase separation, a much slower logarithmic-like growth sets in for super-clusters, as shown in Fig. 2.8(b). Interestingly, the staggered potential discussed above is also responsible for the stalled growth of the Ising domains and, in fact, the smaller checkerboard clusters as well. To understand this unusual freezing behavior, we note that while the strength of the staggered potential increases with the size of the CDW cluster from which it originates, the energy barrier ΔV_{AB} between the two sublattices is also enhanced as more and more checkerboards merge to form larger super-clusters. To demonstrate this effect, we consider different geometrical arrangements of a finite number of checkerboard clusters on a 30×30 lattice, giving rise to different shapes and sizes of super-clusters or Ising domains. MC simulation with ED is then used to compute the resultant effective potential $V(\mathbf{r}_i)$ for the f-electrons. A 30×30 lattice with periodic boundary conditions is evenly divided into 9 blocks; see Fig. 2.10. A small checkerboard cluster of arbitrary shape is placed in each block. Different super-cluster configurations can be realized by rotating and/or

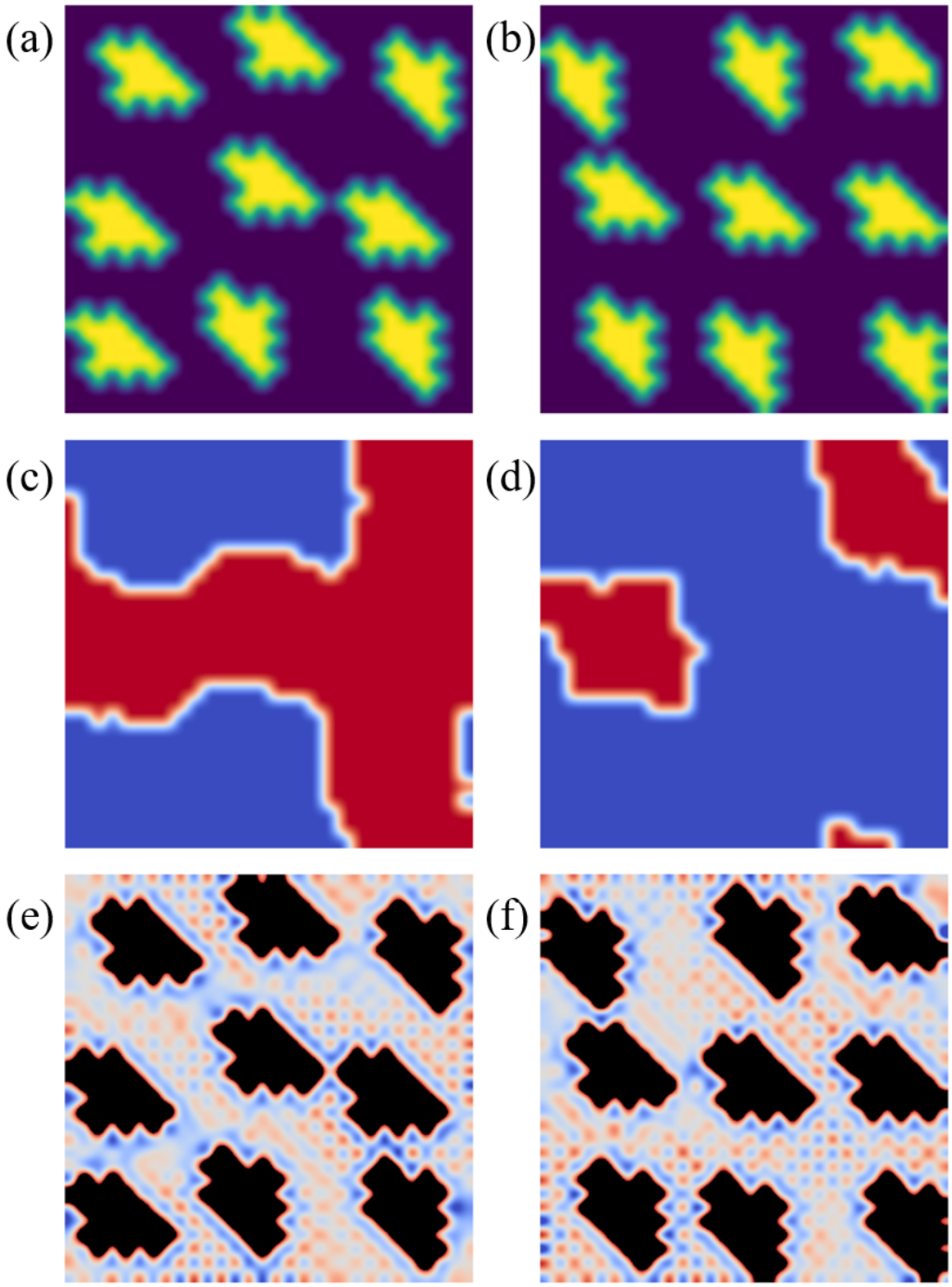

Figure 2.10: Panels (a) and (b) show two examples of f-electron configuration obtained by the block method described in the text. Yellow regions represent the same-shaped checkerboard cluster, randomly rotated and placed in each of the 9 blocks; the blue background represents the empty sites. Panels (c) and (d) are the corresponding Ising configuration σ_i according to the definition discussed in the main text. Red (blue) regions correspond to $\sigma = +1$ (-1). Panels (e) and (f) show the effective potential experienced by a test f-electron created by the collection of checkerboard clusters (shown in black).

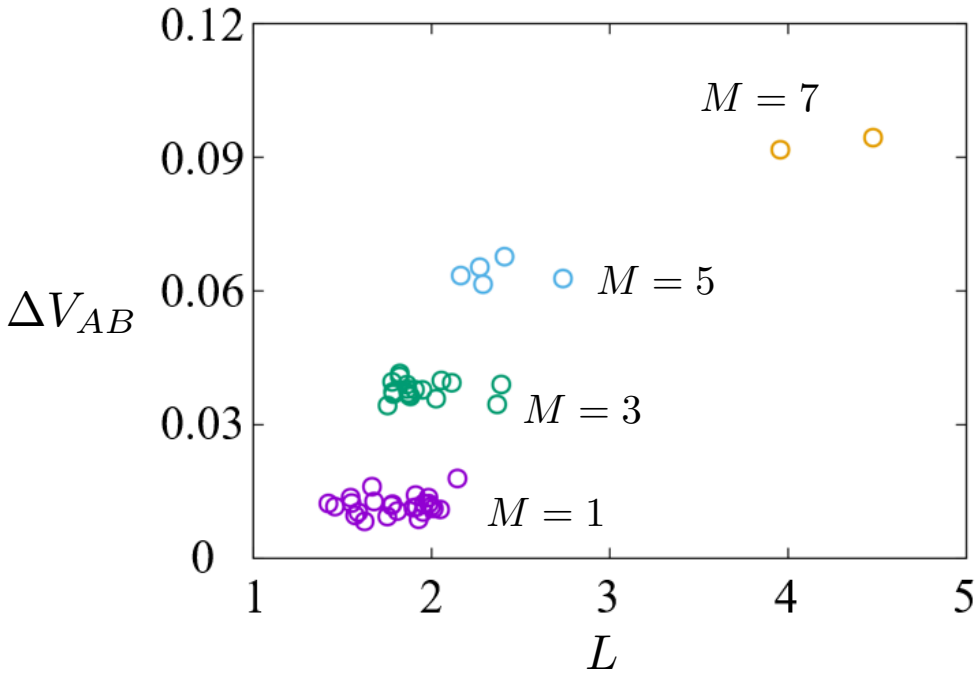

Figure 2.11: (a) Strength of the staggered potential $\Delta V_{AB} = |V_A - V_B|$ versus the characteristic length L of Ising domains from 50 different f-electron configurations generated by the block method. Points of different color share the same block magnetization M, as discussed in the text.

translating the (same-shaped, same-sized) checkerboard cluster within each 10×10 block.

For each f-electron configuration $\{n_i^f\}$ generated by the above procedure, we first compute the corresponding Ising configuration $\{\sigma_i\}$; see Figs. 2.10(c) and (d) for two examples. Next, we compute the resultant characteristic length L of the Ising domains, which represents the typical size of super-clusters, from the structure factor of the Ising variables. For the same f-electron configuration, we also perform MC simulations with ED to compute the effective potential $V_i = -k_B T \log \nu_i$ experienced by a test particle, where ν_i is the frequency with which the test f-electron stays at

site i. The average potential difference between the A/B sublattices is given by

$$\Delta V_{AB} = \left\langle \frac{1}{N/2} \left| \sum_{i \in \mathrm{A}} V_i - \sum_{i \in \mathrm{B}} V_i \right| \right\rangle \tag{2.7}$$

The relation between the staggered potential ΔV_{AB} and the corresponding Ising domain length L is shown in Fig. 2.11 for 50 different f-electron configurations randomly generated by the block method.

Interestingly, the data points seem to fall into four groups, as highlighted by different colors in Fig. 2.11. To understand this unusual result, we note that we can introduce an Ising variable σ_J for the J-th block ($J = 1, 2, \cdots, 9$), which is defined as the majority of Ising spins inside this block. According to the definition discussed above, the majority of Ising variables in a given block is $\sigma = +1$ (-1) if the f-electrons of the small checkerboard cluster inside the block reside on the A (B) sublattice. Different f-electron configurations generated using the block procedure can then be classified according to the block Ising spins $\{\sigma_J\}$. The total "magnetization" of the block Ising spins is given by $M = \sum_{J=1}^{9} \sigma_J$. It turns out that the four different groups in Fig. 2.11 correspond to block magnetizations $|M| = 1, 3, 5$, and 7. Our results thus indicate that the strength of the staggered potential ΔV_{AB} depends mainly on the total block magnetization M. The average potential difference between the A/B sublattices ΔV_{AB} is shown in Fig. 2.9(b) as a function of the numerically obtained characteristic length L of the Ising domains. The staggered potential ΔV_{AB} indeed increases with the size of the super-clusters. Moreover, we observe an intriguing linear

dependence of the potential barrier ΔV_{AB} on the effective Ising order parameter ϕ, as shown in Fig. 2.9(c).

Importantly, the fact that the energy barrier ΔV_{AB} increases with the size of the Ising domains also explains the freezing behaviors observed in our ML-kMC simulations. As the size L of super-clusters increases with time, the potential difference between the two sublattices becomes so strong that individual f-electrons are deeply trapped at one sublattice and cannot hop to the neighboring sites. For example, consider a checkerboard cluster on sublattice-A in Fig. 2.9(a) and a test particle sitting at a site that belongs to the lower-energy sublattice-A. Although the checkerboard at the center has a strong affinity to the new particle, as evidenced by the rather low potential energy at the edge of the cluster, the large energy barrier at B-sublattice prevents the f-electron from joining the cluster. The reduced mobility of the f-electron thus results in an arrested coarsening of both the super-clusters as well as the smaller-sized checkerboard clusters.

Importantly, the fact that the energy barrier ΔV_{AB} increases with the size of the Ising domains also explains the freezing behaviors observed in our ML-kMC simulations. As the size L of super-clusters increases with time, the potential difference between the two sublattices becomes so strong that individual f-electrons are deeply trapped at one sublattice and cannot hop to neighboring sites. For example, consider a checkerboard cluster on sublattice A in Fig. 2.9(a) and a test particle sitting at a site that belongs to the lower-energy sublattice A. Although the checkerboard at

the center has a strong affinity for the new particle, as evidenced by the relatively low potential energy at the edge of the cluster, the large energy barrier at the B-sublattice prevents the f-electron from joining the cluster. The reduced mobility of the f-electron thus results in arrested coarsening of both the super-clusters and the smaller checkerboard clusters.

2.5 Conclusion and Outlook

In summary, by employing modern ML techniques, we have successfully developed a NN energy model that enables the first-ever large-scale KMC simulation of the well-studied FK model. Our findings reveal a novel phase-ordering phenomenon, characterized by domain coarsening occurring simultaneously at two different scales: the growth of checkerboard clusters and the expansion of Ising domains associated with a hidden broken sublattice symmetry. The interplay between these two processes results in an anomalously slow phase separation. Several intriguing dynamical phenomena, such as the early-stage avalanche domain growth and the decelerated coarsening of super-clusters, require further investigation and are left for future work.

Unusual domain coarsening has been reported in classical systems, often linked to frustrated interactions or quenched disorder [114–118]. In this work, we describe a new freezing mechanism that arises from the interaction between itinerant c-electrons and classical f-electrons. Similar glassy dynamics could be a general feature of phase ordering in other correlated electron systems. A characteristic feature of correlated electron materials is the coexistence of fast electron quasiparticles and slow bosonic

or collective degrees of freedom. The nontrivial interplay between these two sets of variables may lead to novel dynamical phenomena unique to correlated electron systems. Given the complexity of such systems, we envision ML techniques as indispensable tools for multi-scale modeling of nonequilibrium dynamics driven by electron correlation effects.

Chapter 3

Machine Learning for Itinerant Electron Magnets on Lattice

3.1 Introduction

Itinerant frustrated magnets with electron-mediated spin-spin interactions frequently exhibit complex non-collinear or non-coplanar spin textures. Among these textures, particle-like objects such as magnetic vortices and skyrmions are of particular interest, not only for their fundamental role in magnetism but also for their significant technological potential in the emerging field of *spintronics* [19–25]. These nanometer-sized, localized spin textures are characterized by nontrivial topological invariants, making them stable objects with long lifetimes. In itinerant electron magnets, skyrmions can be manipulated—moved, created, or annihilated—by applying magnetic fields or electrical currents, owing to electron-spin interactions. Such versatility highlights their potential for practical applications. Furthermore, these complex spin textures can give rise to intriguing electronic and transport properties, such as the topological Hall effects and topological Nernst effects [25, 119–121]. These effects originate

from the nontrivial Berry phase acquired by electrons when traversing closed loops of non-coplanar spins [122].

Dynamical modeling of complex textures in itinerant spin systems, however, is a computationally challenging task. While magnetic moments in most metallic skyrmion materials can be well approximated as classical spin vectors, the local effective magnetic fields, analogous to forces in molecular dynamics, originate from exchange interactions with itinerant electrons and must be computed quantum mechanically. Dynamics simulations of such itinerant magnets thus require solving an electronic structure problem associated with the instantaneous spin configuration at every time step. Repeated quantum calculations would be prohibitively expensive for large-scale simulations. Consequently, empirical classical spin Hamiltonians, from which the local fields can be explicitly calculated, are often employed in large-scale dynamical simulations of skyrmion magnets [123, 124]. Yet, such classical spin models often fail to capture the intricate long-range spin-spin interactions mediated by electrons, limiting their accuracy in describing complex magnetic phenomena.

The computational complexity of the above quantum approaches to spin dynamics is akin to that of *ab initio* MD methods. Unlike classical MD, which relies on empirical force fields, quantum MD calculates atomic forces by integrating out electronic degrees of freedom on-the-fly as atomic trajectories are generated [87]. Various many-body methods, notably density functional theory (DFT), have been employed for force calculations in quantum MD. However, the high computational cost associated with

repeated electronic structure calculations severely restricts the feasible scale of atomic simulations. To address this challenge, ML methods have been leveraged to develop force-field models that accurately emulate the computationally expensive many-body calculations, enabling large-scale MD simulations with quantum-level accuracy.

Crucial to the remarkable scalability of ML-based force-field models is the divide-and-conquer approach proposed in the pioneering works of Behler and Parrinello [91], and Bartók *et al.* [92]. In this approach, the total energy of the system is decomposed into local contributions, $E = \sum_i \epsilon_i$, where ϵ_i represents the atomic energy and depends solely on the local environment of the i-th atom [91, 92]. The atomic forces are subsequently derived from the predicted energy: $\mathbf{F}_i = -\partial E/\partial \mathbf{r}_i$, where $\mathbf{r}_i$ is the atomic position vector. The complex dependence of the atomic energy ϵ_i on its local neighborhood is efficiently approximated by an ML model, trained to ensure that both the predicted individual forces $\mathbf{F}_i$ and the total energy E align with quantum calculations [91–94, 96, 97, 125–130]. By leveraging the principle of locality [100, 101] mentioned above, the ML model can focus on local atomic environments, significantly reducing the computational cost while preserving quantum-level accuracy.

The tremendous success of ML methods in quantum MD simulations has spurred similar approaches to multi-scale dynamical modeling of other functional electronic systems in condensed matter physics [58, 59, 98, 99, 105, 131, 132]. In particular, the Behler-Parrinello (BP) ML scheme [91,92] was generalized to build effective magnetic energy or torque-field models with the accuracy of quantum calculations for itinerant

electron magnets [98,99,133,134]. Notably, large-scale dynamical simulations enabled by such ML models uncovered intriguing phase separation dynamics that results from the nontrivial interplay between electrons and local spins. While the conventional BP scheme can only represent conservative forces, a generalized potential theory for the Landau-Lifshitz equation allows one to extend the BP scheme to describe non-conserved spin torques that are crucial to the dynamical modeling of out-of-equilibrium itinerant spin systems [132].

In this chapter, we present ML torque models for itinerant magnets utilizing convolutional neural networks (CNNs) and fully connected neural networks with symmetry-constrained descriptors. CNNs are a class of neural networks characterized by their local connectivity, implemented through finite-sized convolutional kernels. Importantly, the convolution operation inherently incorporates the locality principle into the ML model, offering an efficient implementation of torque models that can be scaled to larger systems. Our CNN model is designed to directly predict the vector torque field at each site, avoiding the need for the introduction of local energies as required in the BP scheme. To incorporate spin-rotational and lattice symmetries, we employ data augmentation techniques, ensuring the CNN model respects these fundamental physical properties. We demonstrate the efficacy of our approach using an itinerant spin model that exhibits a skyrmion crystal phase under intermediate magnetic fields. Our results show that dynamical simulations with magnetic torques computed from the trained CNN model faithfully reproduce the relaxation processes

observed in itinerant spin systems. Furthermore, despite being trained on datasets derived from small systems, the CNN model effectively stabilizes a skyrmion lattice (SkL) in larger systems, demonstrating both the transferability and scalability of our ML approach. This capability highlights the potential of CNN-based torque models for studying complex magnetic textures in practical, large-scale scenarios.

3.2 s-d Model for Itinerant Magnets

The magnetization dynamics in spin systems are governed by the Landau-Lifshitz-Gilbert (LLG) equation [135]:

$$\frac{d\mathbf{S}_i}{dt} = \mathbf{T}_i - \alpha \mathbf{S}_i \times \mathbf{T}_i + \boldsymbol{\tau}_i, \tag{3.1}$$

where $\mathbf{T}_i$ represents the magnetic torque, defined as

$$\mathbf{T}_i = \gamma \mathbf{S}_i \times \mathbf{H}_i. \tag{3.2}$$

Here, γ is the gyromagnetic ratio, $\mathbf{H}_i$ is the effective exchange field acting on i-th spin, α is the damping coefficient, and $\boldsymbol{\tau}_i(t) = \mathbf{S}_i \times \boldsymbol{\eta}_i(t)$ is a fluctuating torque generated by a random local field $\boldsymbol{\eta}_i$ with zero mean. The stochastic field $\boldsymbol{\eta}_i$ is assumed to be a Gaussian random variable with variance determined by α and temperature T, as dictated by the fluctuation-dissipation theorem. LLG simulations are widely employed to study dynamical phenomena in various magnetic systems, including spin waves in exotic magnetic phases and the dynamics of skyrmions and other spin textures.

For adiabatic spin dynamics, the local exchange field is given by the derivative of

the system energy $E = E(\mathbf{S}_i)$:

$$\mathbf{H}_i = -\frac{\partial E}{\partial \mathbf{S}_i}. \tag{3.3}$$

In magnetic insulators, spin interactions are often short-ranged. The resultant magnetic energy typically involves bilinear interactions between a few nearest-neighbor spins on the lattice, such as $E = \sum_{ij} \left(J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j + \mathbf{D}_{ij} \cdot \mathbf{S}_i \times \mathbf{S}_j \right)$, where J_{ij} denotes the isotropic Heisenberg exchange interaction and $\mathbf{D}_{ij}$ represents the anisotropic exchange, also known as the Dzyaloshinskii-Moriya interaction [123,124]. The exchange field in such models can be expressed as $\mathbf{H}_i = -\sum_j \left(J_{ij} \mathbf{S}_j + \mathbf{D}_{ij} \times \mathbf{S}_j \right)$, where the summation is restricted to the few nearest neighbors, making it computationally efficient for large-scale LLG simulations.

In metallic magnets, the exchange fields originate from interactions between local spins and itinerant electrons. We consider spin dynamics under the adiabatic approximation, analogous to the Born-Oppenheimer approximation in quantum molecular dynamics [87]. In this limit, the relaxation of electrons is assumed to be much faster than the dynamics of local magnetic moments. Consequently, the magnetic energy E in Eq. (3.3) can be obtained by freezing the spin configuration and integrating out the electronic degrees of freedom. The resulting spin-dependent energy function $E = E(\mathbf{S}_i)$ can be viewed as a potential energy surface (PES) in the high-dimensional spin space, similar to the PES in Born-Oppenheimer MD simulations. Practically, calculating this magnetic PES involves solving the electronic structure problem, which

depends on the instantaneous spin configuration $\{\mathbf{S}_i(t)\}$.

To illustrate, we consider a generic single-band s-d model for itinerant magnets. The s-d model describes the interaction between itinerant s-band electrons and localized magnetic moments $\mathbf{S}_i$ of d-electrons. The Hamiltonian is given by:

$$\mathcal{H} = \sum_{ij} \sum_{\alpha=\uparrow,\downarrow} t_{ij} c^\dagger_{i\alpha} c_{j\alpha} - J \sum_i \sum_{\alpha,\beta=\uparrow,\downarrow} \mathbf{S}_i \cdot c^\dagger_{i\alpha} \boldsymbol{\sigma}_{\alpha\beta} c_{i\beta}, \tag{3.4}$$

where $c^\dagger_{i\alpha}$ and $c_{i\alpha}$ are the creation and annihilation operators for an electron with spin $\alpha =\uparrow,\downarrow$ at site i, t_{ij} is the electron hopping constant between sites (i,j), and J is the strength of local Hund's coupling between the electron spin and the magnetic moment $\mathbf{S}_i$ of localized d-electrons. For most skyrmion magnets, these local magnetic moments can be well approximated as classical spins of fixed length $|\mathbf{S}_i| = S$.

For weak Hund's coupling $J \ll t_{ij}$, the effective spin energy can be derived by integrating out the electrons via second-order perturbation theory, leading to a long-range oscillatory interaction, similar to the Ruderman-Kittel-Kasuya-Yosida (RKKY) interaction [136–138]. However, for intermediate and strong Hund's coupling, the effective energy used in force calculations in Eq. (3.3) must be obtained by integrating out electrons on-the-fly:

$$E = \langle \mathcal{H} \rangle = \mathrm{Tr}(\rho \mathcal{H}), \tag{3.5}$$

where $\rho = \exp(-\mathcal{H}/k_B T)$ is the density matrix of the equilibrium electron liquid under the adiabatic approximation. Calculating the density matrix, in the absence of electron-electron interactions, amounts to solving a disordered tight-binding Hamilto-

nian for a given spin configuration. The standard approach for solving tight-binding models involves exact diagonalization, which scales cubically with the system size. Therefore, repeated exact diagonalization calculations for large-scale LLG simulations of the s-d model can become computationally prohibitive.

As discussed above, the BP scheme has been extended to develop ML-based models for the effective spin energy $E(\{\mathbf{S}_i\})$ of itinerant magnets [98,99,132–134]. In this approach, the total energy is partitioned into local contributions:

$$E = \sum_i \epsilon_i = \sum_i \varepsilon(\mathcal{C}_i), \tag{3.6}$$

where the local energy $\epsilon_i = \varepsilon(\mathcal{C}_i)$ is associated with the i-th lattice site and depends only on the spin configuration $\mathcal{C}_i = \{\mathbf{S}_j \mid |\mathbf{r}_j - \mathbf{r}_i| < r_c\}$ in its neighborhood. This local energy function $\varepsilon(\mathcal{C}_i)$ serves as the building block of the magnetic PES. Importantly, the complex dependence of the PES on neighboring spins is approximated using ML models [98, 99, 132]. To preserve SO(3) spin rotation symmetry, the inner product between spin pairs $b_{jk} = \mathbf{S}_j \cdot \mathbf{S}_k$ and the scalar product of spin triplets $\chi_{jkl} = \mathbf{S}_j \cdot (\mathbf{S}_k \times \mathbf{S}_l)$ within the neighborhood are used as features for the neural network. Finally, the exchange fields $\mathbf{H}_i$ acting on the spins are obtained by applying automatic differentiation to the ML energy model.

3.3 Machine Learning Architecture

The BP-type schemes described here provide energy-based ML models for force field calculations, with the total energy partitioned into local contributions ϵ_i. These con-

tributions cannot be directly computed from the electronic structure methods used to generate the training dataset, complicating the training process. The loss function L, typically constructed from MSE or force-based metrics (in this case, spin torque fields), depends indirectly on the predicted energy through automatic differentiation. However, introducing such intermediate local energies adds uncertainties that complicate the training of BP-type models.

While BP-type schemes explicitly include the physical constraint of conservative forces, they are inherently restricted to modeling only conservative forces. In this section, we present alternative ML approaches that directly predict vector forces without relying on intermediate energy calculations.

3.3.1 Convolutional Neural Network Model

The fact that spins in metallic magnets are defined on well-known lattices suggests that spin configurations can be treated as generalized "images", which can then be processed using image-processing techniques, such as CNNs. Below, we present a CNN model for directly predicting the torques $\mathbf{T}_i$ that drive spin dynamics. As illustrated in Figure 3.1, the proposed network takes the spin configuration $\{\mathbf{S}_i\}$ on the lattice as input and returns the torques $\{\mathbf{T}_i\}$ as output. The model comprises multiple convolutional layers f_m with associated activation (nonlinearity) layers σ_m, capturing the nonlinear relationship between $\{\mathbf{S}_i\}$ and $\{\mathbf{T}i\}$. The CNN model is described as a composition of these layers: $f_{\text{CNN}} = (\sigma_L \circ f_L) \circ \cdots \circ (\sigma_1 \circ f_1)$, where L is the depth of the CNN.

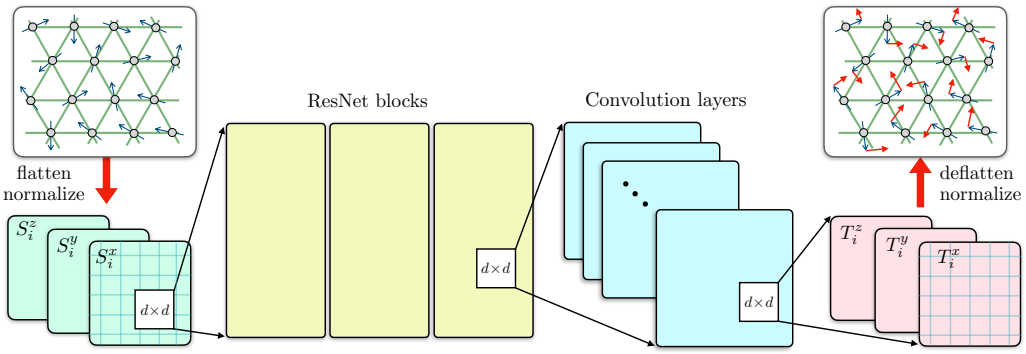

Figure 3.1: Schematic diagram of the CNN-based ML model for spin-torque prediction of itinerant electron magnets. The spin configuration $\{\mathbf{S}_i\}$ on a lattice is first flattened into three arrays, corresponding to the three components of spins, which are input to a series of ResNet blocks. Details of the ResNet are presented in Fig. 3.2. The output of the ResNet blocks is processed through additional convolutional layers, resulting in three arrays that, once reshaped, correspond to the torques $\{\mathbf{T}_i\}$ driving the spin dynamics.

Each convolutional layer f_m maps an input vector field $V \in C^\infty(\mathbb{R}^2, \mathbb{R}^d)$ onto an output vector field $W \in C^\infty(\mathbb{R}^2, \mathbb{R}^k)$ by convolving a *kernel* tensor field $h_m(X) := h(X; \theta_m)$, with trainable parameters θ_m, via the convolution operation:

$$W(\mathbf{r}) := \int_{\mathbb{R}^2} V(\mathbf{q}) h_m(\mathbf{r} - \mathbf{q}) d\mathbf{q}. \tag{3.7}$$

Each vector element of W then undergoes an activation function $\sigma_m : \mathbb{R} \to \mathbb{R}$ to produce the output vector field $A \in C^\infty(\mathbb{R}^2, \mathbb{R}^k)$, called activation maps. In this work, we use the rectified linear unit (ReLU) activation function [139]:

$$\sigma_m(x) := \max(0, x). \tag{3.8}$$

for $m = 1, \ldots, L-1$. Note that the final layer f_L does not have an associated activation function, or technically, $\sigma_L(x) = x$.

Typically, the support of a kernel $\mathrm{supp}(h_m)$, i.e., the region where h_m has nonzero values (also known as the receptive field), is limited to a small region (e.g., 5×5 lattice sites) such that the activation response $W(\mathbf{r})$, and thus $A(\mathbf{r})$, at position $\mathbf{r}$ is influenced only by the input values in the close vicinity of $\mathbf{r}$. This design follows the principle of locality, where local physical quantities, such as the spin torque $\mathbf{T}_i$, are predominantly determined by the local environment:

$$\mathbf{T}_i = \boldsymbol{\mathcal{T}}(\mathcal{C}_i), \tag{3.9}$$

where $\mathcal{C}_i$ represents the magnetic environment around site i, and the function $\boldsymbol{\mathcal{T}}(\cdot)$ is modeled by the CNN. The size of the neighborhood $\mathcal{C}_i$ is determined by the sizes of kernels and the number of convolutional layers.

The hierarchical structure of the convolutional layers enables multi-scale modeling of the spin-torque relationship. Earlier layers represent local, primitive patterns, while deeper layers represent more global, complex patterns. The stacked convolutional layers also produce an effective receptive field (ERF) that grows with the depth of the network, allowing the model to capture interactions over larger spatial extents. A purely convolutional CNN, without fully connected layers, restricts the overall receptive field to a predetermined lattice size, providing the benefit of built-in locality. This design also enables scalability, as the CNN model can be applied to larger lattice systems without retraining, making it naturally suited for studying systems based on locality principles.

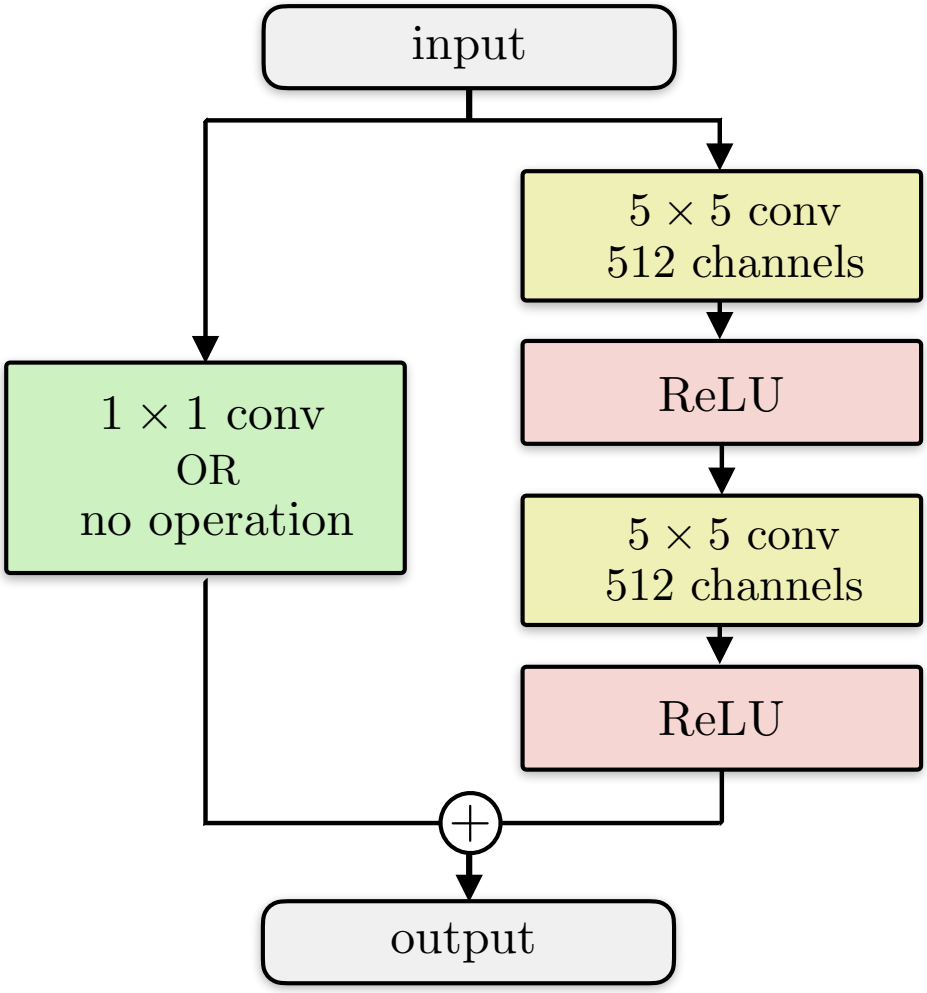

Figure 3.2: Diagram of a ResNet block. The input to a ResNet block goes through two pathways: the skip connection, where no operation is performed if input and output have the same number of channels, or a 1×1 convolution otherwise; and the main connection, where two 5×5 convolution-ReLU activation blocks are stacked. The outputs of both pathways are then added together.

The architecture of the ResNet block is shown in Fig. 3.2. Similar to the original ResNet, the input undergoes two separate pathways. One pathway is a skip connection where the input is directly copied if no dimensionality change is required, or a 1×1 convolution if needed. The other pathway contains two convolutional layers, each followed by ReLU activation [139], which develop feature representations of the input vector field at each local neighborhood. The outputs of these two pathways are then added to produce the overall output of the ResNet block. Note that we do not employ batch normalization, as it was found to overly regularize the network, leading to underfitting.

Our training and testing datasets consist of 60 independent spin dynamics simulations performed on a 48×48 triangular lattice. The parameters used for the s-d Hamiltonian in Eq. (3.4) include a nearest-neighbor hopping term $t_1 = 1$, which serves as the reference unit for energy, and a third-neighbor hopping term $t_3 = -0.85$ to stabilize a triple-Q magnetic order that underpins the SkL phase [140]. The electron-spin coupling constant is $J = 1$, the electron chemical potential is $\mu = -3.5$, and an external magnetic field $H_{\text{ext}} = 0.005$ was applied to break time-reversal symmetry and induce the SkL phase [140]. As discussed above, the exchange fields $\mathbf{H}_i$ acting on the spins are obtained by solving the electron Hamiltonian. Specifically, using Eq. (3.3) and the s-d Hamiltonian in Eq. (3.4), the exchange fields are given by

$$\mathbf{H}_i = J \sum_{\alpha,\beta=\uparrow,\downarrow} \boldsymbol{\sigma}_{\alpha\beta}\, \rho_{i\beta,i\alpha}, \tag{3.10}$$

where $\rho_{i\alpha,j\beta} := \langle c^\dagger_{j\beta} c_{i\alpha} \rangle$ represents the electron correlation function, or single-electron density matrix. The kernel polynomial method (KPM) [90,141] was used to compute the electron density matrix for generating the training dataset. KPM is significantly more efficient compared to ED and is considered numerically exact when a large number of Chebyshev polynomials and random vectors are used.

The timescale of the precessional dynamics of the LLG equation (3.1) is given by $t_0 = (\gamma J S)^{-1}$, where γ is the gyromagnetic ratio, J is the electron-spin coupling, and S represents the magnitude of the localized magnetic moments. The damping term introduces an additional timescale, $t_{\text{damping}} = t_0/\alpha$, which characterizes the rate of energy dissipation, where α is a dimensionless damping coefficient. In this work, simulation time is measured in units of t_0, with a damping coefficient $\alpha = 0.05$.

The initial conditions for the simulations are categorized into two types. The first type, referred to as perturbed SkL, consists of a periodic SkL with random noise added to the spins. The second type consists of randomly initialized spin configurations. For each type of initial condition, a total of 30 simulations were performed, with each simulation comprising 5,000 time steps. To integrate the LLG equation (3.1), we used a semi-implicit second-order scheme [142] that preserves the spin length, with a timestep $\Delta t = 0.1$.

The spins and their corresponding exchange fields at all lattice sites were collected every 10 steps of the simulation. To focus on training for the electron-induced exchange field, the external constant field $H_{\text{ext}} = 0.005$ in the z direction was removed.

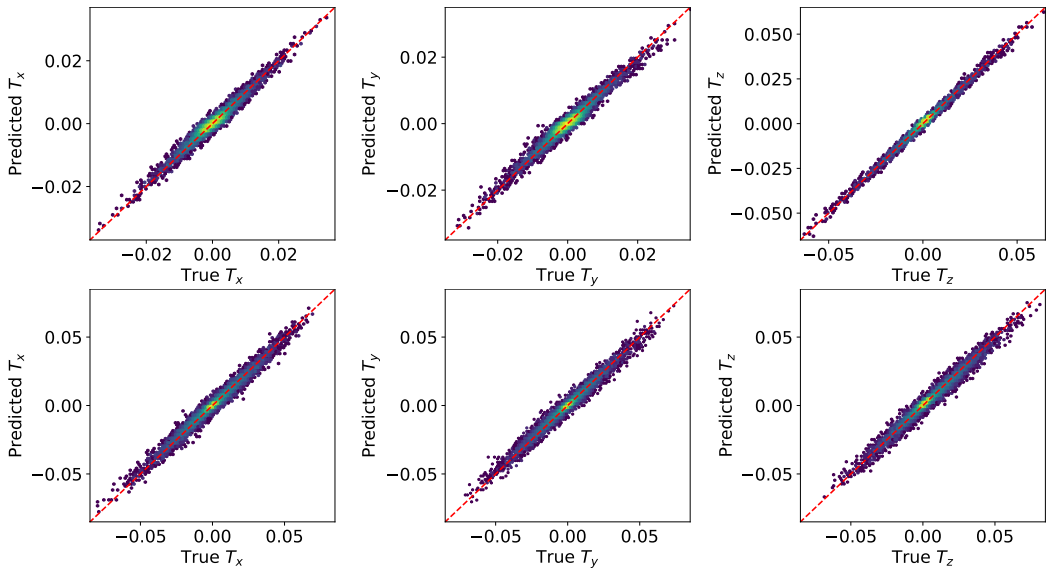

Figure 3.3: Predicted spin torque components (T_x, T_y, T_z) versus ground truth components from the testing set. The red-dotted diagonal lines indicate perfect prediction. The top row shows the prediction results based on spin configurations obtained from LLG simulations of a perturbed SkL. Results from LLG simulations with random initial states are shown in the bottom row.

The field $\mathbf{H}_i$ was then decomposed into components parallel and perpendicular to the spin direction, and only the perpendicular component—equivalent to the torque $\mathbf{T}_i$—was retained, as the parallel component has no effect on the evolution of the spin configuration and is approximately two orders of magnitude larger than the perpendicular component. The perpendicular fields were subsequently normalized to have a mean magnitude of 1 over the entire dataset. 70% of the entire dataset was used for training, while the remaining 30% was used for validation. The split of the dataset was stratified to ensure that both the training and testing sets contained the same proportion of the two types of simulations.

The triangular-lattice s-d Hamiltonian in Eq. (3.4) is invariant under two indepen-

dent symmetry groups: the SO(3)/SU(2) rotation of spins and the D6 point group of the triangular lattice. The rotation symmetry refers to the global rotation of local magnetic moments $\mathbf{S}_i \to \mathcal{R} \cdot \mathbf{S}_i$ (treated as classical vectors), along with a simultaneous unitary transformation of the electron spinor $\hat{c}_{i\alpha} \to \hat{U}_{\alpha\beta}\hat{c}_{i\beta}$, where $\mathcal{R}$ is an orthogonal 3×3 matrix and $\hat{U} = \hat{U}(\mathcal{R})$ is the corresponding 2×2 unitary rotation operator. The ML model, corresponding to an effective force-field model obtained by integrating out the electrons, must preserve the SO(3) rotation symmetry of the spins, implying that under a uniform rotation $\mathcal{R}$ of all spins in the neighborhood, the ML-predicted spin torques should undergo the same rotation transformation $\mathbf{T}_i \to \mathcal{R} \cdot \mathbf{T}_i$. Under a symmetry operation g of the D_6 point group centered at a lattice point, both spins and torques transform according to the D_6 point group as follows: $\mathbf{S}_i \to \mathbf{S}_j$ and $\mathbf{T}_i \to \mathbf{T}_j$, where the lattice points $\mathbf{r}_j = \mathcal{R}(g) \cdot \mathbf{r}_i$, and $\mathcal{R}(g)$ denotes the 3×3 matrix corresponding to g.

To incorporate both symmetries into the CNN model, we introduced data augmentation during the training phase. Specifically, for each input spin configuration and its corresponding torque field, a random SO(3) rotation was applied to the spins $\mathbf{S}_i$, and a random D_6 symmetry operation was applied to the lattice points. The same symmetry operations, for both spin-space and real-space lattice, were also applied to the torque fields $\{\mathbf{T}_i\}$. These additional symmetry-generated input/output configurations were included alongside the original data for supervised training. Unlike previous ML models, where the symmetry is explicitly included through descriptors,

our data-driven approach enforces the symmetry of the itinerant electron Hamiltonian statistically.

Since the torques in the dataset could differ by at most an order of magnitude, we found that conventional mean absolute error or mean square error loss functions did not perform well. Instead, we adopted a mean percentage absolute loss function:

$$L = \frac{1}{N}\sum_{i=1}^{N}\frac{|T_i^x - \hat{T}_i^x| + |T_i^y - \hat{T}_i^y| + |T_i^z - \hat{T}_i^z|}{|\hat{\mathbf{T}}_i|}, \tag{3.11}$$

where N is the total number of lattice sites within each batch, summed across all lattices, $\hat{\mathbf{T}}_i$ is the ground truth field vector at the i-th lattice site, and $\mathbf{T}_i = (T_i^x, T_i^y, T_i^z)$ is the predicted field vector.

The model was trained using the Adam optimizer [143], with an initial learning rate of 10^{-3}, which was later reduced to 10^{-6} when the loss plateaued on the testing set. No regularization methods, such as dropout or weight decay, were used, and there was no evidence of overfitting when comparing loss values between the training and testing sets. The model and training process were implemented in PyTorch [144], and training was performed on an Nvidia A100 GPU for approximately 72 hours.

The spin torques $\mathbf{T}_i$ predicted by the trained CNN model are compared against the ground truth in Fig. 3.3 using configurations from the test dataset. Two types of testing data are employed for this benchmark: LLG simulations of an initially perturbed SkL state, and LLG simulations starting from random spin configurations. In both cases, the predicted torque components closely follow the ground truth with

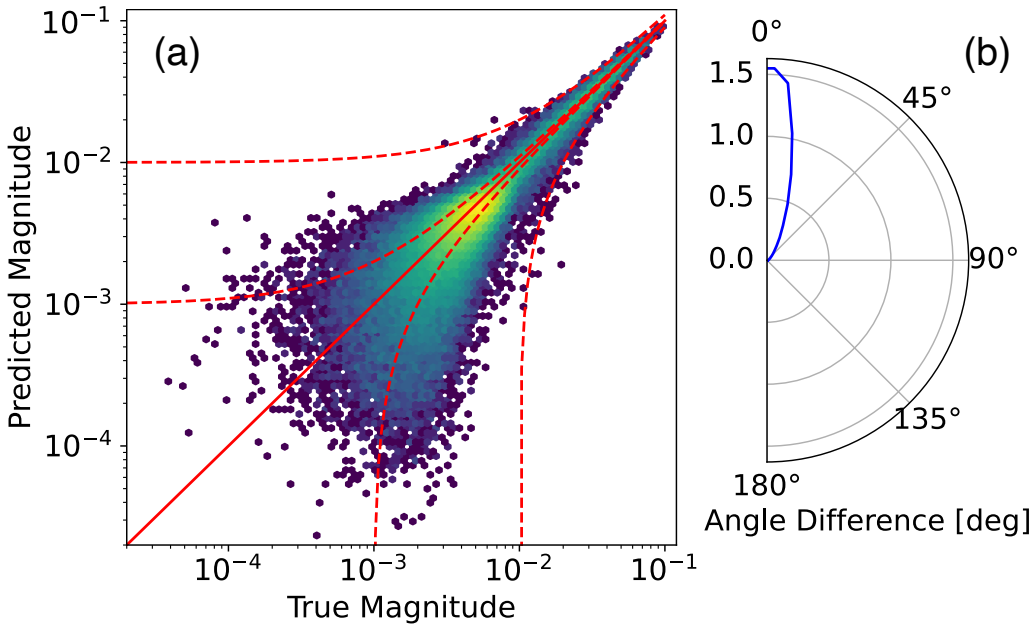

Figure 3.4: (a) Comparison of predicted field vector magnitude against the ground truth. The red line indicates perfect agreement between predictions and ground truth, while the outer red dotted line represents a deviation of 10^{-2} from the ground truth magnitude, and the inner one represents a deviation of 10^{-3}. The color indicates the logarithmic density. (b) Angular difference between the ground truth field vector and the predicted field vector.

comparable variance across the entire range. It is notable that the torque component values in the random spin case span a range nearly twice as large as that of the SkL case. As expected, the ML model performs better for the SkL simulations since these spin configurations correspond to a relatively small and specialized subset of the overall state space. Nevertheless, a reasonably good agreement was obtained even for the test dataset with completely random initial spins.

We further examine the magnitude of the predicted torques compared to the ground truth, as well as the angle between the predicted field vector and the ground

truth vector in Fig. 3.4. Again, overall satisfactory agreement was observed, with the majority of predictions closely aligned with or symmetrically distributed around the ground truth values. It should be noted that, due to the distortion introduced by the logarithmic function, the same deviation from the ground truth at large and small magnitudes may appear asymmetric, creating an apparent "bias" towards smaller values. To address this, two red dotted lines indicating constant deviations of 10^{-2} (outer) and 10^{-3} (inner) have been added in Fig. 3.4(a). Even at large magnitudes, where the error of the ML model is greatest, the difference in field vector magnitude is almost always less than 10^{-2}. At smaller magnitudes, the difference in field vector magnitude is typically less than 4×10^{-3}, with most values around 10^{-3}. No noticeable bias was found in our ML prediction results. The ML-predicted vectors are also well aligned with the ground truth field vectors. As shown in Fig. 3.4(b), most vectors exhibit an angular difference of less than 10°, and it is extremely rare for a predicted vector to deviate by more than 30° from its ground truth counterpart.

3.3.2 Lattice Descriptor

To ensure strict symmetry preservation for the triangular lattice, which belongs to the D_6 point group, we employ bond and chirality variables around a spin $\mathbf{S}_i$ as illustrated in Fig. 3.5. These variables provide a basis for a high-dimensional representation of the D_6 group, which can be decomposed into its IRs. Specifically, linear combinations of bond and chirality variables form the basis of each IR, denoted by f_r^{Γ}, where Γ labels the IR and r indicates the multiplicity within the IR.

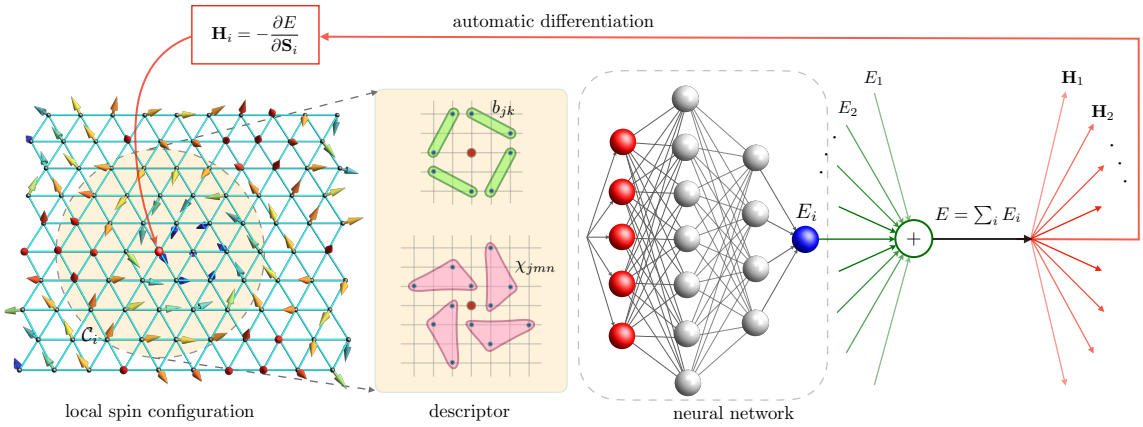

Figure 3.5: The architecture for integrating a NN with LLLG dynamics simulation of the s-d model using the BP scheme. Under the adiabatic approximation, for a given local spin during an LLG step, we train a fully-connected NN, denoted as $y_\theta(\mathbf{x})$, to predict the local energy E_i based on the configuration $\{\mathcal{C}_i\}$ of the surrounding spins, resulting in $E_i = y_\theta(\mathbf{x}|\{\mathcal{C}_i\})$. By combining the BP scheme with Eq.(3.3) and Eq.(3.6), the local exchange field $\{\mathbf{H}_i\}$ is computed through automatic differentiation of the total energy $E = \sum_i E_i$. The input $\mathbf{x}$ for the NN includes descriptors such as the bond $\{b_{jk}\}$ and chirality $\{\chi_{jmn}\}$, which are selected to maintain the symmetry properties of the system.

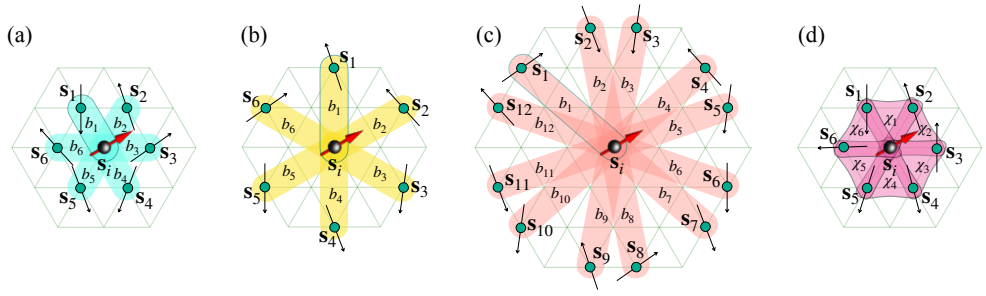

Figure 3.6: Examples of different neighborhood spin configurations used to compute the symmetry-invariant descriptors. (a) The six bond variables $b_j = \mathbf{S}_i \cdot \mathbf{S}_j$, where $\{\mathbf{S}_j || \mathbf{r}_i - \mathbf{r}_j| = a\}$, form the basis of a six-dimensional reducible representation of the D_6 point group. (b) Similarly, the six bond variables $b_j = \mathbf{S}_i \cdot \mathbf{S}_j$, where $\{\mathbf{S}_j || \mathbf{r}_i - \mathbf{r}_j| = \sqrt{3}a\}$, form the basis of a reducible six-dimensional representation of D_6. (c) The twelve bond variables $b_j = \mathbf{S}_i \cdot \mathbf{S}_j$, where $\{\mathbf{S}_j || \mathbf{r}_i - \mathbf{r}_j| = \sqrt{7}a\}$, form the basis of a twelve-dimensional reducible representation. (d) The six scalar chirality variables $\chi_j = \mathbf{S}_i \cdot \mathbf{S}_m \times \mathbf{S}_n$, where $\{\mathbf{S}_{m,n} || \mathbf{r}_i - \mathbf{r}_{m,n}| = a\}$, form the basis of a six-dimensional reducible representation.

Moreover, the generalized coordinates $\{G_l\}$, which are invariant under both lattice symmetry operations and SO(3) rotations, can be derived from the amplitudes and relative phases of these IR bases. In this framework, the generalized coordinates $\{G_l\}$ serve as inputs to the NN to predict the local energy E_i. Both the input and output of the NN are designed to ensure that the symmetries of the s-d model are preserved, which is crucial for obtaining physically consistent results.

Next, the magnetic descriptor is derived from the local spin environment through the following process.

$$\{\mathcal{C}_i\} \rightarrow \{b_{jk}, \chi_{jmn}\} \rightarrow \{\boldsymbol{\xi}_r^\Gamma\} \rightarrow \{G_l = p_r^\Gamma, \beta_r^\Gamma\} \tag{3.12}$$

where $\{\mathcal{C}_i\}$ represents the spin configurations in the neighborhood of i-th site, $\{b_{jk}, \chi_{jmn}\}$ are the bond $b_{jk} = \mathbf{S}_j \times \mathbf{S}_k$ and chirality variable $\chi_{jmn} = \mathbf{S}_j \cdot \mathbf{S}_m \times \mathbf{S}_n$ calculated from

the $\{\mathcal{C}_i\}$, $\{\xi_r^\Gamma\}$ comprises basis functions composed of IRs, and $\{G_l = p_r^\Gamma, \beta_r^\Gamma\}$ represents the generalized variables, or the magnetic descriptor, which serve as inputs to the NN model.

The bond, chirality, and local energy variables determined by the ML model possess invariance under any spin rotation. Consequently, the generalized variables $\{G_l\}$ that are constructed using the IRs are only required to maintain the symmetries of the triangular lattice. However, as the number of spins in the neighborhood increases, the number of bond and chirality variables also increases factorially. Therefore, only bond and chirality variables that meet specific criteria are utilized, as follows:

$$\begin{cases} b_{ij} = \mathbf{S}_i \cdot \mathbf{S}_j & |\mathbf{r}_i - \mathbf{r}_j| \leq r_1 \\ b_{mn} = \mathbf{S}_m \cdot \mathbf{S}_n & |\mathbf{r}_i - \mathbf{r}_{m,n}| \leq r_1 \\ \chi_{imn} = \mathbf{S}_i \cdot \mathbf{S}_m \times \mathbf{S}_n & |\mathbf{r}_m - \mathbf{r}_n| \leq r_2 \end{cases} \tag{3.13}$$

where the cutoff radii are set to $r_1 = 6a$ and $r_2 = 2a$ (a is the lattice constant) in our implementation.

The incorporation of discrete point group symmetry into the descriptor is achieved using the bispectrum method, which involves the use of triple products of the basis functions or coefficients of the IRs of the symmetry group. To obtain all the IR bases of the magnetic environment, the finite reducible representation, as determined by the bond and scalar chirality variables previously discussed, is considered. As the distance between a bond or chirality variable from the central site is maintained under the operations of the discrete symmetry group, the IRs are calculated in blocks

corresponding to bonds or chiralities of a fixed distance. For the triangular lattice, the dimension of each block is either 6 or 12. For example, consider the case shown in Fig. 3.6(a), where the 6 bonds $b_j = \mathbf{S}_i \cdot \mathbf{S}_j$ are between the nearest neighbor spins of i-th site $\{\mathbf{S}_j || \mathbf{r}_i - \mathbf{r}_j| = a\}$ and spin $\mathbf{S}_i$. This reducible representation $\{b_j\}$ can be decomposed into $6 = A_1 \bigoplus B_1 \bigoplus E_1 \bigoplus E_2$ with the following basis:

$$\begin{aligned}
\xi^{A_1} &= b_1 + b_2 + b_3 + b_4 + b_5 + b_6 \\
\xi^{B_1} &= b_1 - b_2 + b_3 - b_4 + b_5 - b_6 \\
\xi_x^{E_1} &= \tfrac{1}{2}(b_1 - b_2 - 2b_3 - b_4 + b_5 + 2b_6) \\
\xi_y^{E_1} &= \tfrac{\sqrt{3}}{2}(-b_1 - b_2 + b_4 + b_5) \\
\xi_x^{E_2} &= \tfrac{1}{2}(b_1 + b_2 - 2b_3 + b_4 + b_5 - 2b_6) \\
\xi_y^{E_2} &= \tfrac{\sqrt{3}}{2}(b_1 - b_2 + b_4 - b_5)
\end{aligned} \tag{3.14}$$

Similarly, for the case shown in Fig. 3.6(b), the 6 bonds $b_j = \mathbf{S}_i \cdot \mathbf{S}_j$ connect the second-nearest neighbor spins of i-th site $\{\mathbf{S}_j || \mathbf{r}_i - \mathbf{r}_j| = \sqrt{3}a\}$ and spin $\mathbf{S}_i$. This reducible representation $\{b_j\}$ can be decomposed into $6 = A_1 \bigoplus B_2 \bigoplus E_1 \bigoplus E_2$ with the following basis:

$$\begin{aligned}
\xi^{A_1} &= b_1 + b_2 + b_3 + b_4 + b_5 + b_6 \\
\xi^{B_2} &= b_1 - b_2 + b_3 - b_4 + b_5 - b_6 \\
\xi_x^{E_1} &= \tfrac{\sqrt{3}}{2}(b_2 + b_3 - b_5 - b_6) \\
\xi_y^{E_1} &= \tfrac{1}{2}(2b_1 + b_2 - b_3 - 2b_4 - b_5 + b_6)
\end{aligned}$$

$$\begin{aligned}\xi_x^{E_2} &= \tfrac{1}{2}(2b_1 - b_2 - b_3 + 2b_4 - b_5 - b_6)\\ \xi_y^{E_2} &= \tfrac{\sqrt{3}}{2}(-b_2 + b_3 - b_5 + b_6)\end{aligned} \tag{3.15}$$

In the case shown in Fig. 3.6(c), the 12 bonds $b_j = \mathbf{S}_i \cdot \mathbf{S}_j$ connect the fourth-nearest neighbor spins of i-th site $\{\mathbf{S}_j | |\mathbf{r}_i - \mathbf{r}_j| = \sqrt{7}a\}$ and spin $\mathbf{S}_i$. This reducible representation $\{b_j\}$ can be decomposed into $12 = A_1 \bigoplus A_2 \bigoplus B_1 \bigoplus B_2 \bigoplus 2E_1 \bigoplus 2E_2$ with the following basis:

$$\begin{aligned}\xi^{A_1} &= b_1 + b_2 + b_3 + b_4 + b_5 + b_6 + b_7 + b_8 + b_9 + b_{10} + b_{11} + b_{12}\\ \xi^{A_2} &= b_1 - b_2 + b_3 - b_4 + b_5 - b_6 + b_7 - b_8 + b_9 - b_{10} + b_{11} - b_{12}\\ \xi^{B_1} &= b_1 - b_2 - b_3 + b_4 + b_5 - b_6 - b_7 + b_8 + b_9 - b_{10} - b_{11} + b_{12}\\ \xi^{B_2} &= b_1 + b_2 - b_3 - b_4 + b_5 + b_6 - b_7 - b_8 + b_9 + b_{10} - b_{11} - b_{12}\\ \xi_x^{E_1} &= \tfrac{1}{2}(2b_1 - 2b_2 + b_3 - b_4 - b_5 + b_6 - 2b_7 + 2b_8 - b_9 + b_{10} + b_{11} - b12)\\ \xi_y^{E_1} &= \tfrac{\sqrt{3}}{2}(-b_3 + b_4 - b_5 + b_6 + b_9 - b_{10} + b_{11} - b_{12})\\ \xi_x^{E_1\prime} &= \tfrac{\sqrt{3}}{2}(b_3 + b_4 + b_5 + b_6 - b_9 - b_{10} - b_{11} - b_{12})\\ \xi_y^{E_1\prime} &= \tfrac{1}{2}(2b_1 + 2b_2 + b_3 + b_4 - b_5 - b_6 - 2b_7 - 2b_8 - b_9 - b_{10} + b_{11} + b_{12})\\ \xi_x^{E_2} &= \tfrac{1}{2}(2b_1 + 2b_2 - b_3 - b_4 - b_5 - b_6 + 2b_7 + 2b_8 - b_9 - b_{10} - b_{11} - b_{12})\\ \xi_y^{E_2} &= \tfrac{\sqrt{3}}{2}(-b_3 - b_4 + b_5 + b_6 - b_9 - b_{10} + b_{11} + b_{12})\\ \xi_x^{E_2\prime} &= \tfrac{\sqrt{3}}{2}(b_3 - b_4 - b_5 + b_6 + b_9 - b_{10} - b_{11} + b_{12})\\ \xi_y^{E_2\prime} &= \tfrac{1}{2}(2b_1 - 2b_2 - b_3 + b_4 - b_5 + b_6 + 2b_7 - 2b_8 - b_9 + b_{10} - b_{11} + b_{12})\end{aligned} \tag{3.16}$$

Note that the chirality variable changes sign under certain lattice rotations, such

as $\chi_{jmn} = -\chi_{jnm}$. For the 6 chirality variables $\chi_j = \mathbf{S}_i \cdot \mathbf{S}_m \times \mathbf{S}_n$ where $\{\mathbf{S}_{m,n} || \mathbf{r}_i - \mathbf{r}_j| = a\}$, shown in Fig. 3.6(d), this reducible representation $\{\chi_j\}$ can be decomposed into $6 = A_2 \bigoplus B_1 \bigoplus E_1 \bigoplus E_2$ with the following basis:

$$\begin{aligned}
\xi^{A_2} &= \chi_1 + \chi_2 + \chi_3 + \chi_4 + \chi_5 + \chi_6 \\
\xi^{B_1} &= \chi_1 - \chi_2 + \chi_3 - \chi_4 + \chi_5 - \chi_6 \\
\xi_x^{E_1} &= \tfrac{1}{2}(2\chi_1 + \chi_2 - \chi_3 - 2\chi_4 - \chi_5 + \chi_6) \\
\xi_y^{E_1} &= \tfrac{\sqrt{3}}{2}(-\chi_2 - \chi_3 + \chi_5 + \chi_6) \\
\xi_x^{E_2} &= \tfrac{\sqrt{3}}{2}(\chi_2 - \chi_3 + \chi_5 - \chi_6) \\
\xi_y^{E_2} &= \tfrac{1}{2}(2\chi_1 - \chi_2 - \chi_3 + 2\chi_4 - \chi_5 - \chi_6)
\end{aligned} \tag{3.17}$$

By applying the same procedures to all invariant blocks, one can obtain all the IRs of the bond/chirality variables $\{b_{jk}, \chi_{jmn}\}$ in the neighborhood $\{\mathcal{C}_i\}$. For ease of use, the basis functions of a given IR in the decomposition are arranged into a vector $\{\boldsymbol{\xi}_r^\Gamma\}$, where Γ labels the IR and r enumerates its multiple occurrences. With these basis functions, a set of invariants known as the power spectrum, $p_r^\Gamma = |\boldsymbol{\xi}_r^\Gamma|^2$, can be immediately obtained, representing the amplitudes of each individual IR coefficient. However, it is important to note that feature variables based solely on the power spectrum are incomplete, as they ignore the relative phases between different IRs. For instance, the relative "angle" between two IRs of the same type, $\cos\theta = (\boldsymbol{\xi}_{r_1}^\Gamma \cdot \boldsymbol{\xi}_{r_2}^\Gamma)/|\boldsymbol{\xi}_{r_1}^\Gamma||\boldsymbol{\xi}_{r_2}^\Gamma|$, is also an invariant of the symmetry group. Without this phase information, the NN model may incur additional error due to spurious

symmetry, as two IRs can rotate independently of each other.

A comprehensive approach for incorporating all necessary invariants, encompassing both amplitudes and relative phases, is the bispectrum method. By utilizing the coefficients of all relevant IRs $\{\boldsymbol{\xi}_r^\Gamma\}$, the bispectrum coefficients can be determined as follows:

$$B_{r,r_1,r_2}^{\Gamma,\Gamma_1,\Gamma_2} = C_{\alpha,\beta,\gamma}^{\Gamma;\Gamma_1,\Gamma_2} \xi_{r,\alpha}^{\Gamma} \xi_{r_1,\beta}^{\Gamma_1} \xi_{r_2,\gamma}^{\Gamma_2} \tag{3.18}$$

where $C_{\alpha,\beta,\gamma}^{\Gamma;\Gamma_1,\Gamma_2}$ are the Clebsch-Gordan coefficients of the point group. To address the issue of an excessive number of bispectrum coefficients, a modified descriptor has been implemented in this study. This descriptor utilizes reference basis functions $\xi_{\text{ref}}^{\Gamma}$ for each IR type of the point group. These reference basis functions are obtained by averaging large blocks of bond and chirality variables, thereby rendering them less susceptible to minor variations in the neighboring spin configurations. The relative "phase" of an IR is then defined as the projection of its basis functions onto the reference basis, $\beta_r^\Gamma = (\boldsymbol{\xi}_r^\Gamma \cdot \boldsymbol{\xi}_{\text{ref}}^\Gamma)/|\boldsymbol{\xi}_r^\Gamma||\boldsymbol{\xi}_{\text{ref}}^\Gamma|$. The generalized variables are a combination of the power spectrum coefficients and the relative phases, $\{G_l = p_r^\Gamma, \beta_r^\Gamma\}$. Using the aforementioned approach, we ensure that the input and output of the NN model remain invariant under the operations of the system's symmetry.

3.3.3 Neural Network

As shown in 3.5, a multilayer perceptron (MLP) NN model $f_\theta(\mathbf{x})$ was implemented using PyTorch. The model consisted of eight hidden layers, each containing a specific

number of neurons, namely $2048 \times 1024 \times 512 \times 256 \times 128 \times 64 \times 64 \times 64$. The input layer of the model was determined by the number of feature variables $\{G_l\}$, which in this case was 1806. The output layer of the model consisted of a single neuron, which was used to predict the local energy E_i. The ReLU activation function was employed between layers of the NN model. Given that the torque $\mathbf{T}_i = \mathbf{S}_i \times \mathbf{H}_i$ is the key component to derive the LLG dynamics, the MSE loss function used in the training of the model was

$$L = \sum_{i=1}^{N} |\mathbf{S}_i \times \mathbf{H}_i^{\text{exact}} - \mathbf{S}_i \times \mathbf{H}_i^{\text{ML}}|^2 \tag{3.19}$$

The optimization of the NN parameters, θ, was conducted using the Adam stochastic optimization algorithm with a learning rate that decreased exponentially as the number of training iterations increased, starting at 0.1 and decreasing to 0.0001. During the initial phase of training, the learning rate was progressively decreased by an order of magnitude for each increment of 10 epochs, up to a maximum of 30 epochs. Specifically, for the first 10 epochs, the learning rate was 0.1, for the next 10 epochs, the learning rate was 0.01, and for the final 10 epochs, the learning rate was 0.001. Subsequently, for all remaining epochs, the learning rate was fixed at 0.0001. To train the neural network, a dataset consisting of 40,000 snapshots obtained from precise simulations was utilized, and the training process was conducted over 200 epochs. To prevent overfitting, a 5-fold cross-validation strategy was employed.

Through training a NN model with 40,000 spin configurations from 40 independent

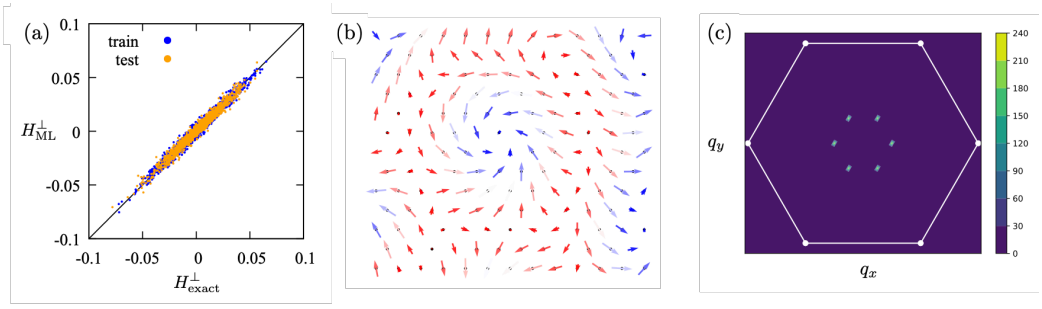

Figure 3.7: (a) Comparison between the exact local exchange force field $\mathbf{H}^{\perp}_{i,\mathrm{exact}}$ and the NN prediction $\mathbf{H}^{\perp}_{i,\mathrm{ML}}$ perpendicular to the same site spin $\mathbf{S}_i$ with the training datasets (blue) and the test datasets (orange). (b) Display examples of the intermediate configuration for the skyrmion phase from the ML-LLG simulations. (c) Demonstrate the structure factor of the final state of skyrmion phase in the first BZ.

simulations, we achieved a model with a MSE of 9.82×10^{-7} and observed no signs of overfitting, as depicted in Fig. 3.7(a). The intermediate spin configuration, which evolved from randomly initialized spins, clearly illustrates a swirling configuration of chiral spin structures, consistent with the skyrmion order, as shown in Fig. 3.7(b). Since there are three degenerate wave factors, we expect six structure factor (defined in the next section) peaks in the first Brillouin zone (BZ), corresponding to the k-point $Q(\pi/3, 0)$, due to the unit cell size of 12, as depicted in Fig. 3.7(c).

3.4 Results

A critical benchmark for evaluating the performance of the trained ML model is its ability to accurately capture the dynamical evolution of the itinerant spin model. To assess this, we incorporated the trained CNN and the NN with descriptors into LLG dynamics simulations and compared the outcomes with those obtained from

LLG simulations using the KPM [90, 141]. Specifically, we focused on simulations of a thermal quench process, where an initially random magnetic configuration was quenched to nearly zero temperature at time $t = 0$. The parameters of the s-d Hamiltonian were chosen to stabilize a spontaneous SkL ground state. Notably, the emergence of the SkL breaks both spin-rotation symmetry and lattice translational symmetry. The periodicity of this spatial modulation, particularly the lattice constant of the SkL, is determined by the underlying electron Fermi surface. While an SkL state can be conceptually described as a periodic array of particle-like spin textures, in practice, SkL phases often emerge due to an instability induced by quasi-nesting of the electron Fermi surface, leading to a multiple-Q magnetic order [140, 145, 146].

In this study, the geometry of the Fermi surface at a chemical potential of $\mu = -3.5$ exhibits significant segments that can be connected by three wave vectors: $\mathbf{Q}_1 = (\pi/3a, 0)$ and $\mathbf{Q}_{2,3} = \mathcal{R}_{\pm 2\pi/3} \cdot \mathbf{Q}_1$, which are related by symmetry operations of the D_6 group. Here, a represents the lattice constant of the underlying triangular lattice. This configuration indicates that the maximum energy gain through electron-spin coupling is achieved by spin helical orders associated with one of these three wave vectors. Further analysis suggests that the electron energy is minimized when all three wave vectors order simultaneously, resulting in the formation of an emergent triangular lattice of skyrmions.

3.4.1 Dynamics for Convolutional Neural Network Model

While the trained CNN model yields relatively accurate predictions for spin torques, small residual errors are still present, as described in the previous section. Statistically, these errors resemble the stochastic noise term, $\boldsymbol{\tau}_i(t)$, in the LLG equation (3.1). These site-dependent, fluctuating random torques are akin to thermal forces in Langevin dynamics, both arising from thermal fluctuations induced by coupling to a thermal bath. Consequently, while the temperature in the ML-LLG simulations was set to exactly zero, a very low but nonzero temperature $T = 0.001$ was introduced in the exact LLG dynamics to mimic the prediction error.

The relaxation of the magnet following the thermal quench is primarily governed by the formation of a triangular SkL. A perfect SkL is characterized by six Bragg peaks at $\mathbf{q} = \pm\mathbf{Q}_1$, $\pm\mathbf{Q}_2$, and $\pm\mathbf{Q}_3$ in momentum space. However, due to the local nature of spin interactions, the crystallization of skyrmions is inherently incoherent. Small skyrmion crystallites form randomly, separated by large domains of disordered structures. To quantitatively characterize this crystallization process, we compute the time-dependent spin structure factor, defined as the square of the Fourier transform of the spin field:

$$\mathcal{S}(\mathbf{q}, t) = \left\langle \left| \frac{1}{N} \sum_{i=1}^{N} \mathbf{S}_i(t) \exp(i\mathbf{q} \cdot \mathbf{r}_i) \right|^2 \right\rangle, \tag{3.20}$$

where $\langle \cdots \rangle$ denotes averaging over the thermal ensemble as well as initial conditions. The structure factor represents the Fourier transform of the spin-spin correlation

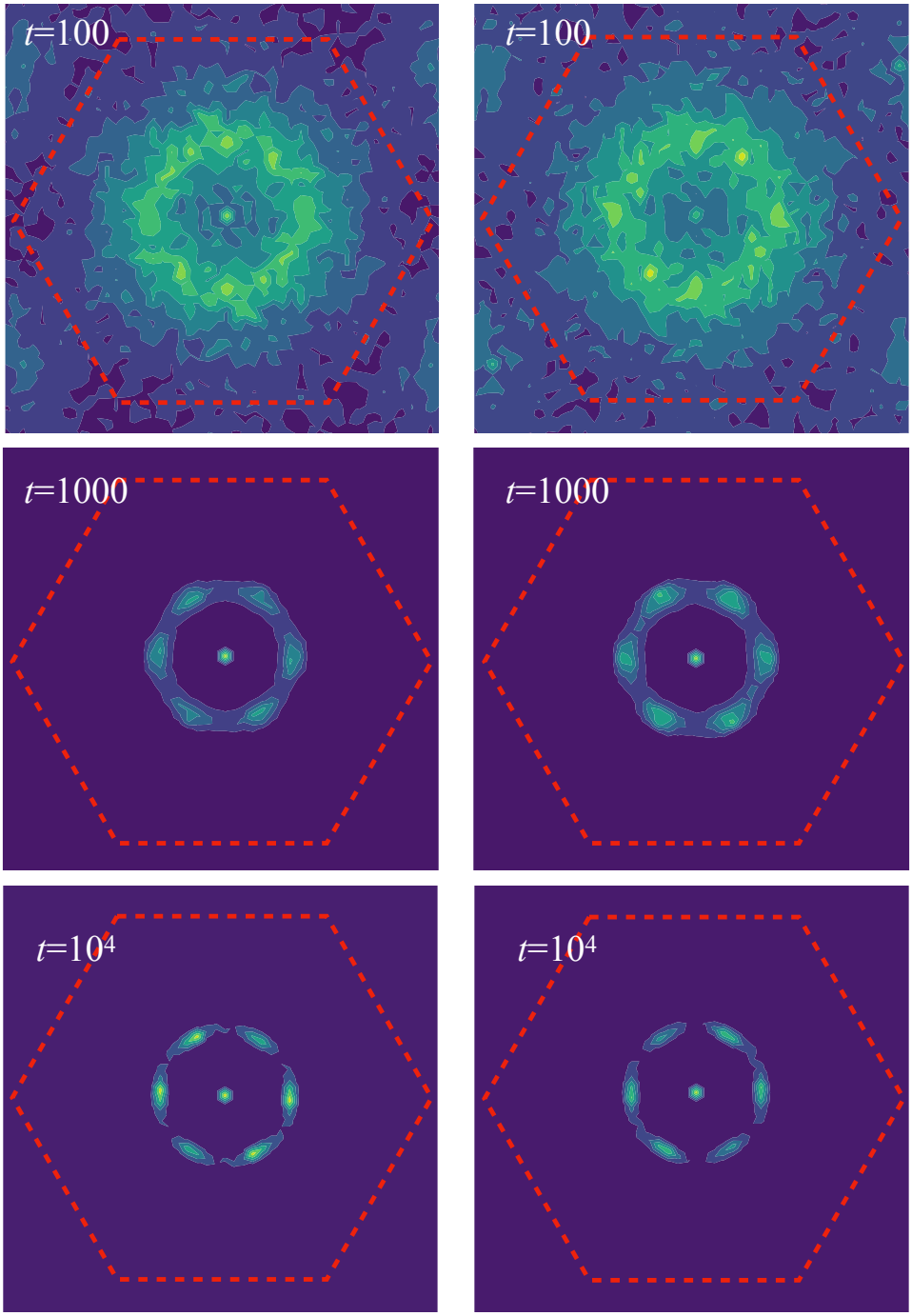

Figure 3.8: Comparison of spin structure factors obtained by averaging 30 independent LLG simulations based on KPM (left) and the CNN model (right). The same set of random initial conditions on a 48×48 triangular lattice was used in both simulations. The red dashed lines indicate the first BZ of the momentum space.

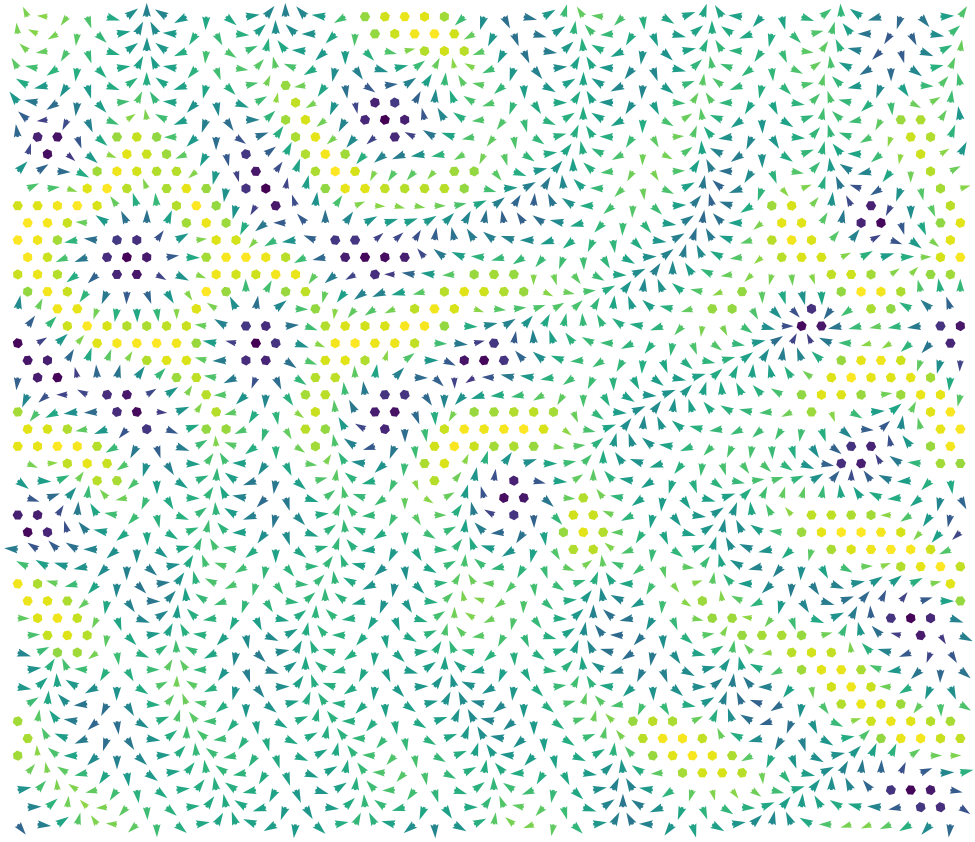

Figure 3.9: Snapshot of the spin configuration at the end of the LLG simulation with random initial conditions on a 48×48 triangular lattice.

function in real space and can be directly measured in neutron scattering experiments. The spin structure factors at various times after the quench, obtained from LLG simulations based on both KPM and CNN models, are shown in Fig. 3.8. Due to the stochastic nature of these simulations, the results were obtained by averaging over 30 independent runs. Overall, the LLG simulations using the trained CNN model exhibit strong agreement with those obtained from the numerically exact KPM.

Both simulations reveal that a ring-like structure emerges rapidly in the structure factor following the quench. The radius of the ring is approximately equal to the

magnitude of the three nesting wave vectors $\mathbf{Q}_\eta$, indicating the initial formation of skyrmions. As the system relaxes toward equilibrium, the ring-like structure becomes more distinct, and spectral weight begins to concentrate at the six spots corresponding to the $\pm\mathbf{Q}_\eta$ wave vectors. Physically, the emergence of the six broad segments reflects the growth of SkL domains. The size of these intermediate skyrmion crystallites can be inferred from the width of the six spots. However, both simulations indicate that even at a late stage of equilibration, the structure factor exhibits only six diffuse peaks at the nesting wave vectors, rather than the sharp Bragg peaks characteristic of a perfect SkL. The broad peaks at the late stage of phase ordering suggest an arrested growth of SkL domains in real space. A snapshot of the real-space spin configuration at $t = 10^4$ after the quench is shown in Fig. 3.9. The snapshot reveals small triangular clusters of skyrmions coexisting with stripe-like structures of varying orientations. These stripes, or helical spin states, correspond to the single-Q magnetic order, which represents metastable states of the s-d model.

This intriguing freezing phenomenon can be partially attributed to the frustration electron-mediated spin interactions. Another contributing factor is the degeneracy between skyrmions of opposite vorticity, or circulation of in-plane spins. The two opposite circulations correspond to topological winding numbers $w = \pm 1$ for the skyrmions. As previously discussed, spin-rotation symmetry is decoupled from the lattice in the s-d Hamiltonian (3.4), which serves as a minimal model for centrosymmetric itinerant magnets without spin-orbit coupling. Consequently, skyrmions with clock-

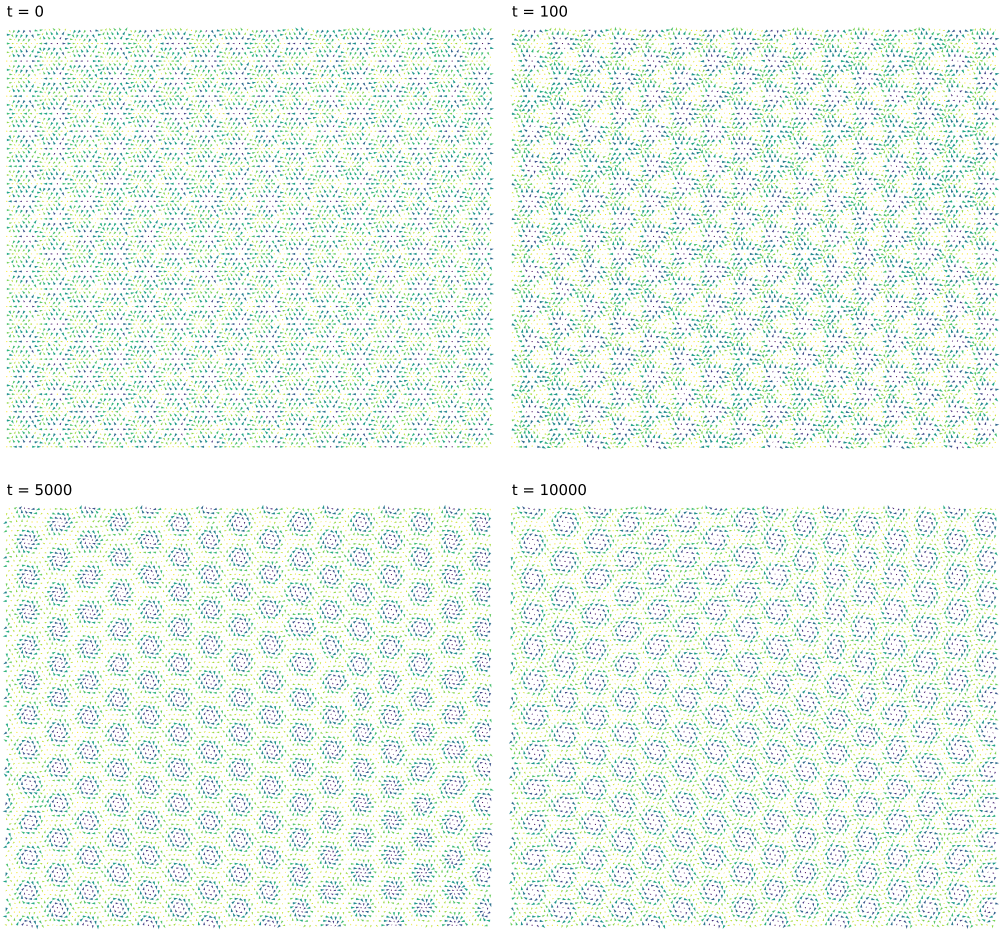

Figure 3.10: CNN-based LLG simulation on a 96×96 lattice demonstrating the restoration of a perturbed SkL. The CNN model was originally trained on a 48×48 lattice. The initial spin configuration is given by the SkL ansatz (3.21) with additional site-dependent random phases and amplitudes of S_z.

wise circulation are energetically degenerate with those exhibiting counter-clockwise circulation. This degeneracy implies that SkL domains of opposite circulations are nucleated with roughly equal probability after the thermal quench, and subsequent annihilation of skyrmions with opposite vorticity prevents the formation of a large, coherent SkL.

As discussed in Sec. 3.3.1, the CNN model, due to its locality property and fixed-

size kernels, can be directly scaled to larger lattice systems without retraining, enabling large-scale dynamical simulations beyond conventional approaches. Here, we demonstrate the scalability of the CNN spin-torque model by applying it to LLG simulations of large-scale SkL phases. Specifically, we conducted LLG simulations of a perturbed SkL state on a 96×96 lattice using a CNN model trained from simulations on a 48×48 lattice. The triangular SkL, characterized by three nesting wave vectors, can be understood as a superposition of three helical spin orders. Explicitly, a perfect SkL can be approximated by the following ansatz [140, 146]:

$$\begin{aligned} \mathbf{S}_i \sim & \left(\cos \mathcal{Q}_{1i} - \frac{1}{2} \cos \mathcal{Q}_{2i} - \frac{1}{2} \mathcal{Q}_{3i} \right) \hat{\mathbf{e}}_1 \\ & + \left(\frac{\sqrt{3}}{2} \cos \mathcal{Q}_{2i} - \frac{\sqrt{3}}{2} \cos \mathcal{Q}_{3i} \right) \hat{\mathbf{e}}_2 \\ & + \left[A \left(\sin \mathcal{Q}'_{1i} + \sin \mathcal{Q}'_{2i} + \sin \mathcal{Q}'_{3i} \right) + M \right] \hat{\mathbf{e}}_3, \end{aligned} \tag{3.21}$$

where $\hat{\mathbf{e}}_{1,2,3}$ are three orthogonal unit vectors, $\mathcal{Q}_{\eta i} = \mathbf{Q}_\eta \cdot \mathbf{r}_i$, and $\mathcal{Q}'_{\eta,i} = \mathcal{Q}_{\eta,i} + \phi$ are phase factors of the three helical orders, with ϕ, A, and M as fitting parameters. To demonstrate that the ML model can indeed stabilize the SkL, which is the ground state of our chosen s-d Hamiltonian, we initialize the system with a perturbed array of skyrmions, as shown in Fig. 3.10(a). The randomness in the initial state was introduced by allowing site-dependent parameters ϕ_i, A_i, and m_i, which were randomly generated in the SkL ansatz (3.21). Contrary to the completely random spin configurations used in the previous dynamical benchmark, this initial state preserves a coherent structure of skyrmion winding numbers. Since these topological numbers

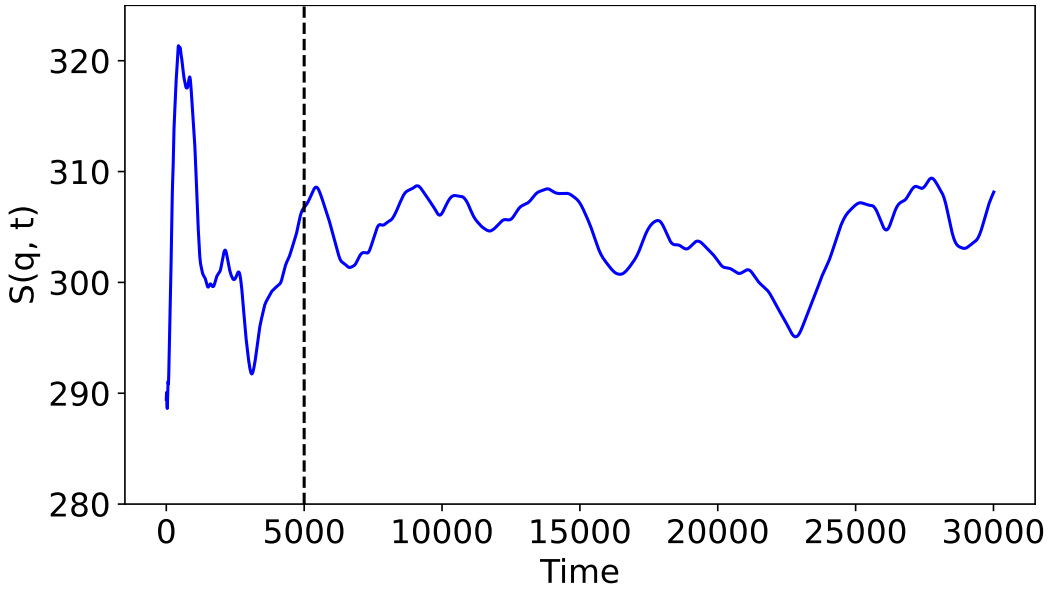

Figure 3.11: Evolution of the structure factor over time for a Skyrmion initial condition in a simulation extending far beyond the training duration. The dashed black line indicates the duration of the training simulation. Our ML-based LLG simulation maintains a stable structure factor for more than five times the duration of the training simulation.

must be conserved, the relaxation of the system is free from random annihilation of skyrmions. As shown in Fig. 3.10, our ML-based LLG simulations successfully restore and stabilize a nearly perfect SkL over an extended simulation period.

We further investigate the scalability in the temporal domain by extending our ML-based LLG simulation well beyond the training duration. Fig. 3.11 shows a nearly constant structure factor long after the training duration. While a significant reduction in the structure factor was observed between $t = 15,000$ and $t = 23,000$, it quickly recovered to its original stable value ($\mathcal{S}(\mathbf{q}, t) \approx 305$). These temporal fluctuations can be attributed to the prediction errors of the ML model, which, as previously discussed, play a role akin to stochastic noise in Langevin-type dynamics simulations.

Our results demonstrate the robustness of the SkL against small random perturbations. Notably, this benchmark underscores the scalability of our ML models not only in spatial domains (larger lattices) but also across temporal scales (significantly longer simulation times).

To incorporate the underlying symmetries of a physical system into an ML model, appropriate biases (prior knowledge) need to be introduced during the statistical learning process. Two primary approaches to achieve this are: (i) *data augmentation* based on the symmetry group of the system, and (ii) constructing *symmetry-invariant* descriptors or *equivariant* neural network architectures with respect to the symmetry group. These approaches correspond to introducing *observational* and *inductive* biases, respectively, in the context of physics-informed ML (see, e.g., [147–152]). As discussed in Section 3.3.1, the *local* symmetries of our system, namely spin-space and real-space lattice symmetries (also known as internal (gauge) and space-time symmetries [153]), consist of G-valued fields over the underlying lattice, where $G = \mathrm{SO}(3) \times \mathrm{D}_6$. In this work, we adopted the data augmentation approach to enforce the symmetry constraints for the reasons explained below.

First, we briefly summarize the theoretical justification for how data augmentation during the training phase introduces the aforementioned symmetries into the supervised learning process (see [149,154,155] for details). For simplicity, let $(\mathbf{S}, \mathbf{T}) = (\{\mathbf{S}_i\}, \{\mathbf{T}_i\})$ denote a pair of spin configurations and their corresponding torque field, collectively represented as $\mathbf{F}$. Our training data $\mathbf{F}_1, \cdots, \mathbf{F}_n$ consist of independent

identically-distributed samples from a probability distribution $\mathbb{P}$ over the space of all spin-torque fields. It is fundamentally important that the probability distribution $\mathbb{P}$ remains *invariant* under the *action* of each local symmetry $g \in \mathcal{G}$, where $\mathcal{G}$ denotes the space of all local symmetries of the system[1]. Thus, the data augmentation process can be viewed as enriching our set of samples from the probability distribution $\mathbb{P}$ by adding transformed spin-torque fields $g \cdot \mathbf{F}$, $g \in \mathcal{G}$.

During training, at each step t, a minibatch B_t of spin-torque samples $(\mathbf{S}, \mathbf{T})$ of size $|B_t|$ is selected, and a random local symmetry $g_{t,b} \in \mathcal{G}$ is applied to each spin $\mathbf{S}_b$ and torque $\mathbf{T}_b$ field, $b \in B_t$. Then, according to the stochastic gradient descent (SGD) algorithm, the parameters θ of the CNN model f_θ are updated as

$$\theta_{t+1} = \theta_t - \frac{\eta}{|B_t|} \sum_{b \in B_t} \nabla_\theta L \left(f_\theta(g_{t,b} \cdot \mathbf{S}_b) \, , \, (g_{t,b} \cdot \hat{\mathbf{T}}_b) \right) \, , \tag{3.22}$$

where L denotes the loss function given by Eq. (3.11), and η is the learning rate. In other words, the augmented SGD can be interpreted as the minimization of the *empirical risk* associated with the following *augmented loss function*:

$$\int_{\mathcal{G}} L \left(f_\theta(g \cdot \mathbf{S}) \, , \, (g \cdot \hat{\mathbf{T}}) \right) \, d\mathbb{Q}(g) \, , \tag{3.23}$$

where the average is taken along the entire orbit of the group action with respect to a probability distribution $\mathbb{Q}$ over $\mathcal{G}$. It has been proven that data augmentation based on the underlying symmetry group reduces the variance of general estimators

[1] The action of a local symmetry $g \in \mathcal{G}$ on a spin $\mathbf{S}$ and a torque $\mathbf{T}$ field is the induced transformation by g. We denote it by $g \cdot \mathbf{F} := (g \cdot \mathbf{S} \, , \, g \cdot \mathbf{T})$.

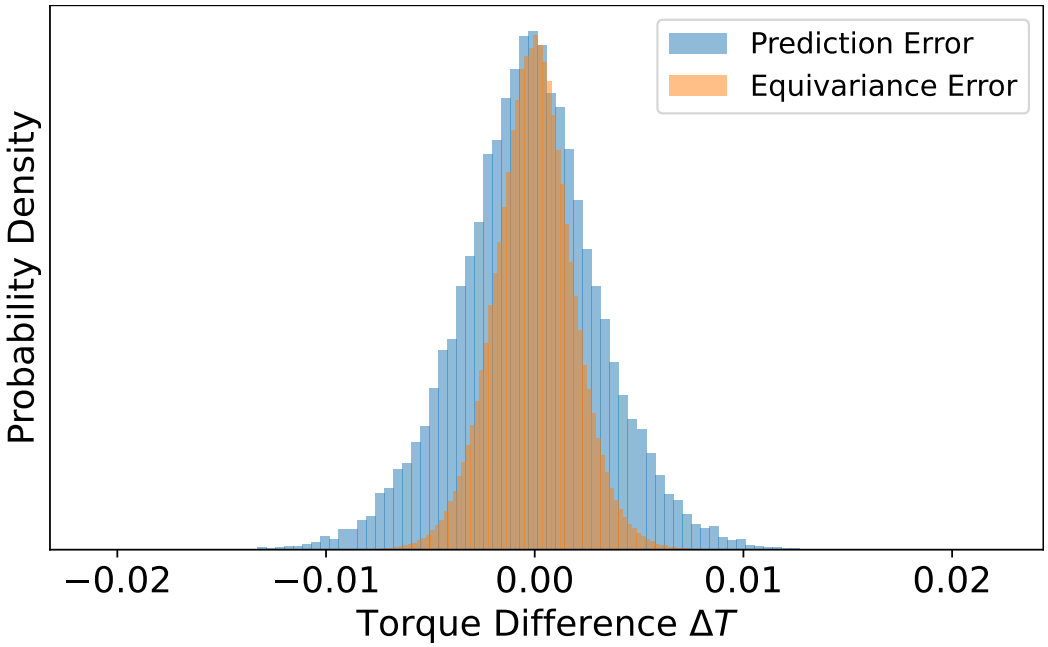

Figure 3.12: Distribution of the equivariance error $err_{eq} := f_{\text{CNN}}(\{R\mathbf{S}_i\}) - Rf_{\text{CNN}}(\{\mathbf{S}_i\})$, where R is an arbitrary rotation. The overall prediction error ("predicted torque" minus "ground truth torque") on the test dataset is superimposed for comparison. The equivariance error is significantly smaller than the overall prediction error, indicating that the model effectively preserves the underlying symmetry of the physical system.

and improves their generalizability [149].

The theoretical justification above can also be validated empirically. Fig. 3.12 shows the typical prediction error (blue), representing the difference between predicted and ground truth torques, and the equivariance error (orange), defined as $f_\theta(g \cdot \mathbf{S}) - g \cdot f_\theta(\mathbf{S})$. As observed in the figure, the equivariance error is smaller than the typical prediction error, indicating that the data augmentation employed effectively preserves the underlying symmetry of the physical system.

To further validate the locality principle, we analyze the receptive field of our CNN model in this section. As discussed in Section 3.3.1, the receptive field of a convolution

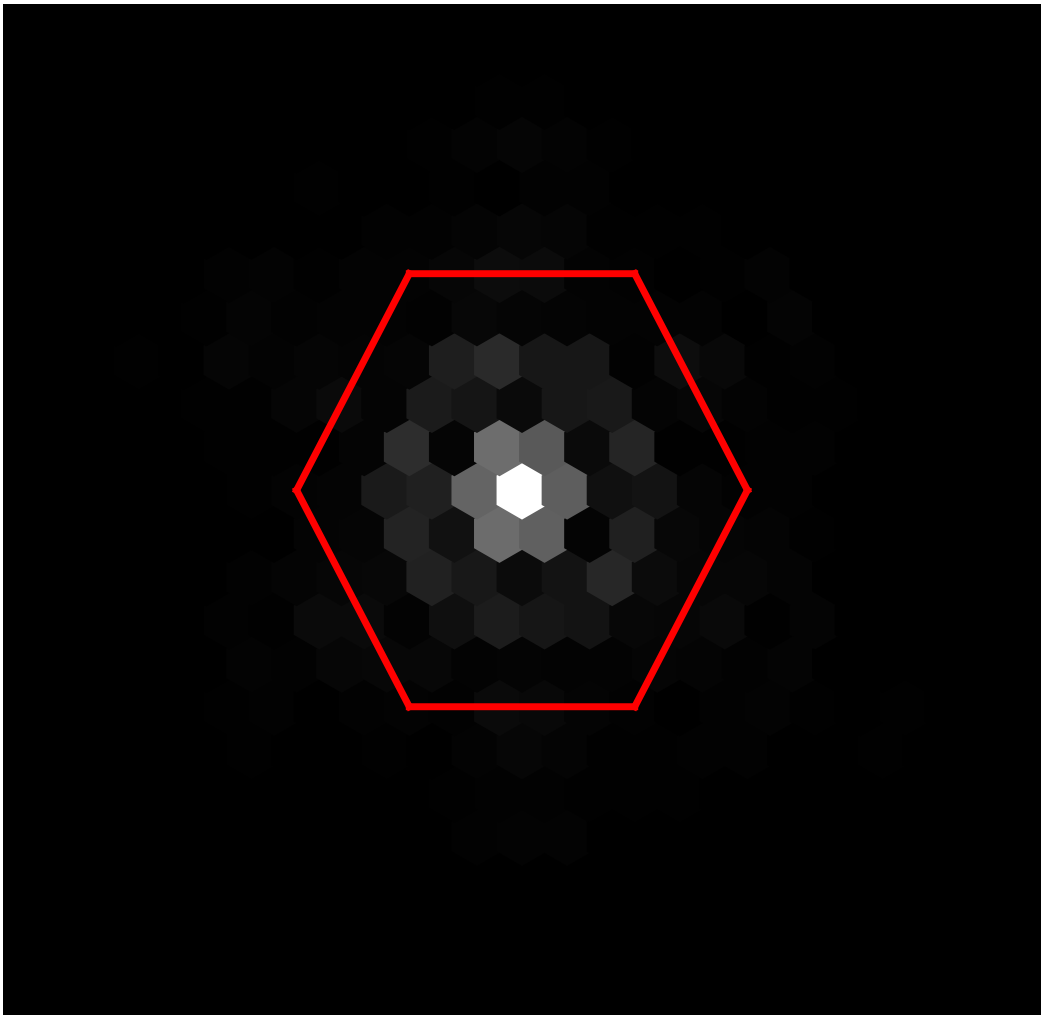

Figure 3.13: ERFs of the ML model. Since both the input and output tensors at each lattice site have three components, there are a total of nine ERFs corresponding to the nine partial derivatives of the three output torque components with respect to the three input spin components. The sum of the absolute values of these derivatives is presented in this figure, with darker pixels indicating smaller derivative values. The red line roughly traces the nonzero value regions of the ERF.

layer f_m is defined as the support $\mathrm{supp}(h_m)$ of the corresponding convolution kernel h_m, i.e., the region where the function values of h_m are nonzero. The receptive field of the entire CNN model is computed as the Minkowski sum of the receptive fields of individual convolution layers, or $\mathrm{RF} = \mathrm{supp}(h_1) \oplus \cdots \oplus \mathrm{supp}(h_L)$. For our model, consisting of 10 layers in depth, each comprising 5×5 convolution kernels with stride 1, the theoretical receptive field size is calculated to be 41. This implies that, in principle, the spin directions within a 41-site neighborhood can influence the torque prediction at lattice site i.

However, the naive computation of the receptive field size may be misleading, as the kernel size of a convolution layer indicates only the theoretical maximum of the receptive field. The actual region of nonzero values may be significantly smaller than the theoretical receptive field size. To address this, we used the approach of [156] to compute the extiteffective receptive field size, within which the function values are practically nonzero. Fig. 3.13 presents the result of this calculation for the trained CNN model. The red hexagonal line delineates the region within which function values are practically nonzero, while values outside this region are essentially zero. The grayscale values inside the hexagon indicate different levels of influence of neighboring spins on the computed torque vector. At lattice site i, located at the center of the red hexagon, the weighting factor is largest, indicating that $\mathbf{T}_i$ is predominantly determined by $\mathbf{S}_i$. The immediate neighbors, or 1-neighborhood $\mathcal{N}_1(i)$, also exhibit high intensity values, implying that the relative configuration of the spin direction $\mathbf{S}_i$

with respect to its neighboring spins $\mathbf{S}_j$ at $j \in \mathcal{N}_1(i)$ has a significant influence on the predicted torque $\mathbf{T}_i$. Similarly, the spin directions in the 3-neighborhood have a substantial influence, with smaller influences detectable up to the 6-neighborhood.

These results are consistent with previous ML spin-torque models based on symmetry-invariant descriptors [98,99], which demonstrate that the spin dynamics of similar s-d models can be accurately captured by bond-product BP-type models employing fully connected NNs with inputs from neighborhoods up to $r_c \sim 5$ lattice constants. Physically, as discussed earlier, the finite size of the ERF is due to the local nature of spin torques. However, the range of locality can only be indirectly inferred from exact calculations. In practice, the cutoff radius is treated as an ad hoc parameter in BP-type ML models or determined through trial and error. It is noteworthy that the CNN model provides a systematic and rigorous method for determining this crucial physical characteristic of electronic models.

3.4.2 Dynamics for Neural Network Model with Descriptor

After concluding the preliminary phases of experimentation as described in Sec. 3.3.3, our investigation progressed to examining the skyrmion phase to further assess the NN's performance with descriptors in analyzing ML-LLG dynamics. Initially, we compared the structure factor along the high-symmetry path $\Gamma(0,0) \to K(\frac{4\pi}{3},0) \to M(2\pi,0)$ by executing 50 independent simulations employing both the ML-LLG and KPM-LLG methodologies. These simulations, initiated from random spin configurations, were conducted with LLG parameters set to $dt = 0.1$ and $\alpha = 0.2$ for both

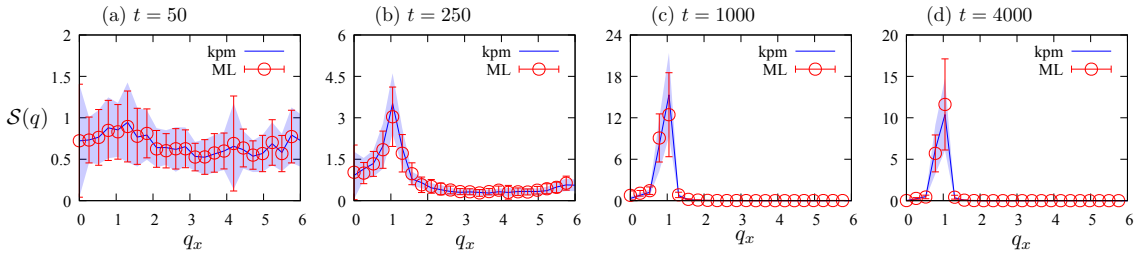

Figure 3.14: Comparison of the structure factor $\mathcal{S}(q)$ along the high-symmetry path $\Gamma(0,0) \to K(\frac{4\pi}{3},0) \to M(2\pi,0)$ obtained from KPM-LLG and ML-LLG simulations at different time steps from $t = 50$ to 4000. The blue lines represent the average value of $\mathcal{S}(q)$ from 50 individual KPM-LLG simulations on a 48×48 lattice, with the blue regions indicating the standard deviation. The red dots with error bars depict the average $\mathcal{S}(q)$ and its standard deviation from ML-LLG simulations.

approaches. The findings, illustrated in Fig. 3.14, reveal that both approaches manifest similar dynamics in the structure factor peaks of the skyrmion phase at the commensurate wave number $Q = (\frac{\pi}{3}, 0)$, with comparable average values and standard deviations.

The temporal evolution of the magnetic phases in the system is depicted in Fig. 3.14. In the initial stage (Fig. 3.14(a)), two peaks, designated as $Q(\frac{\pi}{3},0)$ and $M(2\pi,0)$, grow, while significant fluctuations are observed at $K(\frac{4\pi}{3},0)$. This behavior indicates the formation of antiferromagnetic stripes and the emergence of a vortex phase. As the growth of the stripes stabilizes, the vortex phase becomes more dominant, as shown in Fig. 3.14(b). Subsequently, the skyrmion phase emerges from the vortex phase, as illustrated in Figs. 3.14(c) and (d). This progression of magnetic phases highlights the dynamic evolution of the system and the complex interplay between different magnetic phases.

Given the favorable outcomes of the previous ML-LLG simulations, we applied

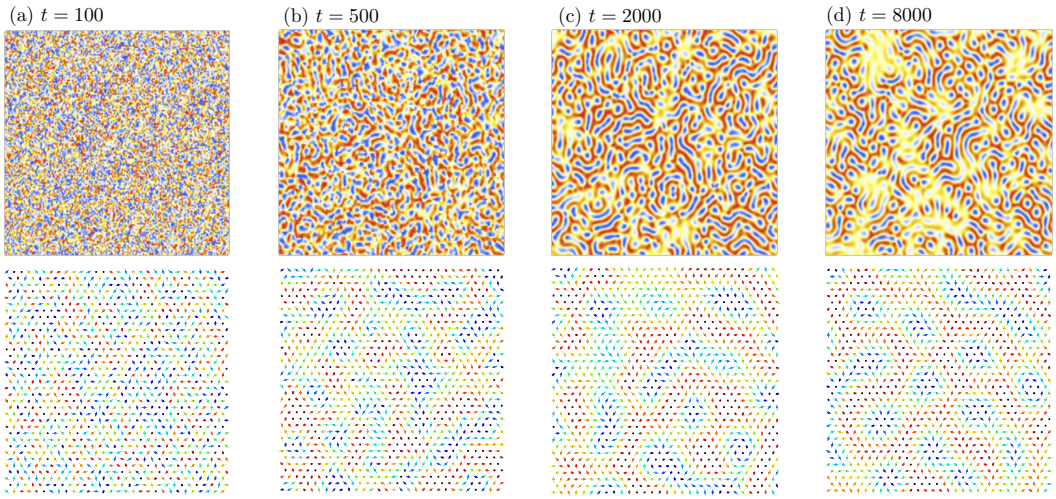

Figure 3.15: Top panels: Density plots of the spin configuration $\{\mathbf{S}_i^z\}$ on a 150×150 lattice at different time steps from $t = 50$ to 4000. Bottom panels: Vector arrows representing the in-plane spin configuration $\{\mathbf{S}_i^{xy}\}$ at the top-right corner of the respective upper panels, with $\{\mathbf{S}_i^z\}$ represented by the arrow colors.

this method to a larger triangular lattice of size $N = 150 \times 150$, as the approach has demonstrated scalability to arbitrary lattice sizes. Fig. 3.15 illustrates the spin dynamics at different time steps. The top panels show density plots of the z-component of the spins $\{\mathbf{S}_i^z\}$, while the bottom panels display vector plots for the in-plane spin components $\{\mathbf{S}_i^{xy}\}$ in the top-right corner of the upper panels. The dynamics observed in Fig. 3.15 reveal three distinct stages in the evolution of the spin system. Initially, the random spins form multiple stripes at various angles. In the intermediate stage, these stripes, with opposing ferromagnetic order in the z-direction, begin to form skyrmions. Finally, the formation of stripes and skyrmions stabilizes, and the growth of skyrmion clusters is arrested.

The corresponding structure factor of the system is shown in Fig. 3.16(a), obtained

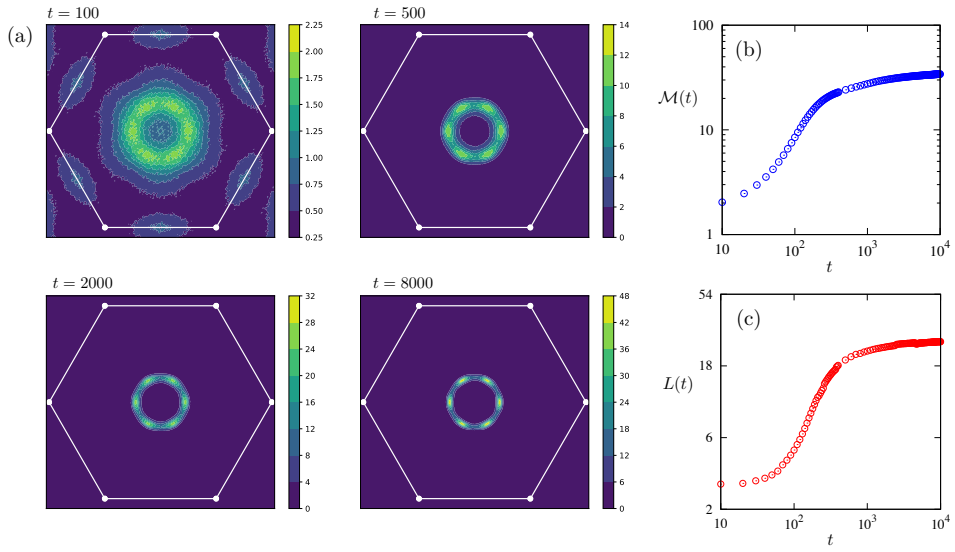

Figure 3.16: Left four panels (a): Structure factor density plots of the ML-LLG simulations on a 150×150 lattice in the first BZ from $t = 50$ to 4000. (b) Skyrmion structure factor peak order $\mathcal{M}$ as a function of time from $t = 10$ to 10^4. (c) Characteristic length of the ML-LLG simulations from $t = 10$ to 10^4.

by averaging over 50 independent simulations. These results are consistent with the conclusions previously discussed. Additionally, to quantify the growth of the structure factor peaks, we defined the skyrmion structure factor peak order and the characteristic length scale of the skyrmions as $\mathcal{M}(t) = \langle S(\mathbf{k}_{pr}, t)\rangle_{\mathbf{k}_{pr}}$ and $L(t) = \frac{\sum_{\mathbf{k}} S(\mathbf{k},t)|\mathbf{k}-\mathbf{k}_p|}{\sum_{\mathbf{k}} S(\mathbf{k},t)}$, where $\mathbf{k}_p$ is the structure factor peak, and $\langle \cdots \rangle$ represents averaging over $\mathbf{k}_{pr}$ such that $|\mathbf{k}_{pr} - \mathbf{k}_p| \leq 0.1$. The evolution of $\mathcal{M}(t)$ and $L(t)$ with time is illustrated in Figs. 3.16(b) and (c). Both quantities exhibit similar trends: slow growth during the initial stage, rapid growth during the intermediate stage with $\mathcal{M}(t), L(t) \propto t$, and eventual saturation during the final stage.

3.5 Conclusion and Outlook

In this chapter, we introduced a CNN model alongside a NN model equipped with descriptors to predict spin torques from input spin configurations in large-scale LLG dynamics simulations of itinerant magnets. Our CNN is purely convolution-based, avoiding fully connected (dense) layers, offering an inherent benefit in terms of localization. Each layer of the CNN applies a convolution using a kernel or filter, akin to a Green's function that portrays the finite response to a nearby source. Thanks to its finite number of trainable parameters, the CNN model is suitable for running dynamical simulations on larger systems without needing to retrain or rebuild the neural network. We demonstrated the practicality of our ML models with a triangular-lattice s-d model exhibiting a skyrmion crystal as its ground state. By using the ML-predicted torques in the LLG dynamics simulations, we validated that the trained ML model can effectively replicate the relaxation dynamics of the skyrmion phase in itinerant spin models. Additionally, we showed the scalability and transferability of our method by proving that large-scale LLG simulations using our CNN model can stabilize a disturbed SkL and sustain its stability over long durations. Furthermore, large-scale LLG simulations using the NN model with descriptors uncovered an atypical phase separation of skyrmions.

Unlike ML force-field models that use the BP scheme, our CNN model predicts torques directly, which are analogous to atomic forces in spin systems. BP-type models, employing methods such as Gaussian process regression or fully connected neu-

ral networks, focus on local energy prediction, which does not allow straightforward comparison with exact calculations. In these models, forces are obtained as gradients of the total energy, composed of all local energies. The concept of local energy leverages locality properties and aids in integrating symmetry into ML frameworks. However, deriving forces from energy gradients confines BP-type models to representing only conservative forces and quasi-equilibrium electron systems. In contrast, our CNN model can represent both conservative and non-conservative spin torques. This feature is especially beneficial for simulating out-of-equilibrium systems driven by non-conservative electron-mediated torques, such as spin transfer torques, which are vital in spintronics applications.

In future studies, we plan to investigate approaches for more robustly enforcing constraints derived from symmetry or conservation laws. Previous work on equivariant CNNs within computer vision [157] may provide valuable directions for constraining CNN layers to maintain SO(3) and D_6 symmetries. Furthermore, our current research limits itself to estimating torques based on spin directions at each time step, without directly addressing the LLG equation as presented in Eq. 3.1. Nonetheless, recent developments have focused on resolving the governing partial differential equations (PDEs) of physical systems through the use of physics-informed deep NN, as seen in Nguyen *et al.* [150]. Utilizing these physics-informed CNN techniques, we expect to achieve quicker and more precise estimates of spin dynamics, highlighting a promising path for upcoming research.

Chapter 4

Machine Learning for Metallic Spin Glass

4.1 Introduction

Spin glasses are disordered magnetic systems characterized by frustrated interactions between localized magnetic moments due to quenched disorder [158–161]. As a result of these frustrated interactions, conventional long-range magnetic order, such as ferromagnetic or N'eel order, cannot be established. Nevertheless, spin-glass systems exhibit a collective freezing transition below a characteristic temperature T_f, indicating the emergence of new magnetic states at low temperatures. The term "glass" draws an analogy between the disordered spins in a spin-glass phase at low temperatures and the atomic positional disorder found in structural glasses [162, 163]. Unlike structural glasses, where atomic configurations freeze due to the exponentially slow relaxation of a supercooled liquid, spin-glass behavior arises from quenched randomness, which plays a central role in their physics.

Several theoretical models have been proposed to understand the nature of spin-

glass phases and transitions. A canonical example is the Edwards-Anderson model [164], which describes random exchange interactions between nearest-neighbor Ising spins σ_i on a lattice: $\mathcal{H} = -\sum_{\langle ij \rangle} J_{ij}\sigma_i\sigma_j$. In this model, the coupling coefficients are random variables, with equal probability of being positive (ferromagnetic) or negative (antiferromagnetic). Mean-field theories of spin glasses, such as the Sherrington-Kirkpatrick model [165], have further elucidated the nature of spin-glass order, leading to the concept of replica symmetry breaking and related order parameters for spin-glass phases [166–169]. The physical insights and mathematical techniques developed in the study of spin glasses have found applications in diverse fields such as computer science, biology, and economics [170–172].

Experimentally, prototype spin-glass materials include dilute magnetic alloys, where magnetic impurities such as Fe or Mn are randomly substituted into the lattice of a nonmagnetic metallic host (e.g., Ag, Cu, Pt) [159–161, 173, 174]. These materials are prepared by rapidly cooling the liquid alloy, thereby fixing the magnetic impurities at random positions within the solid. Since the magnetic impurities are typically several lattice constants apart, their effective interactions are mediated by conducting electrons from the metallic host. Due to the itinerant nature of the conducting electrons, the resultant effective spin-spin interactions are long-ranged, exemplified by the well-known RKKY interactions [136–138]. The RKKY interaction also oscillates between ferromagnetic and antiferromagnetic couplings as the distance increases, providing a mechanism for frustration that stabilizes spin-glass phases.

The RKKY type interactions, however, are obtained based on the assumption of a spherical Fermi surface in the limit of weak electron-spin coupling. In general, the electron-mediated interactions in metallic spin glasses depend on the electronic structure of the underlying metallic system. A full dynamical simulation of such disordered itinerant electron magnets, however, is computationally a very challenging task. As the driving forces come from electrons, integration of the LLG equation for spin dynamics requires solving an electron Hamiltonian at every time-step. Such repeated electronic structure calculations could be prohibitively time-consuming for dynamical simulations of large systems. The computational complexity is similar to the quantum MD simulations, where atomic forces are computed from electronic structure solutions [87].

In this chapter, we present a scalable ML framework for large-scale dynamical modeling of metallic spin glass and disordered itinerant electron magnets. Central to our approach is a deep-learning NN, trained by exact solutions from small systems, that can accurately and efficiently predict the effective local magnetic fields acting on spins. This approach is similar in spirit to ML-based force field models which have revolutionized the field of *ab initio* MD simulations [91–94, 96, 97, 125, 127–129, 175–178]. By accurately emulating the calculation of Kohn-Sham equations, ML interatomic potentials offer the efficiency of classical force-field models with the desired quantum accuracy. Similar ML frameworks have recently been developed to enable large-scale dynamical simulations in several condensed-matter lattice systems [58, 59, 98, 99, 132,

179, 180].

Physically, the linear scalability of ML force-field models is based on the principle of locality, or nearsightedness, of electronic matter [100, 101]. A practical approach to incorporate the locality principle into ML models for quantum MD was proposed in the pioneering work of Behler and Parrinello [91]. The central idea of BP type ML structures involves introducing a local energy ϵ_i [91, 92] which, in our case, is associated with individual spins $\mathbf{S}_i$ in a disordered magnet. The local effective field is obtained from the derivative of the total energy. Importantly, invoking the locality principle, the local energy is assumed to depend on the immediate neighborhood of $\mathbf{S}_i$ and a deep-learning NN is trained to capture this complex dependence.

A crucial component of the scalable ML model is the appropriate representation of the local environment which preserves the symmetry of the original quantum systems. Indeed, atomic descriptors play a central role in the field of ML-based quantum MD methods [91, 92, 96, 104, 181–185]. In our case, a proper representation of spin configurations in a local neighborhood needs to be invariant with respect to real-space rotation and translation symmetries, as well as the SU(2) rotation symmetry in the spin-space. To this end, we develop a magnetic descriptor which is a generalization of a widely used atomic descriptor called atom-centered symmetry function (ACSF) for ML-MD simulations [91, 181].

We apply our ML force-field model to study the relaxation dynamics of a metallic spin-glass model with Heisenberg spins. We note that most works on spin-glass phases

are based on lattice models of either Ising or Heisenberg spins with quenched random nearest-neighbor exchange interaction J_{ij} [159]. There are a few remarkable works on the *ab initio* modeling and dynamical simulations of the prototypical MnCu metallic spin-glass alloys [186–189]. In these approaches, however, a classical Heisenberg model is used to describe the magnetic alloys with an RKKY-like exchange interaction determined from first-principles density functional theory calculations [190]. To demonstrate the ML model for electron-driven spin dynamics, here we consider an amorphous generalization of the well-studied s-d model as a model metallic spin glass. ED combined with LLG dynamics are used to simulate this random s-d model and to generate datasets for training of the ML models.

4.2 Adiabatic Dynamics of Itinerant Electron Magnets

As discussed in several previous works, magnetic moments in dilute magnetic alloys such as CuMn can be well described by Heisenberg spins. We consider a metallic spin system with randomly distributed Heisenberg spins, based on the s-d type electron-spin coupling [191]:

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_e\left(\hat{c}, \hat{c}^\dagger\right) - J\sum_i \sum_{\alpha,\beta=\uparrow,\downarrow} \mathbf{S}_i \cdot \left(\hat{c}^\dagger_{i,\alpha}\boldsymbol{\sigma}_{\alpha\beta}\hat{c}_{i,\beta}\right). \tag{4.1}$$

Here $\mathbf{S}_i$ represents a local classical spin at a random position $\mathbf{r}_i$, and $c^\dagger_{i,\alpha}$ ($c_{i,\alpha}$) denotes the creation (annihilation) operator of an electron with spin $\alpha =\uparrow,\downarrow$ localized at $\mathbf{r}_i$. The first term $\hat{\mathcal{H}}_e$ is the Hamiltonian of the electron subsystem, which cor-

responds to the metallic host in dilute magnetic alloys. The electron operators in the parentheses of the second term correspond to the spin operator of the electrons: $\hat{\mathbf{s}}_i = \frac{\hbar}{2}\left(\hat{c}^\dagger_{i,\alpha}\boldsymbol{\sigma}_{\alpha\beta}\hat{c}_{i,\beta}\right)$. Physically, the coefficient J describes a Hund's coupling between electron spin and local moment.

As discussed in Sec. 4.1, in the limit of small electron-spin coupling, one can integrate out the electrons to obtain an effective interaction between local spins. In particular, assuming an electronic system described by a Fermi liquid $\hat{\mathcal{H}}_e = \sum_{|\mathbf{k}|<k_F}\sum_\alpha \varepsilon_{\mathbf{k}}\hat{c}^\dagger_{\mathbf{k},\alpha}c_{\mathbf{k},\alpha}$, with a parabolic dispersion $\varepsilon_{\mathbf{k}} = \hbar^2|\mathbf{k}|^2/2m_e$, one obtains the well-known RKKY interaction [136–138]

$$E = \sum_{ij} J_{\mathrm{RKKY}}(2k_F|\mathbf{r}_i - \mathbf{r}_j|)\mathbf{S}_i \cdot \mathbf{S}_j, \tag{4.2}$$

where $J_{\mathrm{RKKY}}(x) = 9\pi(J^2/\varepsilon_F)(x\cos x - \sin x)/x^4$, k_F and ε_F are the Fermi wavevector and Fermi energy, respectively. For large coupling $J \gtrsim \varepsilon_F$ and more complex electronic models, the effective interaction in general needs to be computed numerically. For example, such effective interactions have been computed using DFT for realistic CuMn alloys [186–188].

The magnetization dynamics for Heisenberg magnets is governed by the LLG equation

$$\frac{d\mathbf{S}_i}{dt} = \gamma\mathbf{S}_i \times (\mathbf{H}_i + \boldsymbol{\eta}_i) - \alpha\mathbf{S}_i \times (\mathbf{S}_i \times \mathbf{H}_i), \tag{4.3}$$

where γ is the gyromagnetic ratio, α is the damping constant, $\mathbf{H}_i$ is the local effective field that drives the spin dynamics, and $\boldsymbol{\eta}_i$ represents a stochastic magnetic field due to

thermal fluctuations. The random fields at different sites and times are uncorrelated. The three independent components of $\boldsymbol{\eta}_i$ are modeled by a Gaussian random variable of zero mean and a variance proportional to $\alpha k_B T$, where T is temperature. For effective spin Hamiltonian, such as the RKKY interaction in Eq. (4.2), the effective local field is given by

$$\mathbf{H}_i = -\frac{\partial E}{\partial \mathbf{S}_i} = -\sum_j J_{\mathrm{RKKY}}(2k_F|\mathbf{r}_i - \mathbf{r}_j|)\mathbf{S}_j$$

Although this interaction is long-ranged, the $1/r^3$ decay is sufficiently fast that a cutoff radius is introduced in practical calculations. The effective field in general can be efficiently computed for such RKKY type interactions.

It is worth noting that spin dynamics based on such effective interactions is a special case of the adiabatic approximation, which is similar to the Born-Oppenheimer approximation for quantum MD simulations [87]. Within the adiabatic approximation, electrons are assumed to quickly relax to quasi-equilibrium with respect to the instantaneous configuration of classical spins. In this limit, the effective energy is given by $E = \langle\hat{\mathcal{H}}\rangle = \mathrm{Tr}(\hat{\rho}_e\hat{\mathcal{H}})$, where $\hat{\rho}_e$ is the electron density operator:

$$\hat{\rho}_e = \exp(-\hat{\mathcal{H}}/k_B T)/\mathcal{Z}, \tag{4.4}$$

and $\mathcal{Z} = \mathrm{Tr}\exp(-\hat{\mathcal{H}}/k_B T)$ is the partition function. The local magnetic field for Hamiltonian (4.1) in this adiabatic limit is given by

$$\mathbf{H}_i = -\frac{\partial\langle\hat{\mathcal{H}}\rangle}{\partial \mathbf{S}_i} = J\sum_{\alpha\beta}\boldsymbol{\sigma}_{\alpha\beta}\langle\hat{c}^\dagger_{i,\alpha}\hat{c}_{i,\beta}\rangle. \tag{4.5}$$

Here we have used the Hellmann-Feynman theorem $\partial\langle\hat{\mathcal{H}}\rangle/\partial\mathbf{S}_i = \langle\partial\hat{\mathcal{H}}/\partial\mathbf{S}_i\rangle$ to relate the local field to the electron correlation function $C_{i\alpha,j\beta} \equiv \langle\hat{c}^\dagger_{j,\beta}\hat{c}_{i,\alpha}\rangle$. The calculation of the correlation function, however, requires solution of the equilibrium electron density matrix $\hat{\rho}_e$, which amounts to solving a disordered electron Hamiltonian for a given spin configuration. In the absence of electron-electron interactions, the standard method for solving the electronic structure is based on ED. However, since the electronic forces have to be computed at every time-step of the LLG dynamics simulation, the $\mathcal{O}(N^3)$ time complexity of ED can be overwhelmingly time-consuming for large systems.

4.3 Machine Learning Force-Field Model For Disordered Spins

In this section we present a scalable ML framework to essentially derive an effective spin Hamiltonian $E(\mathbf{S}_i)$ for metallic spin glass models in Eq. (4.1). It should be noted that this effective energy E, or effective classical spin Hamiltonian, is technically obtained by freezing the spin configuration and integrating out the electrons using the electron density operator in Eq. (4.4). It can be viewed as a more complicated version of the RKKY interaction. The ML methods offer a systematic approach to obtain an accurate and efficient parametrization of this effective spin Hamiltonian.

4.3.1 Behler-Parrinello Machine Learning Framework

Fundamentally, as discussed in Sec. 4.1, linear-scaling electronic structure methods are possible mainly because of the locality nature of many-electron systems [100,101].

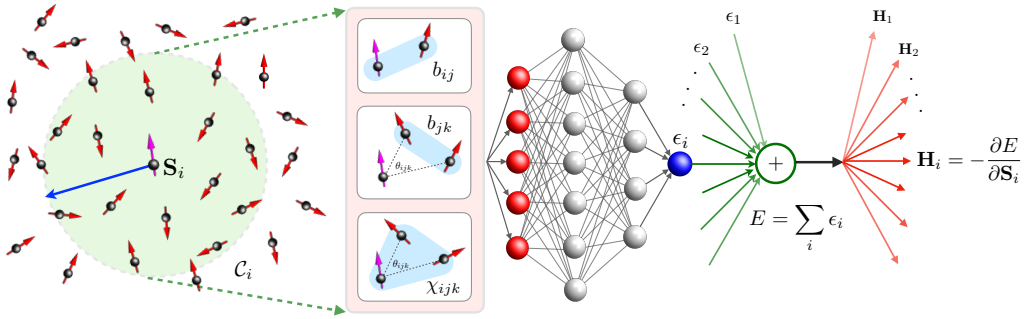

Figure 4.1: Schematic diagram of ML force-field model for metallic spin glass. The input of the ML model is the magnetic configuration $\mathcal{C}_i$ centered at the i-th spin $\mathbf{S}_i$, while the output is the local energy ϵ_i associated with the center spin. The atomic and spin configuration within the neighborhood $\mathcal{C}_i$ is processed using a magnetic version of the ACSF method. The resultant feature variables $\{G_k\}$ are input to a feed-forward fully connected neural network with a single output that is the local energy ϵ_i. The total energy is obtained by applying the ML energy model to all spins in the system. The local effective field is computed using automatic differentiation.

Modern ML techniques provide an explicit and efficient approach to incorporate the locality principle into the implementation of $\mathcal{O}(N)$ methods. In particular, the BP type schemes provide a practical method to incorporate locality and symmetry to the ML models. Indeed, most ML force-field models for quantum MD simulations are based on BP approaches. Here we generalize the BP scheme to implement an ML model for the efficient prediction of system energy E and the local effective fields $\mathbf{H}_i$ for spin dynamics.

A schematic diagram of our ML force-field model is outlined in FIG. 4.1. First, as in the original BP approach, we partition the effective spin energy into local contributions

$$E = \sum_i \epsilon_i = \sum_i \varepsilon(\mathcal{C}_i), \tag{4.6}$$

where ϵ_i is associated with the i-th local spin of the disordered magnet. In the second step of the above equation, we have further assumed that the local energy only depends on the immediate neighborhood, denoted by $\mathcal{C}_i$, of the i-th spin through a universal function $\varepsilon(\cdot)$. The validity of this step naturally relies on the locality principle mentioned above. Practically, the neighborhood $\mathcal{C}_i$ is defined as configuration of spins within a sphere of cutoff radius R_c centered at $\mathbf{S}_i$, i.e. $\mathcal{C}_i = \{\mathbf{S}_j \,|\, |\mathbf{r}_j - \mathbf{r}_i| < R_c\}$; see FIG. 4.1. Importantly, the universal function $\varepsilon(\cdot)$ that relates local energy ϵ_i to the neighborhood is to be approximated by a deep-learning neural network model. As shown in FIG. 4.1, the local magnetic environment is processed to produce a set of symmetry-invariant feature variables, denoted as $\{G_m\}$, which are then fed into the neural network.

Once the total energy E is obtained by applying the same descriptor and neural network model to every spin in the system, the local field acting on spin $\mathbf{S}_i$ can then be efficiently computed through automatic differentiation:

$$\mathbf{H}_i = -\frac{\partial E}{\partial \mathbf{S}_i} = -\sum_j{}' \frac{\partial \epsilon_j}{\partial \mathbf{S}_i}. \tag{4.7}$$

Here the prime in the summation indicates that only local energies ϵ_j within the neighborhood of the i-th spin are considered, again, assuming the locality of the effective field.

Another important insight in the original work of Behler and Parrinello [91] is the introduction of descriptors, or feature variables, that provide an appropriate rep-

resentation of the neighborhood $\mathcal{C}_i$. It is worth noting that, despite the universal approximation power of neural networks, symmetry properties of a function can only be learnt statistically, but not exactly. The goal of a descriptor is to incorporate the symmetry of the original quantum Hamiltonian into the classical effective energy model. Since the output of the ML model in the BP framework is a local energy, which, as a scalar, is invariant under symmetry transformations of the original system, the feature variables $\{G_m\}$ need to be also invariant with respect to the same symmetry group.

In the case of ML-based quantum MD methods, extensive studies have been devoted to the development of atomic descriptors [91, 92, 96, 104, 181–185]. A proper representation of the atomic neighborhood should be invariant under rotational and permutational symmetries, while retaining the faithfulness of the Cartesian representation. A widely used atomic descriptor, which is physically intuitive, is the ACSF representation introduced in the original work of Behler and Parrinello [91]. The ACSFs are built from the relative distances and relative angles among atoms in the neighborhood, which are manifestly invariant under rotations [91, 181]. A more systematic approach to build invariant feature variables is based on the so-called bispectrum coefficients, which are special triple-products of IRs of the symmetry group [92, 103, 104].

The group-theoretical method has also been employed to develop a general theory of descriptors for electronic lattice models in condensed-matter systems [58]. Com-

pared with the MD systems, the SO(3) rotational symmetry of free-space is reduced to discrete point-group symmetries in lattice models. On the other hand, the dynamical degrees of freedom in lattice models, such as local magnetic moments or order-parameters, are characterized by additional internal symmetry group. A proper descriptor for lattice systems thus needs to be invariant with respect to both the internal symmetry group and the lattice point group.

In particular, following the general theory, a magnetic descriptor is developed to incorporate the lattice point-group and SU(2) spin-rotation symmetries into the ML force-field models for lattice models of itinerant electron magnets, such as the s-d or Kondo-lattice models [98, 99, 132]. A two-step approach is used to construct the feature variables. First, to preserve the rotation symmetry in spin-space, bond $b_{jk} = \mathbf{S}_j \cdot \mathbf{S}_k$ and scalar chirality $\chi_{ijk} = \mathbf{S}_i \cdot \mathbf{S}_j \times \mathbf{S}_k$ variables, which are inner product and scalar triple product of spins within a neighborhood, are used as building blocks for the magnetic descriptor. A group-theoretical approach based on reference IR [58], which is a modified bispectrum method, is used to derive feature variables that are also invariant under symmetry operations of the lattice point group.

We note in passing that in order to model atomic systems with finite magnetic moments, atomic descriptors have been generalized in several recent works to include spin degrees of freedom [133, 134, 192–195]. For such magnetic materials, magnetic contributions are crucial for modeling the mechanic phase stability, vibrational properties, and defect dynamics. However, since the goal of these descriptors is to assist

ML-based quantum molecular dynamics, the magnetic structure of atoms are assumed to be fixed throughout the MD simulations. Moreover, often collinear Ising-type spins are considered. Here, on the other hand, we present a magnetic descriptor to be combined with an ML model for spin dynamics, whereas atomic configurations are fixed.

4.3.2 Magnetic Descriptor for Disordered Spins

The symmetry group of the spin glass model described in Eq. (4.1) includes the real-space translation and rotation symmetries of atoms (which carry a local spin $\mathbf{S}_i$) and the global SO(3)/SU(2) rotation symmetry of classical/electron spins. The translation symmetry is readily accounted for in the BP framework, as exactly the same descriptor and neural network are applied to every spin in the system. To account for the rotation symmetry of atoms in real space, we first review the idea of symmetry functions which are built on two fundamental scalars: the relative distance between two atoms $R_{ij} = |\mathbf{r}_j - \mathbf{r}_i|$ and the relative angle between three atoms $\cos\theta_{ijk} = (\mathbf{r}_j - \mathbf{r}_i) \cdot (\mathbf{r}_k - \mathbf{r}_i)/R_{ij}R_{ik}$. As scalars, these two types of variables are manifestly invariant under uniform rotations of atoms in the neighborhood. A two-body symmetry function centered at $\mathbf{r}_i$ is defined as

$$G_2(\Lambda) = \sum_{j \neq i} F_2(R_{ij};\, \Lambda), \tag{4.8}$$

where $\Lambda = \{p_1, p_2, \cdots\}$ denote a set of parameters characterizing G_2, and $F_2(R; \Lambda)$ is a user-defined function, parameterized by Λ to extract atomic structures at certain distances from the center atom. For example, the following function is proposed in

the original work [91] to sample atoms at a distance $d \pm w$ from the center

$$F_2(R; d, w) = e^{-(R-d)^2/w^2} f_c(R). \tag{4.9}$$

Here $f_c(r) = \frac{1}{2}\left[\cos(\frac{\pi r}{R_c}) + 1\right]$ for $R \leq R_c$ and zero otherwise is a soft cutoff function. The two parameters d and w specify the center and width, respectively, of the Gaussian function. The 3-body symmetry functions are defined as

$$G_3(\Lambda) \;=\; \sum_{j,k\neq i} F_3(R_{ij}, R_{ik}, R_{jk}, \theta_{ijk};\, \Lambda), \tag{4.10}$$

An example of the three-body envelop function characterized by three parameters is [91, 127]

$$F_3(R_1, R_2, R_3, \theta; \zeta, \lambda, d, w, d', w') = 2^{1-\zeta}(1 + \lambda\cos\theta)^\zeta$$
$$\times F_2(R_1; d, w) F_2(R_2; d, w) F_2(R_3; d', w'). \tag{4.11}$$

These functions are designed to sample relative orientations between atomic pairs, where the angular resolution is controlled by the parameter ζ. The three F_2 functions are introduced to constrain the distances of the three pairs of spins.

To incorporate the spin degrees of freedom, we first note that assuming the electronic part $\mathcal{H}_e$ in Eq. (4.1) is magnetically isotropic, the metallic spin system is invariant under a global rotation of local magnetic moments $\mathbf{S}_i \to \mathcal{R} \cdot \mathbf{S}_i$, and a simultaneous unitary transformation of the electron spinor $\hat{c}_{i\alpha} \to \hat{U}_{\alpha\beta}\hat{c}_{i\beta}$, where $\mathcal{R}$ is an orthogonal 3×3 matrix and $\hat{U} = \hat{U}(\mathcal{R})$ is the corresponding 2×2 unitary rotation operator. The ML force-field model, which is essentially an effective classical spin

model by integrating out electrons, should preserve the SO(3) spin-rotation symmetry. Similar to magnetic descriptors for lattice models, we use the bond and scalar spin chirality variables $b_{ij} = \mathbf{S}_i \cdot \mathbf{S}_j$ and $\chi_{ijk} = \mathbf{S}_i \cdot \mathbf{S}_j \times \mathbf{S}_k$ as building blocks for characterization of the magnetic environment [98, 99, 132]. These variables are also manifestly invariant under global rotations of all spins.

Based on these building blocks, invariants with respect to rotations in both real space and spin space are obtained by "attaching" these variables into the symmetry functions. A schematic diagram showing the construction of the three types of magnetic symmetry functions is shown in FIG. 4.1. First, we define a magnetic two-body symmetry function

$$G_2^m(\Lambda) = \sum_{j \neq i} F_2(R_{ij};\, \Lambda)\, (\mathbf{S}_i \cdot \mathbf{S}_j). \tag{4.12}$$

Here the superscript m is used to indicate magnetic version of the ACSF. With the F_2 function defined in Eq. (4.9), this magnetic symmetry function accounts for spin-spin correlations between the center $\mathbf{S}_i$ and neighboring spins $\mathbf{S}_j$ at a distance $d \pm w$ from the center. Next, a symmetry function which involves three atoms is defined:

$$G_3^m(\Lambda) = \sum_{jk \neq i} F_3\big(R_{ij}, R_{ik}, R_{jk}, \theta_{ijk};\, \Lambda\big)\, (\mathbf{S}_j \cdot \mathbf{S}_k), \tag{4.13}$$

This symmetry function, however, only involves two spins. Collectively, they describe the two-spin correlations with certain relative angles and distances in the neighborhood of the i-th spin.

The total energy Eq. (4.6) computed from the ML model is now a function explicit

of these magnetic ACSFs, i.e. $E = E(\{G_2^m(\Lambda), G_3^m(\Lambda)\})$. This can also be viewed as an effective classical spin Hamiltonian, which can be formally expanded as

$$E = \sum_{ij} J_{ij}\, \mathbf{S}_i \cdot \mathbf{S}_j + \sum_{ijkl} K_{ijkl} (\mathbf{S}_i \cdot \mathbf{S}_j)(\mathbf{S}_k \cdot \mathbf{S}_l) + \cdots \tag{4.14}$$

For example, the two-body J_{ij} interaction can be obtained through linear combinations of the F_2 and F_3 functions in Eqs. (4.12) and (4.13), respectively:

$$\begin{aligned} J_{ij} = &\sum_{\Lambda} \alpha(\Lambda) F_2(R_{ij}; \Lambda) \\ &+ \sum_{\Lambda'} \sum_{k \neq i,j} \beta(\Lambda') F_3(R_{ki}, R_{kj}, R_{ij}, \theta_{kij}; \Lambda'). \end{aligned} \tag{4.15}$$

Here α and β are coefficients that are determined through training. The first term above, which depends only on the distance R_{ij} between the spin pair, is similar to the RKKY effective interaction discussed in Sec. 4.2. The second F_3 term describes a two-spin interaction which depends not only on the pair distance R_{ij}, but also on the immediate atomic environment of the spin pair. The F_3-type interactions are thus the magnetic analogs of the bond-order potentials, a class of empirical MD potentials that include effects of atomic environment on a chemical bond [196–198]. The higher-order terms in Eq. (4.14) are generated in nonlinear transformations of the neural net. In particular, the four-spin interactions K_{ijkl} are known to play a crucial role in stabilizing non-coplanar spin structures in itinerant electron magnets [199, 200]

Finally, the three-body F_3 functions can also be combined with the scalar spin

chirality χ_{ijk} to form a magnetic symmetry functions centered at the i-th spin

$$G_3^{\chi}(\Lambda) = \sum_{jk\neq i} F_3\big(R_{ij}, R_{ik}, R_{jk}, \theta_{ijk}; \Lambda\big)\,(\mathbf{S}_i \cdot \mathbf{S}_j \times \mathbf{S}_k), \tag{4.16}$$

Since the scalar chirality χ_{ijk} is nonzero only when the three spins are non-coplanar, the above G_3^{χ} variables are also very effective in modeling magnetic structures with non-coplanar spins, similar to the 4-spin terms in Eq. (4.14). However, unlike the bond variables, the scalar chirality is a pseudo-scalar, which changes sign $\chi_{ijk} \to -\chi_{ijk}$ under time-reversal transformation, or simultaneous inversion of three spins. As a result, they are not appropriate for ML modeling of time-reversal symmetric magnetic systems, such as most spin glasses. One could still include the chirality ACSFs in the feature variables and use data augmentation, i.e. by adding time-reversed spin structures to training dataset, to approximate the time-reversal symmetry. This would lead to, e.g. 6-spin terms $\sum L_{ijklmn}\chi_{ijk}\chi_{lmn}$ in the effective spin Hamiltonian. However, our numerical experiments show that inclusion of chiral ACSFs does not lead to significant improvement.

4.4 Relaxation Dynamics of an Amorphous s-d Model

We apply the above ML framework to a random s-d model as an example for metallic spin glasses. The Hamiltonian in Eq. (4.1) describes a general itinerant magnet with s-d type electron-spin coupling. We note that one can also include short-range Heisenberg type interactions $J_{\rm ex}\,\mathbf{S}_i \cdot \mathbf{S}_j$ among localized moments, due to either direct or super-exchange mechanisms, to this Hamiltonian. Since the short-range interaction

can be easily included in LLG dynamics simulations, here we focus on ML modeling for the electron-driven forces. The details of the electron-mediated interactions depend on the electronic Hamiltonian $\hat{\mathcal{H}}_e$. For example, the well-studied s-d model, also known as the Kondo-lattice model for spin-1/2, corresponds to a periodic array of local spins $\mathbf{S}_i$ and a tight-binding Hamiltonian with nearest-neighbor hopping defined on the same lattice [201–203]. The ferromagnetic s-d model in the strong coupling limit, also known as the double-exchange model, plays an important role in the colossal magnetoresistance phenomena [14, 204].

Here we introduce an amorphous generalization of the s-d model. First, instead of placing the spins on a periodic lattice, the local moments are randomly distributed in a three dimensional space. Specifically, the position $\mathbf{r}_i$ of spin $\mathbf{S}_i$ is a random vector uniformly distributed within a 3D cubical box of side L, with the only constraint that the distance between any pair of spins is greater than a minimum, i.e. $R_{ij} = |\mathbf{r}_j - \mathbf{r}_i| > r_{\min}$. For a given random atomic configuration $\{\mathbf{r}_i\}$, a disordered tight-binding model is employed to describe the electronic subsystem $\hat{\mathcal{H}}_e$, giving rise to a disordered s-d Hamiltonian

$$\begin{aligned}\hat{\mathcal{H}} = &\sum_{ij}\sum_{\alpha=\uparrow,\downarrow} t\big(|\mathbf{r}_i - \mathbf{r}_j|\big)\, c^{\dagger}_{i\alpha} c_{j\alpha} \\ &-J\sum_{i}\sum_{\alpha,\beta=\uparrow,\downarrow} \mathbf{S}_i \cdot \left(\hat{c}^{\dagger}_{i,\alpha}\boldsymbol{\sigma}_{\alpha\beta}\hat{c}_{i,\beta}\right).\end{aligned} \tag{4.17}$$

The electron hopping coefficient $t_{ij} = t(R_{ij})$ is a random variable dependent on the distance R_{ij} between an atomic pair through a Yukawa-type exponentially decaying

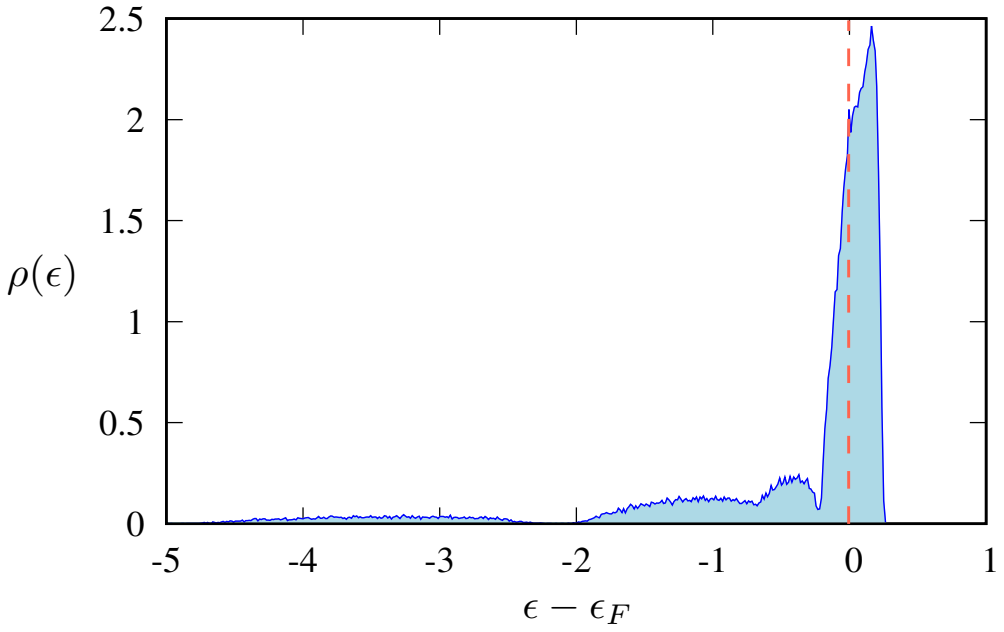

Figure 4.2: Electron density of states of the disordered tight-binding model Eq. (4.17) in the absence of electron-spin coupling $J = 0$. The hopping constant $t_0 = 1$, which serves as the unit of the energy. The number of atoms is $N = 100$ in a cubical box of linear size $L = 5\ell$ with periodic boundary conditions. The electron filling fraction is set at $f = 0.5$. The zero of the energy, indicated by the dashed line, corresponds to the Fermi level.

function

$$t(R) = t_0 \exp(-R/\ell). \tag{4.18}$$

In the following, we set $t_0 = 1$ which also serves as the reference for energy. The characteristic range or length scale of electron hopping is given by the decay length ℓ. It is worth noting that such Yukawa tight-binding models have long been used in the modeling of amorphous systems [205–209].

We first examine the electron density of states (DOS) of the random tight-binding Hamiltonian. To this end, we set the s-d coupling $J = 0$ and use exact diagonaliza-

tion to compute the eigen-energies ϵ_k of the electron hopping Hamiltonian. The DOS, defined as $\rho_e(\epsilon) = \frac{1}{N}\langle\sum_k \delta(\epsilon-\epsilon_k)\rangle$, is obtained by averaging 500 different atomic configurations $\{\mathbf{r}_i\}$. In this calculation, $N = 500$ atoms are randomly distributed within a simulation box of linear size $L = 5\ell$, with a minimum separation $r_{\min} = 0.5\ell$. The resultant DOS is shown in FIG. 4.2. The Fermi energy ϵ_F is determined from the condition of half electron filling. The DOS exhibits a pronounced peak in the vicinity of the Fermi level. The nonzero DOS at Fermi level $\rho(\epsilon_F)$ indicates a gapless electronic state that is susceptible to small perturbations. Moreover, previous large-scale numerical study shows that electron wave functions for the eigenstates in the middle of the band is likely to be delocalized [209]. These extended electron eigenstates near the Fermi level are the dominate contributors to the long-range effective spin-spin interactions.

Our goal is to build a ML model for the disordered s-d system (4.17) with a large electron-spin coupling $J = 6t_0$. We note that the RKKY type perturbation methods cannot be applied to such strong coupling regime of s-d type models. The ML methods thus provide a non-perturbative approach to derive an effective classical spin Hamiltonian and force field that are beyond the analytical methods. To generate the training and testing datasets, ED and ED-LLG simulations are carried out on a system consisting of $N = 100$ atoms with a half-filled electron band in a cubical box of linear size $L = 5\ell$. The minimum separation is again set at $r_{\min} = 0.5\ell$. 10 different realizations of random atomic configurations are used in generating the dataset. For

each realization of the atomic structure, the s-d model is solved using the ED for 250 different spin configurations, including random spins and states obtained from relaxation simulations. It is worth noting that, as the prediction of local field $\mathbf{H}_i$ for each spin counts as single training data point, the size of the effective datasets is $250 \times 10 \times 100 = 2.5 \times 10^5$.

A neural network with 8 hidden layers is constructed using PyTorch [106] to learn the dependence of local energy ϵ_i on the feature variables $\{G_k\}$ that characterize the neighborhood $\mathcal{C}_i$. The number of neurons at the input layer is determined by the number of feature variables and is fixed at 450. The number of neurons in successive hidden layers are: $1024 \times 1024 \times 512 \times 256 \times 128 \times 128 \times 128 \times 128$. The NN performs a series of nonlinear transformations on the input neurons where ReLU [210] is used as the activation function between layers. The NN model is trained based on a loss function including the mean square error (MSE) of both the effective field and total energy. Since only the perpendicular component of the local field $\mathbf{H}_{i,\perp} = \mathbf{S}_i \times \mathbf{H}_i$ contributes to the driving force for spin dynamics, the loss function focuses on the MSE of the perpendicular field. Specifically, for a given spin configuration, the loss function is defined as

$$L = \mu_H \sum_{i=1}^{N} \left| \mathbf{H}_{i,\perp}^{\mathrm{ED}} - \mathbf{H}_{i,\perp}^{\mathrm{ML}} \right|^2 + \mu_E \left| E^{\mathrm{ED}} - E^{\mathrm{ML}} \right|^2, \tag{4.19}$$

where μ_H and μ_E determines the relative weights of the force and energy constraints in the loss function. As shown in Eq. (4.7), the ML local field is obtained from the

derivative of the sum of local energies. This can be efficiently done using automatic differentiation in PyTorch [211]. Trainable parameters of the NN are optimized by the Adam stochastic optimizer [212] with a learning rate of 0.001. A 5-fold cross-validation and early stopping regularization are performed to prevent overfitting.

To incorporate the rotation symmetry, both in real and spin space, to the ML model, a descriptor with all three G_2^m, G_3^m and G_3^χ symmetry functions is developed to characterize local spin configurations $\mathcal{C}_i$. A cutoff radius $R_c = 2.5\ell$ is used for computing these feature variables. First, the two-body symmetry functions $G_2^m(d, w)$ depend on two parameters characterizing a mass shell of thickness w and radius d. The width of the shell is fixed at $w = 0.05\ell$, while 50 different radii d in the range of $[0.5, 2.46]$ (in units of ℓ) are used. For both three-body symmetry functions G_3^m and G_3^χ, there are two parameters ζ, λ for characterizing the angular distribution, similar d, w for a mass shell, and another set d', w' for the distance between two neighboring spins; see Eq. (4.11). Here we set $\zeta = 1$ and consider two $\lambda = \pm 1$ corresponding to an angular function $(1 \pm \cos\theta_{ijk})$ with a peak at a relative angle $\theta_{ijk} = 0$ and π, respectively. For the three F_2 functions in G_3^m and G_3^χ, we use $w' = 1$ and two different $d' = 1$ and 2 in units of ℓ, and the same w, d parameters as in the G_2^m. The total number of feature variables is $50 + 2 \times 200 = 450$.

The benchmark of force prediction for the above ML model with all three G_2^m, G_3^m, and G_3^χ symmetry functions is shown in FIG. 4.3(a) and (b). Excellent agreements between the ML predicted local fields and the exact values are obtained for both train-

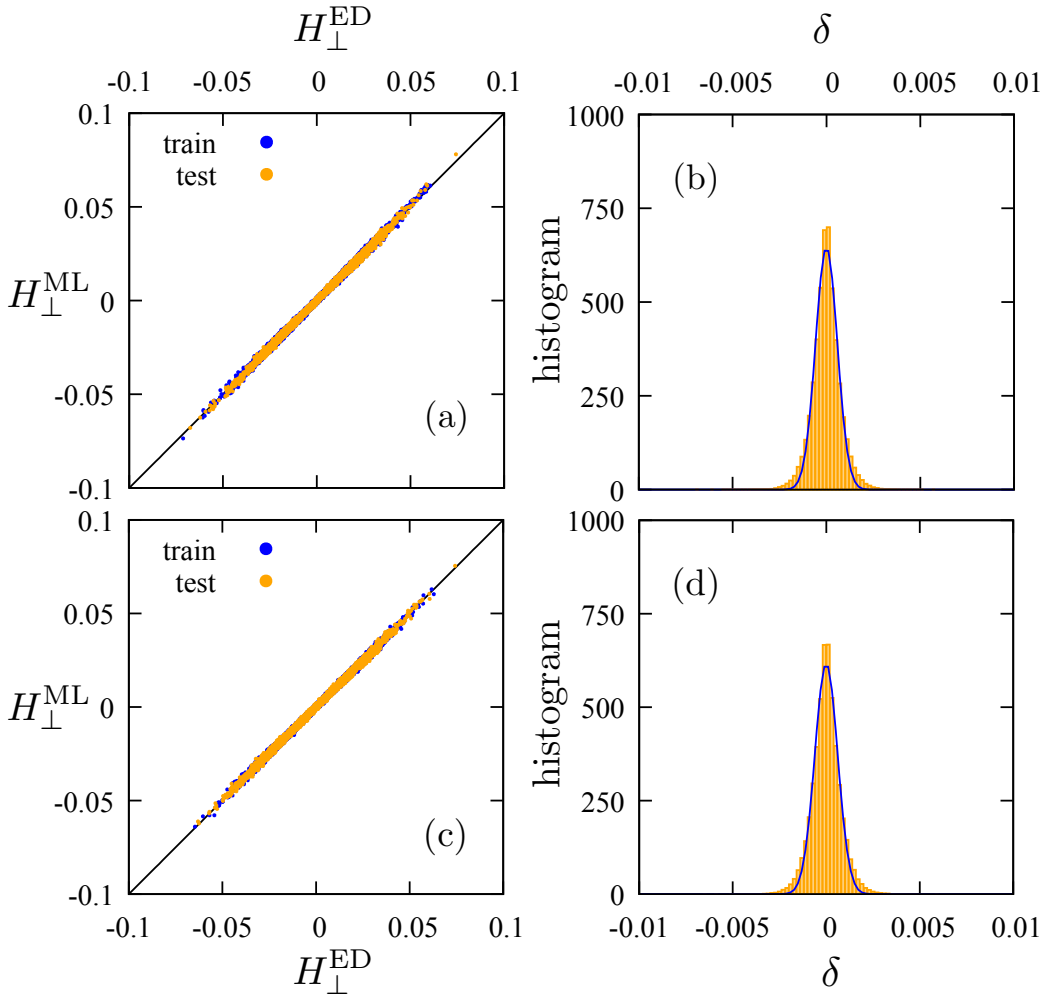

Figure 4.3: Benchmark of ML prediction for the perpendicular local field. Panel (a) shows the predicted components of $\mathbf{H}_\perp$ versus the ground-truth values for the trained NN model based on a magnetic descriptor with all three symmetry functions G_2^m, G_3^m, and G_3^χ. The corresponding histogram of prediction error $\delta = H_\perp^{\rm ML} - H_\perp^{\rm ED}$ is shown in panel (b). The results for ML model with a descriptor that *excludes* the chirality feature variables G_3^χ are shown in panels (c) and (d).

ing and testing datasets. The histogram of the prediction error, shown in FIG. 4.3(b), is characterized by a rather small mean square error $\sigma \approx 0.001$. However, as discussed in Sec. 4.3.2, the chirality feature variables G_3^{χ} change sign under the time-reversal transformation. The resultant effective spin Hamiltonian and the force-field model is not invariant with respect to time-reversal symmetry, which is an intrinsic symmetry of the original s-d model. The time-reversal symmetry can be incorporated into the ML model, albeit inexactly, using data augmentation, i.e. by including both spin configurations $\pm\mathbf{S}_i$ in the dataset. On the other hand, effective interactions involving scalar chirality that is also time-reversal symmetric, $\sum L_{ijklmn}\chi_{ijk}\chi_{lmn}$, is of six-order in spin variables. It is likely that their contribution is negligible compared to that of bond variables.

For comparison, we have also developed a ML model with a magnetic descriptor only involves the two magnetic symmetry functions G_2^m and G_3^m. Since these two feature variables are built from bond variables $b_{ij} = \mathbf{S}_i \cdot \mathbf{S}_j$, hence are invariant under time-reversal, the resultant ML models automatically preserve the time-reversal symmetry. The force benchmark of this ML model *without* chirality variables is summarized in FIG. 4.3(c) and (d). Notably, the ML predicted local fields also agree very well with the ED calculations. This result indicates that the spin chirality variables do not play a major role in the effective spin Hamiltonian. The bond variables alone provide accurate approximations to the amorphous s-d model.

Next we perform dynamical benchmarks of the ML force-field models. To this end,

we integrate the trained ML models with the LLG method and perform thermal-quench simulations of the random s-d model. Specifically, for a given disordered atomic configuration, an initial state of random spins is quenched to a temperature $T = 0.001t_0$ at time $t = 0$. A damping coefficient $\alpha = 0.05\gamma$ in Eq. (4.3), where γ is the gyromagnetic ratio, is used for both the ML- and ED-LLG simulations. The LLG simulations are then repeated for different independent initial random spins and realization of atomic disorder. The ensemble-averaged dynamical evolution of the quenched states is then compared with that obtained from the ED-LLG simulations. In particular, we compute the time-dependent correlation functions

$$C(r_{ij}, t) = \langle \mathbf{S}_i(t) \cdot \mathbf{S}_j(t) \rangle - \langle \mathbf{S}_i(t) \rangle \cdot \langle \mathbf{S}_j(t) \rangle, \tag{4.20}$$

where $\langle \cdots \rangle$ means ensemble average over both atomic and spin configurations. FIG. 4.4 shows the ensemble-averaged spin-spin correlation functions at different times during the quench. The correlations obtained from the ML-LLG simulations agree well with those from the ED methods, showing that the ML force-field model not only accurately predicts the driving forces, but also capture the spin dynamics and relaxation process.

The correlation functions are assumed to depend only on the distance r_{ij} between of a pair of spins, thanks to the rotation and translation symmetry of the disordered states. The ensemble-averaged magnetization is found to be nearly zero, $\mathbf{m} = \langle \mathbf{S}_i \rangle \approx \mathbf{0}$, in both ML and ED-LLG simulations. Moreover, we find that the relaxation

dynamics slows down significantly at large times after the quench. The freezing of spin dynamics thus indicates that the system settles into a local minimum of the effective spin Hamiltonian.

Due to the minimum distance constraint, there are no spin pairs with distance less than $r_{\min} = 0.5\ell$. On the other hand, early in the relaxation, e.g. at $t \lesssim 30$ in units of $(\gamma t_0)^{-1}$, an anti-parallel spin-spin correlation $C(r) \approx -0.4$ quickly develops at the minimum separation $r_{\min}$. This negative correlation then decays with increasing separation and vanishes at $r \approx 1.4\ell$. This result shows that the effective spin-spin interaction is predominantly antiferromagnetic at short distances. Indeed, as has been demonstrated in the case of lattice s-d models at half-filling, the electron-mediated interaction in the large coupling limit is antiferromagnetic [213]. This can be understood from the $J \to \infty$ (or $t_0 = 0$) limit where the ground states are given by configurations with exactly one electron per atom. As atoms are decoupled from each other, the local moment $\mathbf{S}_i$, which is entangled with the spin of localized electron, can point in arbitrary directions. This macroscopic ground-state degeneracy is lifted in the presence of a small t_0. Standard second-order perturbation leads to an effective antiferromagnetic interaction $J_{ij}^{\text{eff}} = t_{ij}^2/J > 0$.

As the system further relaxes toward lower energy states, a weak ferromagnetic correlation starts to develop at an intermediate distance, as indicated by the small peak at $r \sim 1.5\ell$ for the correlation functions at $t \gtrsim 150$. Compared with lattice models where electron hopping is restricted to nearest neighbors, the emergence of

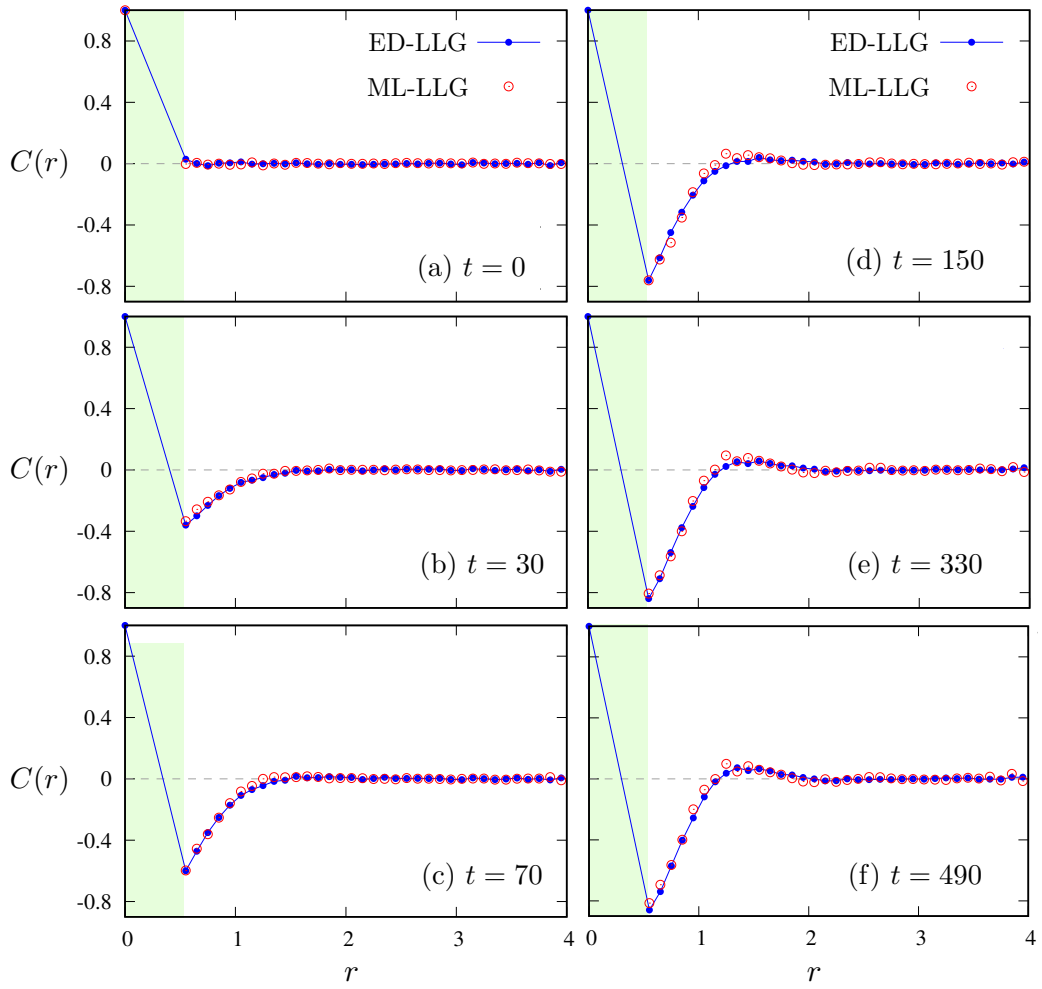

Figure 4.4: Dynamical benchmark of the ML force-field model for the random s-d system. The spin-spin correlation functions $C(r_{ij}) = \langle \mathbf{S}_i \cdot \mathbf{S}_j \rangle$ at various time-steps after a thermal quench are obtained from LLG simulations of $N = 100$ spins using the ML force-field model and the ED method. The radius is measured in units of ℓ, and the shaded area indicates the forbidden region where $r < r_{\text{min}}$. The simulation time is measured in units of $(\gamma t_0)^{-1}$, where γ is the gyromagnetic ratio and t_0 is the energy scale of electron hopping.

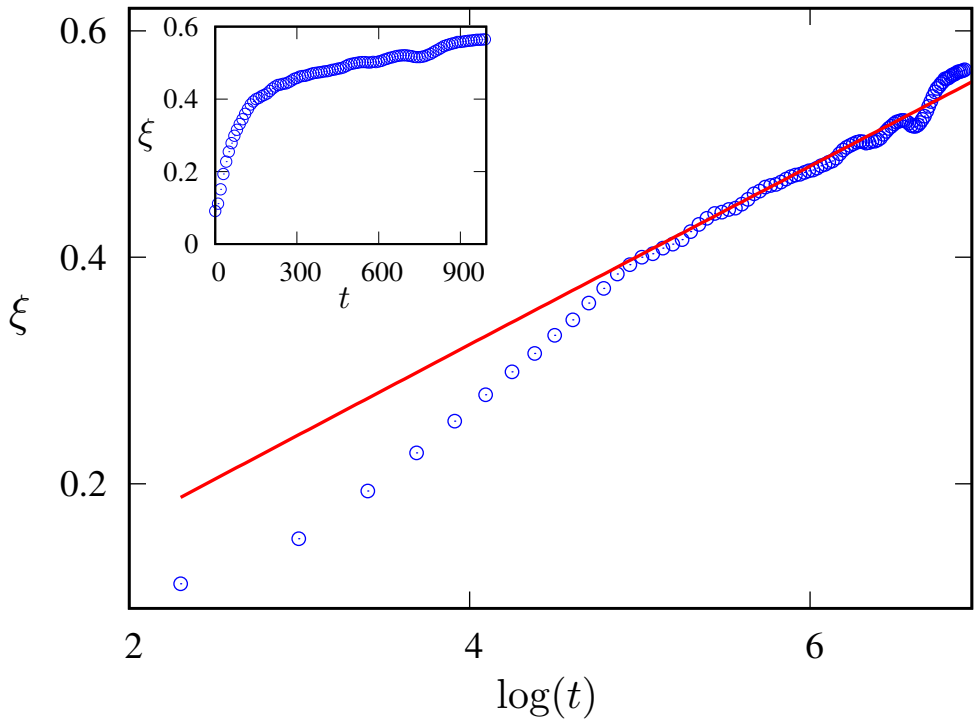

Figure 4.5: The correlation length ξ as a function of the the logarithm of time $\log(t)$ in a log-log plot; the inset shows the $\xi(t)$ in the linear scale. ML-LLG simulations are performed on a random s-d system of $N = 500$ atoms. The length ξ is computed from the resultant ensemble-averaged correlation functions after a thermal quench at $t = 0$. The simulation time is measured in units of $(\gamma t_0)^{-1}$, where γ is the gyromagnetic ratio and t_0 is the energy scale of electron hopping. The red line corresponds to a power law $\xi(t) \sim (\log t)^{1/\psi}$ with an exponent $\psi \approx 1.02$.

this ferromagnetic correlation at intermediate separation indicates the complexity of electron-mediated interactions in disordered s-d systems. Interestingly, our results show that the effective spin-spin interactions exhibits an RKKY-like oscillation, although with a much reduced amplitude, even in the large coupling regime.

To quantify the relaxation process, we compute a time-dependent correlation length $\xi(t)$ from the correlation functions. For a well-defined exponential decaying correlation, $C(r) \sim \exp(-r/\xi)$, the correlation length can be easily computed from

the large r behavior. Due to the more complicated forms of the spin-spin correlation functions in our case, we employ an empirical formula to obtain the correlation length $\xi(t) = \int_0^\infty r\big|C(r,t)\big|dr \Big/ \int_0^\infty \big|C(r,t)\big|dr$, It can be easily shown that this formula correctly reproduces the correlation length for an exponential-decaying function as well as for correlation functions exhibiting a dynamical scaling: $C(r,t) = \mathcal{F}(r/\xi(t))$, where $\mathcal{F}(x)$ is the scaling function. Applying this formula to the correlation functions obtained from ML-LLG simulations for a random s-d model with $N = 500$ atoms, the result is shown in the inset of Fig. 4.5. The ML model is trained from a system of $N = 100$ atoms with the same model parameters as discussed above. The ensemble-averaged correlation functions here are obtained from 10 different random realizations of atomic configurations and 20 independent initial random spin states.

As the system relaxes after the thermal quench, the correlation length quickly increases initially. The late-stage relaxation dynamics of a spin glass, often characterized by aging phenomena, is intimately related to its complex energy landscape. Although there is no conventional long-range order in spin glasses, the relaxation processes can still be viewed as the growth of "ordered" domains in which a local energy minimum is attained. It is worth noting that growth of ordered domains in a conventional symmetry breaking phase is controlled by topological defects of the associated order parameter fields and is often characterized by a power-law behavior [70].

On the other hand, the growth of the spin-glass order has been discussed within the phenomenological droplet model, which predicts a growth law $R(t) \sim (\log t)^{1/\psi}$,

where $R(t)$ is a measure of the linear domain size and ψ is called the barrier exponent which characterizes the activation barrier height at the interface [214, 215]. Using the extracted correlation length as a qualitative measure of the domain size, Fig. 4.5 shows $\xi(t)$ versus $\log(t)$ in a log-log plot. The late-stage domain coarsening can be approximated by a power-law with an exponent $\psi \approx 1.02$. Our preliminary results are consistent with the scaling theory of the droplet model. More systematic investigation of the aging dynamics and domain growth will be left for future work.

4.5 Conclusion and Outlook

In summary, we have introduced a scalable ML force-field model tailored for the spin dynamics in metallic spin glasses. Our methodology extends the BP frameworks that are prevalently employed in ML-based force-field models for quantum molecular dynamics (MD). Our study concentrates on spin glass systems characterized by three-dimensional Heisenberg spins, which are representative of prototype metallic spin-glass materials such as dilute magnetic alloys. To ensure the preservation of rotational symmetry in both real and spin spaces, we have refined the ACSF atomic descriptor, a tool extensively utilized in ML-based MD simulations, by incorporating spin degrees of freedom into the symmetry functions. As a demonstration of our approach, we examine the relaxation dynamics of an amorphous version of the classical s-d model. It is demonstrated that the trained ML model not only precisely predicts the effective local fields but also effectively represents the dynamical evolution of spins when subjected to a thermal quench.

Our work also opens a new avenue to large-scale simulations of nonequilibrium dynamical phenomena in spin glasses. Partly because of the computational complexity, most large-scale studies of spin-glass dynamics are based on short-range random-J spin models on a lattice. Moreover, local Monte Carlo updates are employed as a surrogate dynamics for Ising-type spin-glass models. Here we outline a general approach to derive realistic off-lattice spin-glass model with electron-mediated spin-spin interactions. Our work demonstrates the proof of principle that accurate ML force-field models can be developed for a large class of metallic spin glass systems with local s-d electron-spin coupling. Importantly, the efficiency of ML models, which are essentially effective classical spin models, could enable large-scale simulations which are essential for understanding the dynamical properties of spin glass systems.

In particular, our ML framework could also be used to model dilute magnetic alloys such as CuMn. As discussed in the Introduction, DFT has been applied to solve the magnetic ground states of a few representative magnetic alloys for a given atomic configuration where the magnetic atoms randomly occupy sites of the host lattice. However, such first-principles calculations of the magnetic moments and the corresponding effective fields are very time-consuming even for relatively small systems. In practice, often an effective Heisenberg model is first obtained by fitting the exchange interactions with DFT energy. An alternative approach is to derive an effective s-d Hamiltonian from first-principles calculations, which can then be used to develop a ML force field model as outlined in this work.

Chapter 5

Machine Learning for Silicon Band Structures

5.1 Introduction

Recent advances in electronic structure methods and rapid progress in computational capabilities and artificial intelligence techniques have enabled accurate and fast computation of materials' properties. This new paradigm of materials research is further assisted by the creation of large freely available databases containing many years worth of human knowledge [216, 217]. For example, several ML models have been developed for accurate and efficient structure-property mapping from large databases of Kohn-Sham DFT calculations [39–42]. Another prominent application is the ML based force-field or interatomic potential models trained by dataset from DFT calculations [91–93, 96, 97, 125, 127, 128, 175–177]. Such ML models, which are essentially classical force-field models, yet with a desired quantum accuracy, allows for larger scale and longer time *ab initio* MD simulations. More fundamentally, ML models are also shown to provide accurate approximations of the density functionals or

the Hohenberg-Kohn mapping from external potential to electron density functionals [218–223].

The integration of ML and data science techniques with *ab initio* electronic structure methods have provided a tantalizing prospect of inverse materials design where a novel material of a given functionality can be predicted from available experimental measurements and theoretical calculations [224–227]. Yet, despite tremendous progress, efficient calculation of materials' properties beyond the idealized zero temperature remains a challenge for a successful data-driven design and discovery pipeline. In particular, one important thermal effect is the phonon-induced renormalization of electronic structures [228,229]. This renormalization is the main mechanism for the temperature dependence of band gap energy.

A well-developed first-principles approach to incorporate electron-phonon coupling is based on the density functional perturbation theory (DFPT) [230–232]. For example, both phonon dispersion relations and electron-phonon matrix elements can be obtained from DFPT. These calculations can then be combined with the perturbative Feynman diagram methods to compute, e.g. the phonon-induced electron self-energies, including both the Fan and the Debye-Waller terms [233–236]. A consistent theory of temperature-dependent band structures up to second-order in electron-phonon coupling was developed by Allen, Heine and Cardona (AHC) [234,235]. Various phonon-induced thermal effects, such as band gap renormalization and broadening of absorption edge, can be included within this framework [237–239]. The perturba-

tive calculation of the electron-phonon interaction can also be elegantly integrated with the *GW* approximation [240–242] to partially account for electron-electron interactions. However, as this first-principles approach is built on the Kohn-Sham DFT, it is also restricted to this particular electronic structure method.

The frozen phonon method [243–249] offers an alternative framework for first-principles electron-phonon calculations. This approach is also more general in the sense that it can be used with any electronic structure solver, which implies that systems with weak and strong electronic correlations can be treated on the same footing. In fact, one of the earliest *ab initio* electron-phonon calculations was based on the frozen phonon method [245]. The central idea of this approach is rather straightforward: by comparing electronic structure solutions without and with the atomic displacements of a certain normal mode of the lattice, one can numerically compute the shift in electron eigen-energies or other physical observables induced by this particular phonon modes. This numerical calculation, often based on the finite difference method up to quadratic order, is then repeated for all phonon modes allowed in a supercell [249, 250].

Alternatively, thermal effects of lattice fluctuations can be directly computed by MC sampling of thermal phonon configurations [251–254]. Each sampled configuration of the frozen nuclei defines an electronic structure problem with frozen atomic displacements. Expectation values of physical quantities, such as band gap, at finite temperatures are obtained by averaging over solutions of the sampled configurations.

Instead of MC sampling, dynamical methods such as adiabatic *ab initio* MD [255–257] and path-integral MD [258–261] have also been used to generate the phonon configurations. The vibrational expectation value of physical quantities here is obtained by averaging along dynamical paths.

Compared with the above finite-difference frozen phonon method, the approach of sampling phonon configurations, either stochastically or dynamically, has the advantage that higher-order electron-phonon couplings can be straightforwardly included [253,254]. However, the sampling approach is computationally very demanding as an electronic structure problem has to be solved for each sampled frozen phonon configuration in a supercell, and accurate evaluation of physical quantities requires a large number of configurations. To circumvent this computational difficulty, it was shown that the configurational averaging can be approximated by averaging over a set of special configurations called thermal lines [262, 263] or even by fully deterministic supercell calculations based on a single optimal configuration of the atomic positions [264]. While these approximations significantly reduce the cost of repeated electronic structure calculations, they are exact only in the thermodynamic limit of large supercells. In practical implementations of finite supercells, their accuracy has to be carefully benchmarked.

In this work, we propose a ML approach to solve the efficiency issue of configurational averaging method. The central idea is to build a ML model that accurately predicts the physical quantities of interest corresponding to a configuration of nuclei

in a supercell. MC algorithm based on an *ab initio* phonon model is used to sample atomic configurations, and the trained ML model is employed to efficiently predict the corresponding physical property. As a proof of principle, we apply this MC-ML approach to compute the temperature dependent electronic band gap of silicon crystal. We use silicon as an example as it has been extensively studied by using various methods and thus provides a testbed for validating our methods. Our approach not only produces accurate temperature dependence of the gap energy, but also significantly reduces the number of DFT calculations in the process.

The ML model in our proposed framework can be viewed as a special case of the general ML structure-property mapping models [265–269] which have played a central role in the high-throughput materials design strategy. A proper descriptor of the input structure is an important ingredient of such ML models. In our case, the descriptor is particularly important for preserving the symmetry properties of the original electronic model in the mapping from phonons to properties. To this end, we employ group-theoretical methods to obtain generalized coordinates of atomic configurations in a supercell which are invariant under the point group symmetry of the crystal. The use of the phonon descriptor enhances both the training efficiency and the accuracy of the ML model.

5.2 Stochastic Formalism and Machine Learning Models

Our framework of vibronically renormalized electronic band structures is based on the stochastic frozen-phonon method. There are two major components in our framework: (i) the *ab initio* lattice dynamics model for sampling the phonon configurations in a supercell, and (ii) the ML model that maps the phonon configuration to the physical properties of interest, e.g. the electronic band gap energy.

5.2.1 *Ab Initio* Phonon Models

The fundamental assumption of most *ab initio* electron-phonon calculations, especially the frozen-phonon methods, is the Born-Oppenheimer or adiabatic approximation. The fact that masses of nuclei are much larger than that of electron indicates that the electron velocities are much larger than the nuclear ones, suggesting that electrons can follow the motion of the slow nuclei almost instantaneously. The well-separated time scales of the two sets of degrees of freedom allows one to write the full nuclei-electron wave function as a product form $|\Psi\rangle = |\chi\rangle \otimes |\psi\rangle$, where $|\chi\rangle$ and $|\psi\rangle$ are the phonon and electron wave functions, respectively. Under the adiabatic approximation, one can integrate out the electron degrees of freedom to obtain a potential energy surface $\mathcal{E}(\mathbf{R}_1, \mathbf{R}_2, \cdots, \mathbf{R}_N)$ as a function of the coordinates $\mathbf{R}_i$ of the nuclei. Practically, this involves solving an electronic structure problem with each atomic configuration frozen in $\mathbf{R}_i$ based on DFT calculations, or other electronic structure methods.

The Hamiltonian of nuclei in Born-Oppenheimer approximation is given by $\mathcal{H}_{\text{nucl}} = \sum_i \mathbf{P}_i^2/2M_i + \mathcal{E}(\{\mathbf{R}_i\})$, where $\mathbf{P}_i$ is the momentum operator of the i-th nuclei and M_i is the corresponding mass. However, since the computation of the potential energy for each atomic configuration $\{\mathbf{R}_i\}$ requires solving an electronic structure problem, direct *ab initio* determination of the phonon modes is computationally intractable, if not impossible. As discussed in Sec. 5.1, ML force-field approaches offer an efficient and accurate approximation to the potential energy surface. However, due to the lack of explicit analytical expressions for $\mathcal{E}(\{\mathbf{R}_i\})$, the phonon modes cannot be directly derived from the ML model. Vibrational thermal average has to be done in conjunction with dynamical simulations based on the ML force-field models.

Direct phonon or lattice models can be obtained based on the harmonic approximation to the nuclear Hamiltonian $\mathcal{H}_{\text{nucl}}$. For crystalline systems, nuclei only undergo small amplitude oscillations about their equilibrium positions. We can then write the nuclear positions as $\mathbf{R}_i = \mathbf{R}_i^{(0)} + \mathbf{u}_i$, where $\mathbf{R}_i^{(0)}$ is the equilibrium position and $\mathbf{u}_i$ is the displacement vector. Assuming small displacements, one can then express the potential energy $\mathcal{E}(\{\mathbf{R}_i\})$ in a power series of the displacement vectors. Since the equilibrium positions minimize the potential energy, the linear term in the expansion vanishes. To the leading second-order, the potential energy is $\mathcal{E} = \mathcal{E}_0 + \sum_{ij} \sum_{\alpha,\beta} D_{i\alpha,j\beta} u_i^\alpha u_j^\beta$, where u_i^α denotes the α-component of the displacement vector of the i-th nucleus ($\alpha = x, y,$ and z), and $D_{i\alpha,j\beta} = \partial^2\mathcal{E}/\partial u_{i,\alpha}\partial u_{j,\beta}$, evaluated at the equilibrium positions $\{\mathbf{R}^{(0)}\}$, is the dynamical matrix which encapsulates the

normal modes of the lattice.

The normal modes $\mathbf{V}^{(\nu)}_{s,\mathbf{k}}$ of the crystalline systems are obtained by diagonalizing the harmonic Hamiltonian, here ν is the branch-index, s is the sublattice index, and $\mathbf{k}$ is the reciprocal lattice wave vector. The displacement vectors can then be expressed as

$$\mathbf{u}_i = \sum_{\nu,\mathbf{k}} \xi_{\nu,\mathbf{k}} \, \mathbf{V}^{(\nu)}_{s_i,\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{R}^{(0)}_i}, \tag{5.1}$$

where $\xi_{\nu,\mathbf{k}}$ are the effective or normal-mode coordinates. In terms of the effective coordinates, the vibrational dynamics of the lattice is described by the following phonon Hamiltonian

$$\mathcal{H}_{\text{phonon}} = \sum_{\nu} \sum_{\mathbf{k}} \left(\frac{-\hbar^2}{2} \frac{\partial^2}{\partial \xi^2_{\nu,\mathbf{k}}} + \frac{1}{2} \Omega^2_{\nu,\mathbf{k}} \xi^2_{\nu,\mathbf{k}} \right). \tag{5.2}$$

Here $\Omega_{\nu\mathbf{k}}$ is the eigen-frequency of the normal modes. *Ab initio* calculation of the phonon modes can now be routinely computed using DFT methods, both based on the DFPT or frozen-phonon formalisms. Computation of the dynamical matrices using methods beyond DFT have also been demonstrated [270, 271].

5.2.2 Stochastic Vibrational Average

Next we consider the computation of observables such as in the frozen-phonon framework. Quantum mechanically, an observable corresponds to a Hermitian operator $\mathcal{O}(\{\mathbf{r}_l\}, \{\mathbf{R}_i\})$ which depends on both electron and nuclei coordinates. By integrating out the electronic degrees of freedom within the Born-Oppenheimer approximation,

the observable

$$\mathcal{O}(\{\mathbf{R}_i\}) = \langle\psi(\{\mathbf{r}_l\})|\mathcal{O}(\{\mathbf{r}_l\},\{\mathbf{R}_i\})|\psi(\{\mathbf{r}_l\})\rangle$$

becomes a function of nuclear coordinates only. The electron quantum state $|\psi(\{\mathbf{r}_l\})\rangle$ in general also implicitly depends on the nuclei configuration $\{\mathbf{R}_i\}$. The vibrational expectation value of the observable quantity at finite temperature T is then given by [249]

$$\langle\mathcal{O}(\mathcal{T})\rangle = \frac{1}{\mathcal{Z}}\sum_M \langle\chi_M(\{\mathbf{R}_i\})|\mathcal{O}(\{\mathbf{R}_i\})|\chi_M(\{\mathbf{R}_i\})\rangle e^{-\mathcal{E}_M/k_BT}, \tag{5.3}$$

where the summation is carried over nuclear quantum states $|\chi_M(\{\mathbf{R}_i\})\rangle$ whose energy is $\mathcal{E}_M$, $\mathcal{Z} = \sum_M \exp[-\mathcal{E}_M/k_BT]$ is the partition function of phonons and k_B is the Boltzmann constant.

In this work, the observable of interest is the electronic band gap E_g but the formalism applies to any other observables. Similar formulas can also be derived for matrix elements $\mathcal{A}_{jk}(\{\mathbf{R}_i\})$ between different electronic states; the resultant expectation values will determine the transition rates such as optical absorption and dielectric coefficients [272, 273].

In the harmonic approximation, the quantum number of the nuclei is given by an array of integers $M = \{n_{\nu,\mathbf{k}}\}$, and the corresponding wave function becomes a product state: $\chi_M(\{\mathbf{R}_i\}) \propto \prod_{\nu,\mathbf{k}} H_{n_{\nu,\mathbf{k}}}(\xi_{\nu,\mathbf{k}})\exp[-\xi^2_{\nu,\mathbf{k}}/2]$, where $H_n(x)$ is the n-th order Hermite polynomial, $\mathbf{R}_i = \mathbf{R}_i^{(0)} + \mathbf{u}_i$, and the displacement vectors $\mathbf{u}_i$ is ex-

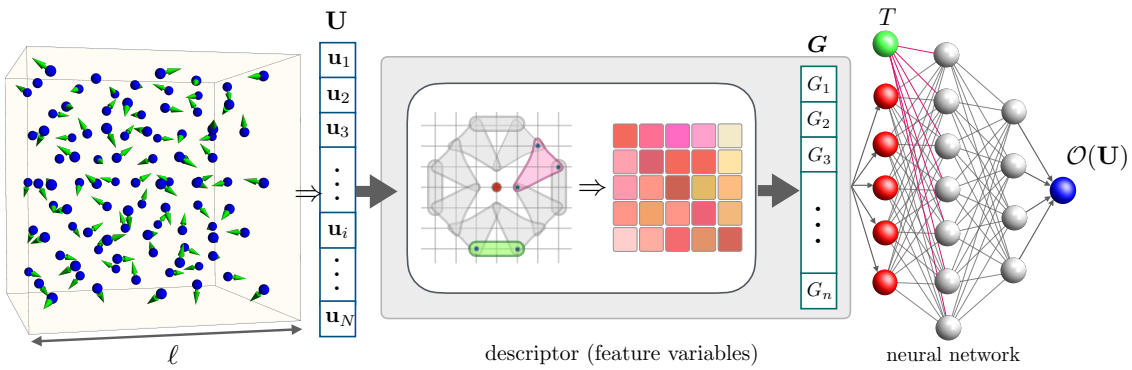

Figure 5.1: Machine learning model for band gap prediction for a given atomic displacement field $\mathbf{u}_i$ within the a box of linear size ℓ. The ML model is composed of two central components: the descriptor and the neural network. The input of the ML model is a cubic block of displacement vectors $\mathbf{u}_i$. The descriptor corresponds to a representation or feature variables $\boldsymbol{G} = (G_1, G_2, G_3, \cdots)$ of the displacement vectors that is invariant under symmetry operations of the point group of the lattice, which in the case of Si is the T_d group. The complex dependence of the band gap on the nuclei configurations is encoded in the NN which takes the symmetry-invariant feature variables $\boldsymbol{G}$ as well as the temperature T as the input, and the predicted gap energy at the output.

pressed in terms of the normal mode coordinates $\xi_{\nu,\mathbf{k}}$ through Eq. (5.1). As the atomic displacements will play a central role in this work, we introduce the notation $\mathbf{U} = \{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_N\}$ to denote the collection of displacement vectors. Within the quadratic approximation, the summation over the integer quantum numbers $n_{\nu,\mathbf{k}}$ in Eq. (5.3) can be recast into integrals over the normal mode coordinates:

$$\langle \mathcal{O}(\mathcal{T}) \rangle = \prod_{\nu,\mathbf{k}} \int d\xi_{\nu,\mathbf{k}} \frac{\exp(-\xi^2_{\nu,\mathbf{k}}/2\sigma^2_{\nu\mathbf{k},T})}{\sqrt{2\pi}\sigma_{\nu\mathbf{k},T}} \mathcal{O}(\mathbf{U}), \tag{5.4}$$

where the width $\sigma_{\nu,\mathbf{k}}$ of the Gaussian integral is given by:

$$\sigma^2_{\nu\mathbf{k},T} = (2n_{\nu\mathbf{k},T} + 1) \left(\frac{\hbar}{2\Omega_{\nu,\mathbf{k}}} \right)^2, \tag{5.5}$$

and $n_{\nu\mathbf{k},T} = [e^{\hbar\Omega_{\nu\mathbf{k}}/k_BT} - 1]^{-1}$ is the Bose-Einstein occupation of the $(\nu, \mathbf{k})$ phonon. The displacement configuration $\mathbf{U}$ in $\mathcal{O}(\mathbf{U})$ of the Gaussian integrals is again expressed as superposition of the phonon modes as in Eq. (5.1).

The computation of the multidimensional Gaussian integral, however, is computationally challenging for observables with a complex dependence on the nuclei configuration. The calculation can be simplified by expanding the observables as a power series expansion of displacement vectors, often up to second order, the Gaussian integral can then be analytically evaluated. Within the DFT framework, the expansion coefficients can be similarly computed using DFPT [230, 274–276] or frozen-phonon methods [243, 246].

An alternative approach, which also allows one to go beyond perturbation theory and DFT, is the stochastic Monte Carlo method [251–253]. To this end, we rewrite the integral in Eq. (5.4) as

$$\langle \mathcal{O}(T) \rangle = \int \mathcal{D}\mathbf{U}\, \pi_G(\mathbf{U}) \mathcal{O}(\mathbf{U}) \approx \frac{1}{\mathcal{N}} \sum_{k=1}^{\mathcal{N}} \mathcal{O}(\mathbf{U}^{(k)}). \tag{5.6}$$

Here $\mathcal{D}\mathbf{U}$ is a simplified notation for the multiple integrals over the effective coordinates $\xi_{\nu,\mathbf{k}}$, and $\pi_G(\mathbf{U})$ is an effective probability density function of $\mathbf{U}$, corresponding to the product of Gaussian density functions. The Metropolis-Hastings algorithm is then used to implement a transition probability $P(\mathbf{U} \to \mathbf{U}') = \min[1, \pi_G(\mathbf{U}')/\pi_G(\mathbf{U})]$ for sampling the displacement configurations. A series of displacement configurations $\mathbf{U}^{(k)}$ sampled from the Markov-Chain Monte Carlo is used to compute the vibrational

average as indicated by the approximation in Eq. (5.6) above.

5.2.3 Machine Learning Model

Having obtained an *ab initio* phonon model, sampling of the nuclei configuration, or more specifically the displacement field $\mathbf{u}$, is relatively straightforward. The bottleneck of the stochastic approach to vibrational thermal averaging is the computation of the observable $\mathcal{O}(\mathbf{U})$ for the sampled configurations. As discussed in Sec. 5.1, to overcome this computational complexity, a fully deterministic method based on the idea of effective vibrational configurations is proposed to replace the stochastic MC sampling. It has been shown that the vibrational average can be approximated by either a single nuclear configuration, called special displacement configuration [264], or a reduced MC sampling along so-called thermal lines [262]. While these methods greatly simplify the problem and have been shown to work reasonably well for a wide range of materials [277], they are essentially based on quadratic approximation of electron-phonon theory and therefore may fail to capture higher order effects in certain class of materials [278]. In addition, it has been shown very recently that the special displacement method fails to correctly account for the renormalization of energy levels of nitrogen vacancy centers in diamond [279]. Hence, it is desirable to develop an efficient vibrational averaging method without sacrificing too much accuracy.

Modern ML methods, especially supervised learning, offer a promising solution to the issue of computational efficiency by providing an accurate and efficient mapping

from the phonon configuration $\mathbf{U}$ to the electronic observable $\mathcal{O}$ of interest. Fundamentally, this approach relies on the universal approximation theorem which shows that deep multilayer NN can be trained to accurately approximate any given high-dimensional functions [280, 281]. A schematic diagram of our proposed ML model is shown in FIG. 5.1; the input of the model is a frozen phonon configuration represented by a displacement field $\mathbf{U}$ and the temperature T, while the output is the predicted observable corresponding to the expectation value $\mathcal{O}(\mathbf{U}) = \langle\psi|\hat{\mathcal{O}}|\psi\rangle$ where $|\psi\rangle$ is calculated from some electronic structure methods. Using an *ab initio* phonon model discussed in Sec. 5.2.1 to generate a series of displacement fields $\mathbf{U}_T^{(r)}$ at a few predetermined temperatures T, the electronic states $|\psi^{(r)}\rangle$ are then solved consistently based on the same first-principle methods. A dataset of the form $\left\{\mathbf{U}_T^{(r)}, T;\, \mathcal{O}^{(r)}\right\}$ is used to train the above ML model.

Broadly speaking, our proposed ML model for $\mathcal{O}(\mathbf{U})$ can be viewed as a special case of ML-based modeling of structure-property relationships in materials science [98, 265–267, 269, 282, 283]. Yet, while the inputs to most conventional ML structure-property model are static disordered or quasi-random structures of the system, a frozen phonon configuration as represented by a displacement field $\mathbf{U}$ describes an instantaneous structure of nuclei. In this sense, our ML model is also similar to the ML-based interatomic potential models which maps a dynamical atomic configuration in a finite range to a local atomic energy [91–93, 96, 97, 125, 127, 128, 175].

As shown in Fig. 5.1, there are two central components of the ML model: (i) a

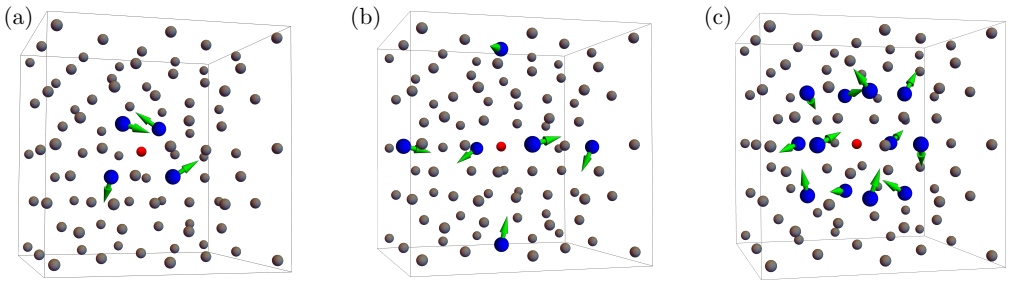

Figure 5.2: Group of atoms sharing same distance to the center of the box form an invariant block of the representation matrix of the symmetry group. Panels (a), (b), and (c) show such invariant groups of 4, 6, and 12 atoms, respectively, in a diamond lattice.

learning model based on deep multi-layer NN and (ii) a descriptor which provides generalized or effective coordinates of frozen phonons. The ML model works as follows. First, the set of displacement vectors $\mathbf{U} = \{\mathbf{u}_i\}$ that characterizes the atomic configuration within a cubic box is mapped to a collection of symmetry-invariant feature variables $\boldsymbol{G} = \{G_1, G_2, \cdots\}$, also known as a descriptor. These feature variables are then fed into the NN which produces the predicted observable $\mathcal{O}(\mathbf{U})$ at the output node, thus bypassing the time-consuming electronic structure calculations. Integrating this ML phonon-to-property model into the stochastic framework offers an efficient yet accurate vibrational thermal average.

5.3 Descriptors for Frozen Phonon Configurations

While a NN-based learning model is used mainly because of its superb approximation capability, the descriptor plays a crucial role not only to facilitate the training efficiency, but also to preserve the symmetry of the original electron-phonon system. In

conventional image processing models, the so-called data augmentation technique is commonly employed to artificially expand the size and diversity of a training dataset by applying various transformations to the original images. These transformations can include rotation, scaling, flipping, translation, or color changes, among others. The goal is to make the model more robust and invariant to these transformations, which helps improve generalization and performance on unseen data. However, despite the universal approximation capability of NN models, symmetries of the original physical Hamiltonian can only be learned approximately even with the augmented datasets.

To ensure that the symmetry of the original electronic system is exactly preserved in the ML model, a proper representation of the frozen phonon configuration is to be fed into the learning model instead of the raw displacement vectors. Indeed, similar symmetry-based descriptors also play a crucial role in both ML structure-property models as well as ML force-field models for *ab initio* MD simulations. This is particularly important for the BP type ML schemes for interatomic potentials. The output of such BP-type ML models is an atomic energy which, as a scalar, is invariant under symmetry operations of the crystalline system including rotation, reflection, and permutation of same-species atoms. Consequently, a proper descriptor of the local neighborhood not only should be able to differentiate distinct atom configurations, but also remain invariant under the above symmetry transformations.

Over the past decade, several learning models based on a variety of atomic descrip-

tors have been proposed [91,92,96,104,181–185]. For example, the ACSF, which is one of the most popular atomic descriptors, is based on the two-body (relative distances) and three-body (relative angles) invariants of an atomic configuration [91, 127]. The group-theoretical method, on the other hand, offers a more controlled approach to the construction of atomic representation based on the bispectrum coefficients [92, 104].

In our ML phonon-to-property models, the continuous translation and SO(3) rotation symmetries of crystalline systems are reduced to discrete point group symmetry associated with a supercell. On the other hand, the discrete rotation, reflection, and mirror symmetries of the point group not only transform the atoms in the supercell, but also act uniformly on all displacement vectors $\mathbf{u}_i$. The output of the ML models is often a scalar, such as band gap energy or optical absorption coefficient. As these quantities are invariant under symmetry operations of the point group, symmetry requirement means that two atomic displacements $\mathbf{U}^{(a)}$ and $\mathbf{U}^{(b)}$ related by symmetry operations, when fed into the ML model, should produce exactly the same scalar observable. It is worth noting that, in the absence of descriptors, such symmetry-related configurations will be treated by the neural net as unrelated inputs. To ensure the invariance of the predicted scalar quantity (gap energy), the feature variables $\boldsymbol{G} = \{G_1, G_2, \cdots\}$ have to remain invariant under symmetry operators of the point group associated with the lattice.

A general theory of descriptors for lattice systems was recently developed based on group-theoretical methods [58, 105]. Implementations of specific descriptors have also

been demonstrated mostly for lattice model systems in condensed matter physics [59, 99,131,132,179]. Here we adopt the group-theoretical approach to develop a symmetry-invariant descriptor for the frozen phonon configuration in a supercell. To this end, the first step is to decompose a given displacement field $\mathbf{U}$ into irreducible representations, from which invariant feature variables can be obtained. In the case of silicon where atoms reside on a diamond lattice, the relevant point group of the site symmetry is T_d. Take an arbitrary lattice point as the center of reference frame, under a discrete rotation, represented by the 3×3 orthogonal matrix $\mathcal{A}$, the displacement vectors transform according to $\mathbf{u}_i \rightarrow \mathbf{u}'_j = \mathcal{A} \cdot \mathbf{u}_i$ where the two lattice points are related by $\mathbf{R}_j = \mathcal{A} \cdot \mathbf{R}_i$. For N atoms in the box, their displacement field thus constitute a $3N$-dimensional reducible representation of the T_d group. Since the distance is preserved by symmetry operations of the point group, the decomposition of this high-dimensional representation can be simplified as the representation matrix is automatically block-diagonalized, with each block corresponding to a fixed distance from the center-site of the block. FIG. 5.2 shows examples of such invariant group of atoms in the diamond lattice.

Since there are three components for each atomic displacement vector $\mathbf{u}_j$, the dimension of each block is related to the number of atoms n_b in the corresponding neighboring group through the relation $D_b = 3 \times n_b$. For the diamond lattice, there are four distinct types of neighboring groups with $n_b = 4$, 6, 12, and 24 atoms. Consider the simplest case of a 4-atoms group, as shown in FIG. 5.2(a). The dis-

placement vectors of the four atoms, denoted as **a**, **b**, **c**, and **d** for simplicity, form a $D_b = 12$ dimensional representation of the T_d group, which can be decomposed as $12 = A_1 \bigoplus E \bigoplus T_1 \bigoplus 2T_2$, where, A_1 denotes 1D representation, E denotes 2D representation and T denotes 3D representation. The A_1 component, for example, is given by the combination

$$\begin{aligned} f^{A_1} &= a_x + a_y + a_z - b_x - b_y + b_z \\ &+ c_x - c_y - c_z - d_x + d_y - d_z. \end{aligned}$$

Another example is the basis functions of one of the T_2 IR:

$$\begin{aligned} f_x^{T_2} &= a_x + b_x + c_x + d_x, \\ f_y^{T_2} &= a_y + b_y + c_y + d_y, \\ f_z^{T_2} &= a_z + b_z + c_z + d_z. \end{aligned}$$

Details of the IR decompositions and the basis functions can be found in Appendix A.

By repeating the same procedures for each block, the displacement field **U** is fully decomposed into the IR of the T_d group. For convenience, we arrange the resultant IR basis functions into a vector

$$\boldsymbol{f}_r^{\Gamma} = (f_{r,1}^{\Gamma}, f_{r,2}^{\Gamma}, \cdots, f_{r,D_\Gamma}^{\Gamma}), \tag{5.7}$$

where Γ labels the type of IR, r enumerates the multiple occurrence of IR-Γ in the decomposition of the displacement configuration, and D_Γ is the dimension of the

corresponding IR. Given these basis functions, one can immediately obtain a set of invariants called power spectrum $\{p_r^\Gamma\}$, which are the amplitude of each individual IR basis, i.e.

$$p_r^\Gamma = \left|\boldsymbol{f}_r^\Gamma\right|^2 . \tag{5.8}$$

The power spectrum coefficients constitute a basic set of feature variables that are invariant under symmetry transformations of the point group. However, a descriptor based only on the power spectrum is incomplete in the sense that the relative phases between different IRs are ignored. For example, the relative "angle" between two IRs of the same type: $\cos\theta = (\boldsymbol{f}_{r_1}^\Gamma \cdot \boldsymbol{f}_{r_2}^\Gamma)/|\boldsymbol{f}_{r_1}^\Gamma||\boldsymbol{f}_{r_2}^\Gamma|$ is also an invariant of the symmetry group. Without such phase information, the NN model might suffer from additional error due to the spurious symmetry, namely two IRs can freely rotate independent of each other.

A systematic approach to include all relevant invariants, including both amplitudes and relative phases, is the bispectrum method [103, 104]. The bispectrum coefficients are triple product of IR basis functions defined as

$$B_{r,r_1,r_2}^{\Gamma,\Gamma_1,\Gamma_2} = C_{\alpha,\beta,\gamma}^{\Gamma;\Gamma_1,\Gamma_2} f_{r,\alpha}^\Gamma f_{r_1,\beta}^{\Gamma_1} f_{r_2,\gamma}^{\Gamma_2} \tag{5.9}$$

where $C^{\Gamma;\Gamma_1,\Gamma_2}$ are the Clebsch-Gordan coefficients of the point group [103]. While a bispectrum descriptor provide a faithful invariant description of the phonon field, the number of all bispectrum coefficients increases cubically with the size of supercell. Moreover, many of them are redundant. In this work, we develop a descriptor similar

to the one used in Ref. [58, 105], that is modified from the bispectrum method. We introduce a set of reference basis functions $\boldsymbol{f}^{\Gamma}_{\text{ref}}$ for each IR Γ of the point group. These reference basis are computed by averaging large blocks of displacement variables, such that they are less sensitive to small changes in the supercell. For example, by dividing the supercell into 24 symmetry-related blocks for the case of T_d point group, we define the average displacement vector of a block-B as $\overline{\mathbf{u}}_B = \sum_{i\in B} \mathbf{u}_i$. The reference basis $\boldsymbol{f}^{\Gamma}_{\text{ref}}$ can then be computed from these block-averaged displacement vectors using exactly the same decomposition formulas.

Given the reference IR, we then define the relative "phase" θ^{Γ}_r of an IR as the projection of its basis functions onto the reference basis:

$$\exp\left(i\theta^{\Gamma}_r\right) \equiv \boldsymbol{f}^{\Gamma}_r \cdot \boldsymbol{f}^{\Gamma}_{\text{ref}} / |\boldsymbol{f}^{\Gamma}_r|\,|\boldsymbol{f}^{\Gamma}_{\text{ref}}|. \tag{5.10}$$

The relative phases between two IR basis of the same type can then be obtained via their respective phases with the reference. Finally, the relative phases between *different* IR basis are provided by the bispectrum coefficients computed entirely from the reference IRs. To summarize, the lattice descriptor are comprised of the following three types of variables:

$$\boldsymbol{G} = \{G_\ell\} = \{p^{\Gamma}_r,\, e^{i\theta^{\Gamma}_r}, B^{\Gamma,\Gamma_1,\Gamma_2}_{\text{ref}}\}. \tag{5.11}$$

Here $B^{\Gamma,\Gamma_1,\Gamma_2}_{\text{ref}}$ is the bispectrum coefficient in Eq. (5.9) where all three IRs are computed from the reference. These feature variables are not only invariant under the point-group symmetry, but also provide a faithful representation of phonon configu-

rations (module symmetry operations) in the supercell.

5.4 Computational Details and Results

5.4.1 DFT Calculations

First principles DFT calculations were done using Quantum Espresso (QE) [284] package. Perdew-Zunger (PZ) exchange-correlation functional [285] within the local density approximation (LDA) was used in all the calculations. Full lattice relaxation of primitive cell of Si crystal was performed which gave lattice parameter of 5.398 Å consistent with previous works [249]. The primitive BZ was sampled by using k-mesh of $24 \times 24 \times 24$ and energy cutoff of 30 Rydberg was used after careful convergence tests. The relaxed lattice was used to perform *ab initio* phonon calculation using DFPT [230–232] within the harmonic approximation as implemented in QE. For phonon calculations, we used a q-mesh of $8 \times 8 \times 8$. From the information of phonon eigen-frequencies $\Omega_{\nu,\mathbf{k}}$ and eigenvectors $\mathbf{V}^{(\nu)}_{s,\mathbf{k}}$, we generated atomic supercell configurations of size $6 \times 6 \times 6$ primitive unit cells containing 432 Si atoms using importance sampling Monte Carlo according to Eq. (5.4). Subsequently, DFT self-consistent calculation were performed on about 10% of the distorted atomic supercell configurations using k-mesh of $3 \times 3 \times 3$ to generate a database of electronic band gap as a function of atomic configurations.

In FIG. 5.3, we show the electron and phonon dispersion relations in the ground state of silicon crystal obtained from DFT calculations. These results are consistent with previous literature [286]. DFT calculations show an indirect band gap of 500

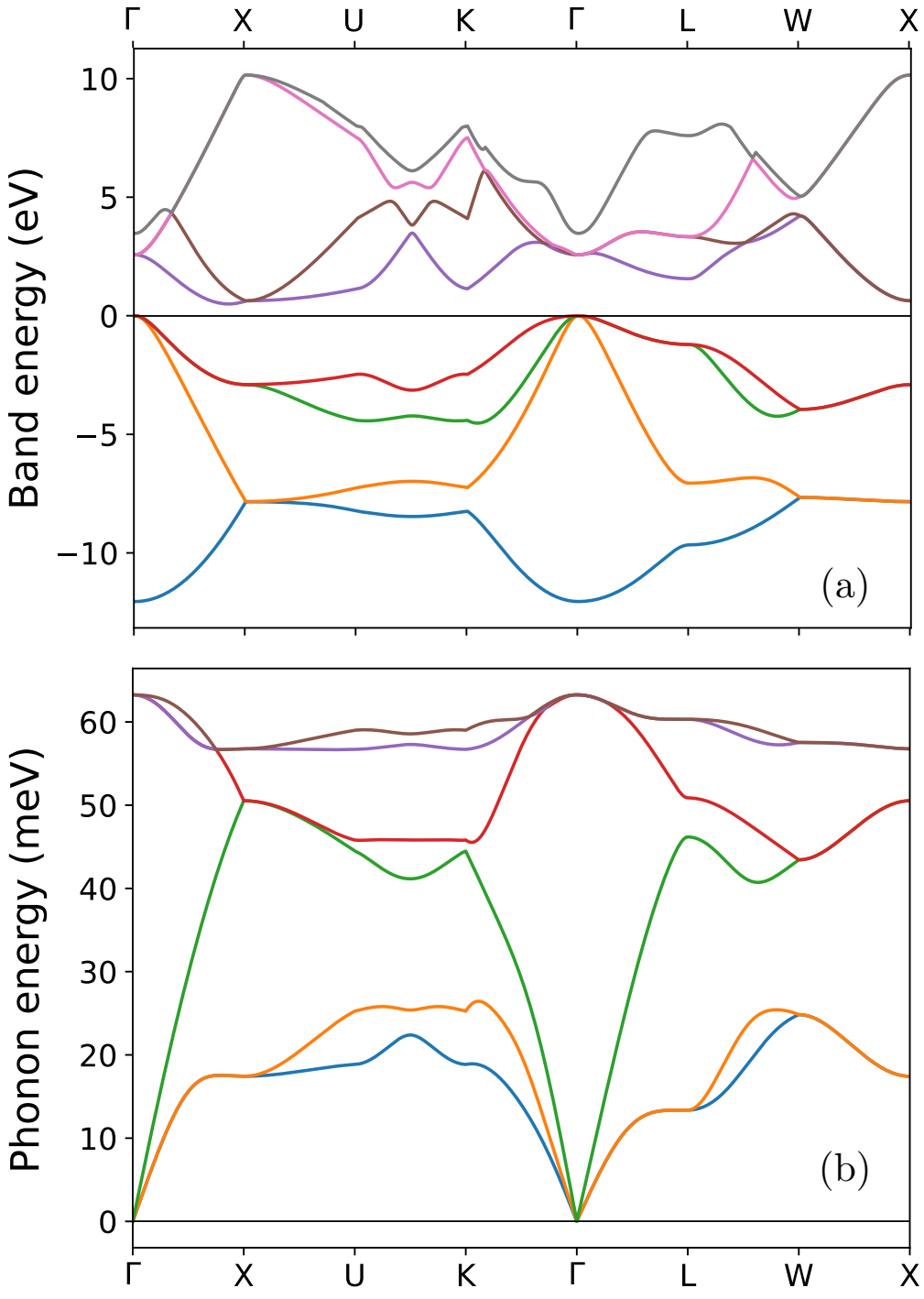

Figure 5.3: Ground state (a) electron and (b) phonon dispersion of Si crystal along high-symmetry directions of the lattice.

meV which is severely underestimated compared to the experimental band gap of 1.12 eV. Such band gap underestimation of insulators and semiconductors is a well known problem in DFT and more sophisticated methods like GW calculations are known to give better estimation of the band gap. In practice, GW correction of band gap is similar to the "scissor shift" of bands. Since we are interested in the band gap reduction induced by thermal phonons, this underestimation of band gap should not be an issue as the same GW-correction (to leading order) appears in both the perfect and perturbed crystal structures. Hence, in the following, we will use phonon induced band gap correction $\Delta E_g = E_g^{\text{eq}} - E_g^{\text{distorted}}$ as our observable of interest, where E_g^{eq} is the band gap of the perfect crystalline structure and $E_g^{\text{distorted}}$ is the band gap of the distorted structure.

The phonon dispersion shown in FIG. 5.3 is also consistent with previous studies. By sampling the normal mode coordinates $\xi_{\nu\mathbf{k}}$ according to Eq. (5.4), a displacement configuration $\mathbf{U} = \{\mathbf{u}_i\}$, or a frozen phonon configuration, within the supercell is constructed according to Eq. (5.1). Fixing the nuclei positions at $\mathbf{R}_i = \mathbf{R}_i^{(0)} + \mathbf{u}_i$ in the supercell, DFT calculations were performed to obtain the optimized electron density distribution. An energy gap $E_g(\mathbf{U})$ corresponding to a particular phonon configuration is computed as the difference between the lowest unoccupied and the highest occupied states. In FIG. 5.4, we show histogram plot of the band gap correction ΔE_g for a sample size of 100 at two different temperature values. As seen from the plots, the MC method has not fully converged and needs large number of samples to get

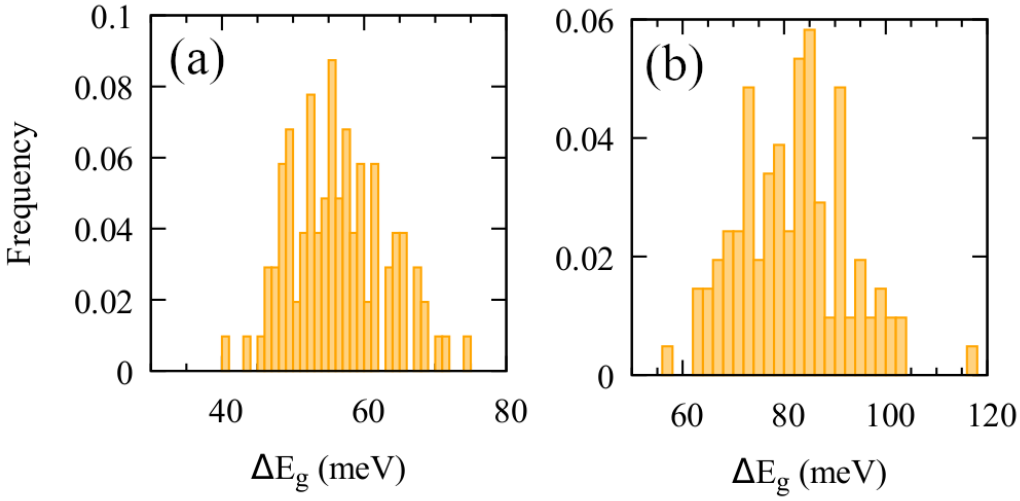

Figure 5.4: Histogram showing distribution of phonon-induced band gap correction at (a) $T = 0$K and (b) $T = 200$ K using 100 data points for Si-supercell of size $6 \times 6 \times 6$.

a better estimation of the band gap. In the following, we show that it is possible to get a better estimation (i.e. smaller standard deviation) of the band gap without incurring additional computational cost by first using the DFT calculated dataset to train a ML phonon-to-band gap mapping, and then using the ML model to predict band gap for a large number of configurations to obtain thermal average.

5.4.2 Neural Network Architecture and Model Training

As shown in FIG. 5.1, our learning model is based on a fully-connected NN, also known as a multilayer perceptron model. A fully-connected NN comprises of a series of fully connected linear layers with non-linear activation functions that are applied at each layer. The parameters of the j-th layer are the matrix of network weights $\mathbf{W}^{(j)}$ and bias vector $\mathbf{b}^{(j)}$. The weights introduce coupling between neurons of adjacent

Layer	Network
Input Layer	[838,2048][a]
Hidden Layer 1	[2048,1024]
Hidden Layer 2	[1024,512]
Hidden Layer 3	[512,256]
Output Layer	[256,1]

[a]Fully connected layer with arguments [input size, output size].

Table 5.1: The architecture of neural network. GeLU activation function and drop-out rate of 0.3 was used in all layers except output layer.

layers. The mapping of neurons $\boldsymbol{x}$ from one layer to the next is achieved by a linear transformation followed by a nonlinear activation function as

$$x_m^{(j+1)} = f_{\text{av}}\left(\sum_n W_{mn}^{(j)}\, x_n^{(j)} + b_m^{(j)}\right). \tag{5.12}$$

A commonly used activation function is the ReLU [287] defined as $f_{\text{av}}(x) := \max(x, 0)$. However, sometimes a large number of ReLU neurons remains inactivated in the model. Hence, often some variants of leaky ReLU [288] are used instead. In this work, we used the Gaussian Error Linear Units (GeLU) [289] activation function, which shows better performance than the standard ReLU.

We build a fully-connected NN with four hidden layers. The dimension of the hidden layers are (2048, 1024, 512, 256); see Table 5.1 for details. Moreover, a dropout layer with 0.3 dropout rate is introduced. As for the initialization of each layers, we utilize the Xavier uniform for the weights and the Normal distribution for the bias. Adam optimizer with initial 0.0001 cosine learning rate and l_2 regularization

coefficient 5×10^{-9} is used.

As noted above, the phonon configurations are sampled from a supercell consisting of $6 \times 6 \times 6$ primitive unit cells. Here a primitive unit cell corresponds to a parallelepiped formed from primitive vectors $\mathbf{a}_1 = (0, a/2, a/2)$, $\mathbf{a}_2 = (a/2, 0, a/2)$, and $\mathbf{a}_3 = (a/2, a/2, 0)$, where a is the length of a conventional cubic unit cell. The skewed primitive unit cell, however, does not preserve the cubic symmetry of the silicon crystal. To facilitate the incorporation of the point-group symmetry, here we instead use an input block comprised of $3 \times 3 \times 3$ cubic cells with a total of 279 atoms.

To further improve the statistics of the ML predictions, we apply the same ML model to multiple overlapping cubic blocks of the supercell. Specifically, we choose sites with coordinate $(i, j, k)a$ and $(i+1/2, j+1/2, k+1/2)a$ as the center, where i, j, k are integers ranging from 0 to 2. For these 54 sites, we consider their surrounding sites within a block of $3 \times 3 \times 3$ cubic cells which includes 279 sites. Since our descriptors keep the same dimension as the configuration, the number of the input features is $838 = 279 \times 3 + 1$, where first 837 features are the descriptors of 279 sites configuration and the rest feature is the temperature. The band gap energy corresponding to the phonon configuration $\mathbf{U}$ of the supercell is obtained by averaging ML predictions from these 54 blocks. Accordingly, the loss function used for training the NN is given by the following MSE

$$L = \left\langle \left| E_g^{\mathrm{DFT}} - \frac{1}{M} \sum_{k=1}^{M} \hat{E}_g^{(k)} \right|^2 \right\rangle, \tag{5.13}$$

where $\hat{E}_g^{(k)}$ denotes the ML predicted band gap energy for the k-th cubic block, $M =$

54 is the number of blocks, and $\langle \cdots \rangle$ indicates averaging over the training dataset. We have generated 103, 108, 113, and 204 distinct phonon configurations for four different temperatures at $T = 0, 100, 200$, and 300K, respectively, and performed fully self-consistent DFT calculations on these configurations to estimate the band gap E_g^{DFT}. For each temperature, 80 randomly chosen configurations and the corresponding band gap energies are used to train the ML model, while the rest of the configurations are kept as the validation and test dataset. After 600 epochs of training process with each batch including only one configuration, the loss function value is converged to around 0.0024297 for the training dataset and 0.00410986 for the validation dataset. The whole process takes 540 seconds on NVIDIA GPU A100 workstation.

FIG. 5.5 shows the parity plot of ML predicted band gap corrections Δ_g versus those from DFT calculations for four temperatures used in NN training. As expected, the phonon-induced correction to the band gap energy is enhanced at higher temperatures due to the increased number of thermal phonons. While an overall agreement was obtained between the ML predicted values and DFT results, the prediction accuracy varies with the temperatures. For a more detailed comparison, histogram pots of the prediction error $\delta = \Delta E_g^{\mathrm{ML}} - \Delta E_g^{\mathrm{DFT}}$ are shown in FIG. 5.6(a)–(d) for four different temperatures. The error is larger for higher temperature because the width of the Gaussian distribution used in MC sampling increases with temperature. Yet, it is remarkable that the error from NN trained on ~ 100 data points is merely of the order of 10 meV, thus demonstrating the robustness of our ML model.

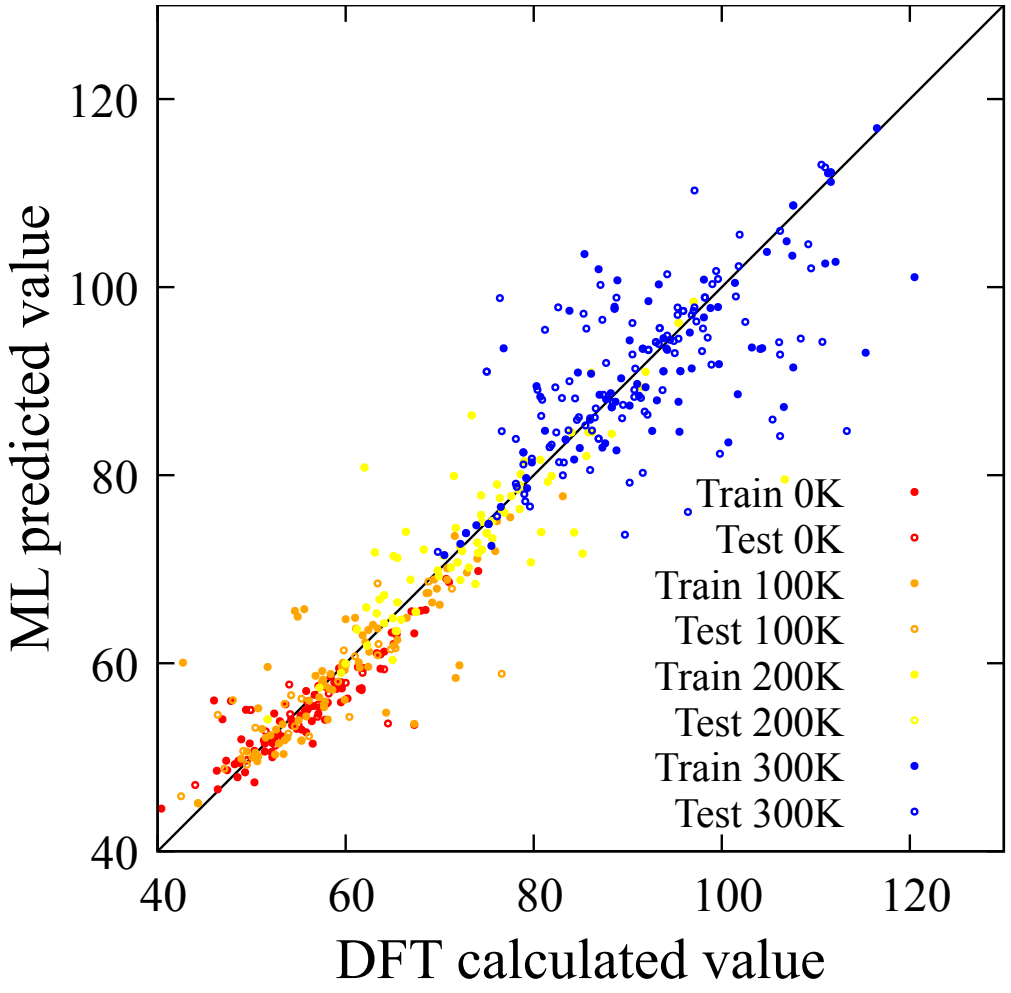

Figure 5.5: Scatter plot showing phonon-induced band gap correction (in meV units) predicted by ML model versus DFT calculation for $6 \times 6 \times 6$ Si-supercell at four different temperature values.

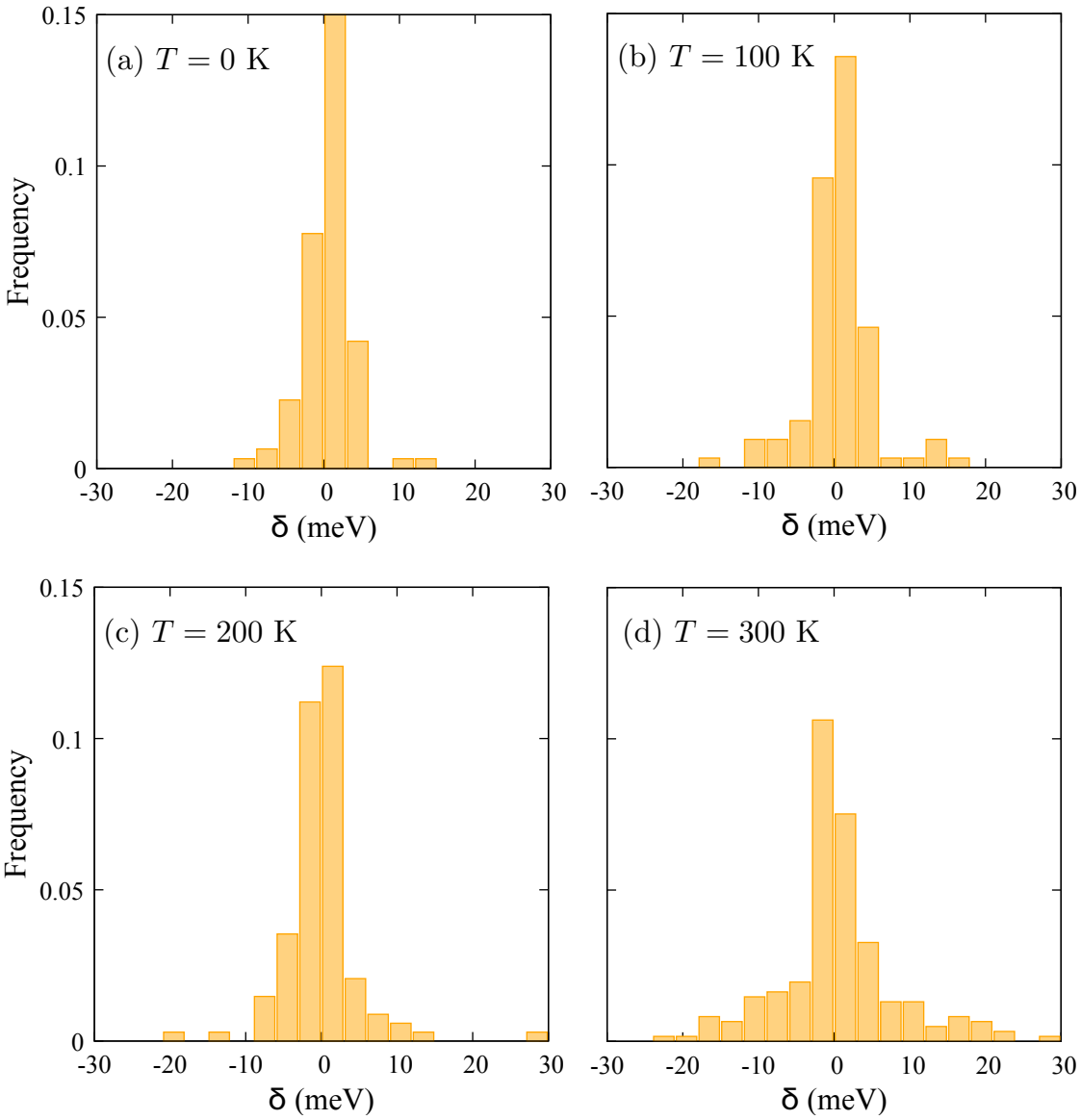

Figure 5.6: Histogram of the prediction error defined as $\delta = \Delta E_g^{\mathrm{ML}} - \Delta E_g^{\mathrm{DFT}}$ for temperatures at (a) $T = 0$ K, (b) 100 K, (c) 200 K, and (d) 300 K.

After NN training and optimization, we used the trained NN to predict band gap for a large number of MC generated supercell configurations at the four temperature values at which DFT calculations were performed to train the ML model. FIG. 5.7 shows the temperature dependence of the phonon-induced band gap correction of Si using DFT and NN prediction. For comparison, we also show the experimentally measured band gap correction. It should be noted that, although there are noticeable discrepancies between experiment and computational results for higher temperature region ($T \gtrsim 150$ K), the ML predictions agree quite well with the DFT calculations. Importantly, by using ML-predicted data set, we were able to reduce the errorbars at $T = 0, 100, 200$ and 300 K significantly compared to the DFT calculation.

It is remarkable that both DFT calculation and ML prediction correctly capture the temperature dependence of band gap, especially in the low temperature regime. The discrepancy at higher temperature is likely due to the anharmonic and lattice thermal expansion effects which are not incorporated in our calculations. Inclusion of these effects in *ab initio* calculations and application of more accurate exchange correlation functionals could yield better comparison with experiments [290, 291]. As discussed above, by training a ML model using more accurate *ab initio* data, our proposed framework can also incorporate such many-body effects into the electron-phonon couplings and the phonon-induced temperature dependence.

Moreover, since temperature serving as a conditional control is another input to our NN (see FIG. 5.1), we also apply the trained ML model to predict the band gaps

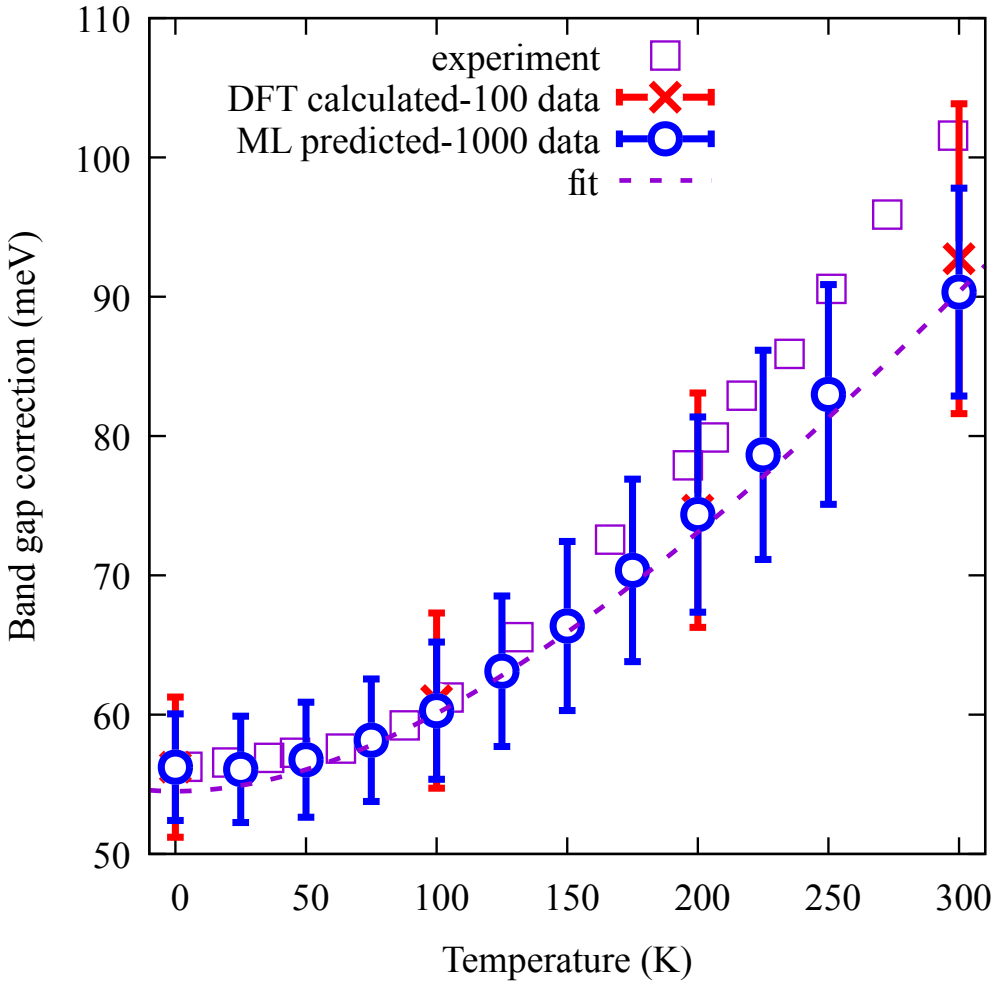

Figure 5.7: Temperature dependence of the phonon-induced band-gap correction ΔE_g of Si. In total, 528 DFT calculations on 432-atom Si supercell were done. ML prediction was done on 1000 configurations at each temperature. Experimental data is extracted from Ref. [1]. The experimental data is shifted by DFT calculated zero-point band gap correction energy.

at intermediate temperature values (see FIG. 5.7). The inclusion of temperature as a control parameter to the ML model, which was trained by datasets from multiple temperatures, could impose further constraints by enforcing a consistency between predictions at different temperatures. We have also trained ML models only trained by dataset from a particular temperature and obtained better prediction accuracies at that temperature. However, such models not only lack the generality for other temperatures, but also likely suffer from over-fitting. On the other hand, the ML models with additional conditioning from the temperature is shown to produce consistent trend of temperature dependent band gap energy. The prediction accuracies at the intermediate temperatures are of the same order as those used for the training. We note that the overall prediction accuracy can be improved by increasing the number of training data points or the scale of the neural network. Our results show that, by leveraging the power of transfer learning, band gap corrections at intermediate temperatures (where training data are not available) can also be accurately predicted by our trained ML model without incurring any significant computational cost. The only cost is in generating supercell configurations which is nominal.

5.5 Conclusion and Outlook

In summary, we have proposed a ML-based *ab inito* framework to incorporate phonon-induced renormalization of electronic band structures and other electronic properties. Our ML framework is based on the well-established stochastic approach to temperature dependent band structures and transition processes due to electron-phonon

coupling. In this approach, frozen phonon configurations within a supercell are first sampled from Monte Carlo or dynamical simulations based on an *ab initio* phonon model. Supervised learning methods are then applied to build a ML model that accurately maps the sampled frozen phonon configuration within a supercell to the physical properties of interest. With the efficient sampling of phonon configurations and prediction of the corresponding physical properties, accurate vibrational temperature dependence can be obtained based on moderate datasets of *ab initio* calculations.

A crucial component of our ML framework is a phonon descriptor for incorporating the point-group symmetry of the crystalline system into the ML model. Contrary to atomic descriptors used in most ML-based *ab initio* molecular dynamics methods, where the objects of interest are the atomic species and relative coordinates to a central atom, the dynamical variables in our case are atomic displacement vectors (from their equilibrium position in a perfect crystal). To properly account for these differences, we have employed the group-theoretical methods to obtain feature variables for the input to the neural network which is used as the learning model in our framework.

In this work, we apply the proposed ML framework to obtain the temperature dependence of phonon-induced corrections to the electronic band gap in crystalline silicon. We demonstrate that the prediction errors of the energy gaps can be significantly reduced from less than 100 datasets thereby reducing the number of expensive DFT computations necessary for better estimation. Moreover, applying the trained

ML model at temperatures not included in the training dataset, we show that accurate and consistent trend of the band-gap corrections are obtained, highlighting the power of transfer learning. While our work focuses on the prediction of band gap energy, the proposed ML framework is of general purpose and can be used for thermal vibrational effects on other electronic properties or transition coefficients as long as Born-Oppenheimer approximation is valid.

While the ML model developed in this work is based on a specific flavor of DFT (local density approximation for electron and phonon calculations) and density-functional perturbation theory-based *ab-initio* phonon model, as discussed above, the frozen phonon approach is independent of the underlying electronic structure methods. This means that many-body effects beyond DFT can be readily included in this approach. However, the *ab initio* phonon model used to generate the atomic configurations should be consistent with the first-principle method employed for training the ML model.

Chapter 6

Nonequilibrium Dynamics in Charge Density Wave Insulators

6.1 introduction

Complex adaptive materials with multiple resistant states are widely exploited as building blocks for next generation electronic and information technology. The operation of conventional metal-oxide semiconductor field-effect transistors (MOSFET) relies on the electrostatic charging of free electrons in thermal equilibrium, leading to a switching speed bounded by 60 mV per decade, also known as the Boltzmann limit. In contrast, controllable switching between many-body states with different resistances offers the possibility of high-speed and low-power transistor operations. In particular, it has been suggested that ultrafast switching can be achieved by the collective response of electrons in an insulator-to-metal transition (IMT) [292–294]. Indeed, an emerging technology called the "Motronics" aims at achieving electronic and logical operations by controlling electron correlation in the Mott insulators [295–297]. The high-resistant state of these materials results from the localization of electrons due

to the short-range Coulomb repulsion. The Mott insulator can be induced to a low-resistant state via a first-order transition by reducing the electron repulsive energy through electrostatic doping.

A charge density wave (CDW) state [298–300] is another many-body state that attracts enormous attention due to their potential electronic and optoelectronic applications. A CDW phase is a macroscopic quantum state consisting of a periodic modulation of the electronic charge density accompanied by a periodic structural distortion. Early interests in CDW focused on the sliding phase of the quasi-one-dimensional incommensurate states with nonlinear conductivity at low applied voltage [299]. Other novel properties include giant dielectric response, multi-stable conducting states, and proximity to unusual superconductivity. The interest in CDW is rekindled in recently years due to the advent of quasi-2D van der Waals materials, where CDW phases can manifest themselves at room temperatures and above [301–309]. Moreover, as these CDW phases can be readily manipulated by temperature, strain, bias voltage, and other stimuli, these quasi-2D CDW materials are emerging as a new paradigm for multifunctional devices.

In particular, electrically induced transitions between different CDW phases, as well as CDW-metal transitions have been demonstrated in the 1T polymorph of tantalum disulfide (TaS_2) [310–315], with a timescale of a few nanoseconds, making this material a promising candidate for high-speed energy-efficient electronic applications [315–317]. The voltage or electric field induced IMT or resistive transition is

known to be a highly complex process which involves a large variety of microscopic mechanisms including carrier dynamics, lattice distortions, heat transport, and electron correlation effect. The electrical switching from CDW to metallic state is no exception. On the other hand, recent experiments indicated that this process is driven mainly by motion of complex domain walls [311] and is qualitatively different from the resistive switching in complex oxides or chalcogenide glass which are characterized by highly inhomogeneous intermediate states [318–324].

Despite its fundamental importance and technological implications, theoretical modeling of the electrically induced CDW transitions remains mostly at the phenomenological level. This is understandably a difficult task due to the multi-scale nature of the problem. On one hand, one needs to describe the dynamical evolution of the lattice degrees of freedom, while accounting for the nonequilibrium nature of the electron system. On the other hand, large-scale real-space simulation is required in order to capture the transient pattern formation and domain-wall propagation. Finally, for systems with strong electron correlation, the need of proper many-body techniques further adds the computational complexity.

In this chapter, we present the first-ever large-scale dynamical simulations of the CDW phase transition, focusing on the interplay of lattice dynamics and the influence of the out-of-equilibrium electrons. We consider the adiabatic limit of the Holstein model in which the lattice degrees of freedom can be treated as classical variables, and the model can then be exactly solved numerically. By performing large-scale

quantum Brownian dynamics simulation with the forces computed from the nonequilibrium Green's function method, we demonstrate a reversible gating-induced CDW-to-metal transition that is driven by the domain-wall motion. We further obtain a dynamical phase diagram and characterize the voltage and temperature dependence of the domain-wall dynamics.

6.2 Gating Induced Charge Density Wave to Metal Transition

We consider a capacitor structure with a square-lattice Holstein model sandwiched by two electrodes shown in Fig. 6.1(a). The right electrode serves as the substrate, while a gate voltage V is applied at the left electrode. The total Hamiltonian of the system is $\mathcal{H} = \mathcal{H}_{\text{Hols}} + \mathcal{H}_{\text{res}}$, where $\mathcal{H}_{\text{Hols}}$ is the Holstein Hamiltonian for the square-lattice in the center, and $\mathcal{H}_{\text{res}}$ describes the electrodes, reservoir degrees of freedom. The Holstein Hamiltonian reads [325]

$$\hat{\mathcal{H}}_{\text{Hols}} = -t_{\text{nn}} \sum_{\langle ij \rangle} \hat{c}_i^\dagger \hat{c}_j - g \sum_i Q_i \left(\hat{n}_i - \frac{1}{2} \right) \tag{6.1}$$

where $\hat{c}_i / \hat{c}_i^\dagger$ is the annihilation/creation operators of spin-less electron at site-i, and $\hat{n}_i \equiv \hat{c}_i^\dagger \hat{c}_i$ is the corresponding number operator, Q_i describes a local structural distortion at site-i, such as the breathing mode of the oxygen octahedron. The first-term describes nearest-neighbor hopping t_{nn} of electrons, and the second term denotes phonon-electron interaction with a coupling constant g. As discussed above, here we treat the Holstein phonon Q_i as classical variables with the following elastic en-

ergy [325]

$$\mathcal{V} = \frac{k}{2}\sum_i Q_i^2 + \sum_{\langle ij\rangle} \kappa\, Q_i Q_j\,, \tag{6.2}$$

where k is the effective on-site spring constant, and κ describes a nearest-neighbor repulsion between the local distortions. The Holstein model is one of the most studied electron-phonon model systems, and is widely used to investigate the physics of polarons, superconductivity, and CDWs [26–34].

In the adiabatic limit, similar to the Born-Oppenheimer approximation widely used in quantum or *ab initio* molecular dynamics [87], the electron relaxation is assumed to be much faster than the lattice dynamics. To this end, we employ an over-damped Langevin dynamics, or the Brownian dynamics (BD) to model the time evolution of the lattice distortions [34, 326, 327]

$$\frac{dQ_i}{dt} = -\frac{1}{\gamma}F_i + \zeta_i(t), \tag{6.3}$$

where γ is an effective friction constant, and $\eta_i(t)$ denotes a stochastic force described by a delta-correlated stationary Gaussian process. The driving force of a conservative system is given by the derivative of a potential energy: $F = -\partial E/\partial Q_i$. For electron forces in thermal equilibrium, the effective energy is given by $E = \langle \hat{\mathcal{H}}_{\text{Hols}}\rangle = \text{Tr}\big(\hat{\rho}_{\text{eq}}\hat{\mathcal{H}}_{\text{Hols}}\big)$. However, for a driven electronic system, the energy is not well defined, and the force in general is non-conservative. Nonetheless, the force can still be

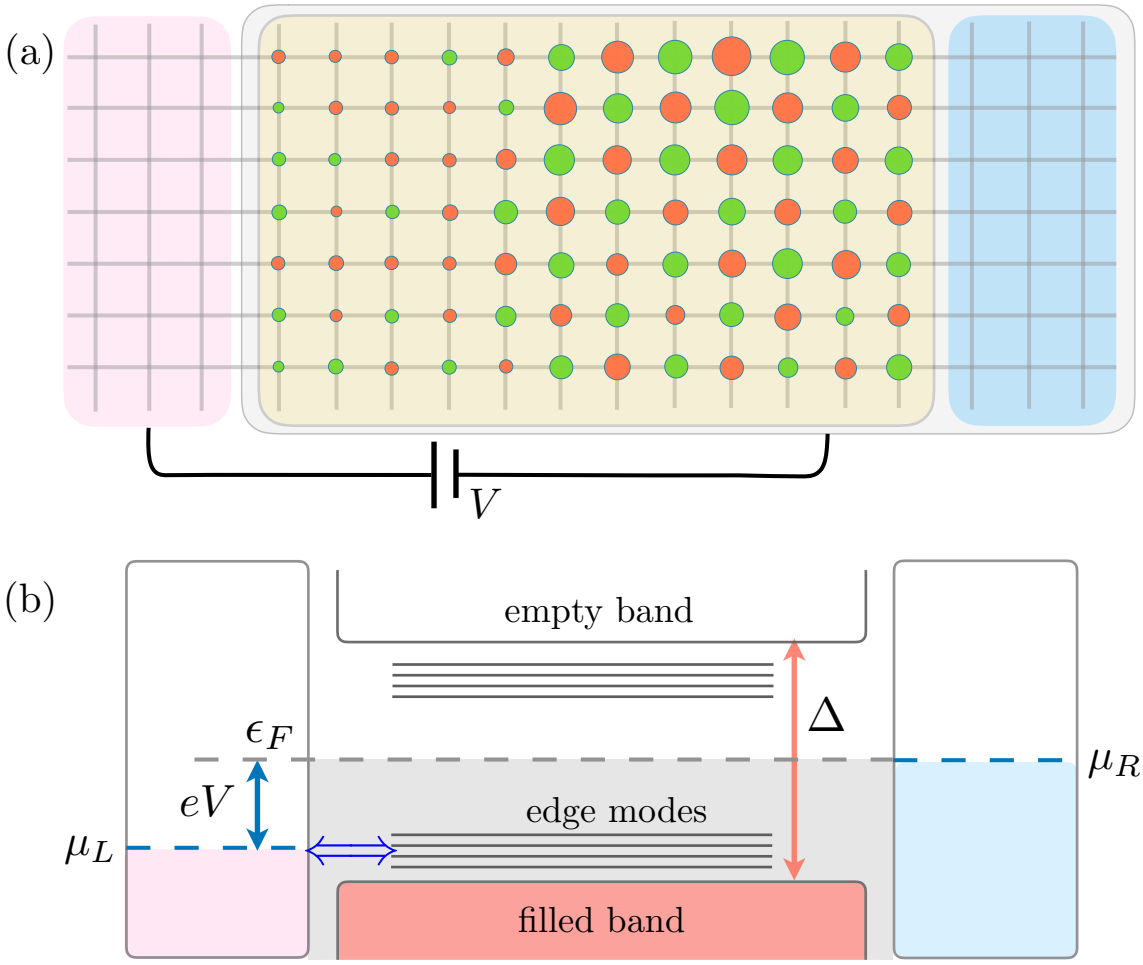

Figure 6.1: (a) Schematic diagram of the Gating-induced CDW to metal transition in the adiabatic Holstein model. (b) Energy diagram of the two electrodes and the center Holstein model. The CDW state is a band insulator with an energy gap $E_g = 2gQ_0$. Both the Fermi level ϵ_F at the bulk and the chemical potential μ_R of the right electrode are set at the middle of the energy gap, while the chemical potential of the left-electrode is lowered from the Fermi level by the gating voltage: $\mu_L = \epsilon_F - eV$.

obtained from a generalized Hellmann-Feynman theorem [328–332]

$$F_i = -\left\langle \frac{\partial \hat{\mathcal{H}}_{\text{Hols}}}{\partial Q_i} \right\rangle = g\langle \hat{n}_i \rangle \tag{6.4}$$

where the expectation value $\langle \cdots \rangle$ is evaluated with respect to the quasi-steady-state, yet out-of-equilibrium, electron subsystem.

6.3 Nonequilibrium Green's Function

We next employ the nonequilibrium Green's function (NEGF) method [333–336] to compute these expectation values. To this end, we consider the following explicit Hamiltonian for the electrodes and reservoir

$$\mathcal{H}_{\text{res}} = \sum_{\xi,i} \varepsilon_\xi \, d^\dagger_{i,\xi} d_{i,\xi} - \sum_{i,\xi} V_{\xi,i} \big(d^\dagger_{i,\xi} c_i + \text{h.c.} \big). \tag{6.5}$$

Here $d_{i,\xi}$ represents non-interacting fermions from the bath (i inside the bulk) or the leads (for i on the two open boundaries), and ξ is a continuous quantum number. For example, ξ can be used to represent the band-structure of the two leads. After integrating out the reservoir fermions in both leads and bath, the retarded Green's function matrix for the central region is given by $\mathbf{G}^r(\epsilon) = (\epsilon \mathbf{I} - \mathbf{H} - \mathbf{\Sigma}^r)^{-1}$, where $H_{ij} = t_{ij} - g\delta_{ij} Q_i$ is the tight-binding matrix of the Holstein model Eq. (6.1) and

$$\Sigma^r_{ij}(\epsilon) = \delta_{ij} \sum_\xi |V_{i,\xi}|^2 / (\epsilon - \epsilon_\xi + i0^+) \tag{6.6}$$

is the dissipation-induced self-energy. The resultant level-broadening matrix given by $\mathbf{\Gamma} = i(\mathbf{\Sigma}^r - \mathbf{\Sigma}^a)$ is diagonal with $\Gamma_{ii} = \pi \sum_\xi |V_{i,\xi}|^2 \delta(\epsilon - \epsilon_\xi)$. For simplicity, we assume

flat wide-band spectrum for the reservoirs, which leads to a frequency-independent broadening factor. Next, using the Keldysh formula, the lesser Green's function is obtained from the retarded/advanced Green's functions: $\mathbf{G}^<(\epsilon) = \mathbf{G}^r(\epsilon)\mathbf{\Sigma}^<(\epsilon)\mathbf{G}^a(\epsilon)$, and the lesser self-energy is related to the $\Sigma^{r/a}$ through dissipation-fluctuation theorem: $\Sigma^<_{ij}(\epsilon) = 2i\,\delta_{ij}\,\Gamma_i\,f_{\rm FD}(\epsilon-\mu_i)$. Here $\Gamma_i = \Gamma_{\rm lead}$ or $\Gamma_{\rm bath}$ depending on whether site-i is at the boundaries or in the bulk. The local chemical potential $\mu_i = \epsilon_F$ for the bath, and $\mu_i = \mu_{L/R}$ for the two electrodes. Finally, the nonequilibrium force in Eq. (6.4) is proportional to the on-site electron density $\langle\hat{n}_i\rangle = \langle c_i^\dagger(t)c_j(t)\rangle$, which is the diagonal element of the equal-time lesser Green's function

$$F_i(t) = g\,G^<_{ii}(t,t) = g\int_{-\infty}^{+\infty} G^<_{ii}(\epsilon;t)\,d\epsilon. \tag{6.7}$$

In our implementation, the integral is replaced by a Riemann summation with an $\Delta\epsilon = 0.003$, i.e. with up to 4000 energy values in the summation.

6.4 Results

We apply the NEGF-BD method outline above to simulate the gating-induced IMT of the Holstein model on a 40×30 lattice. The initial state of the simulations was obtained first using the equilibrium Brownian dynamics simulations with half-filled electrons $\overline{n} \equiv \sum_i \hat{n}_i/N = 0.5$ per site. This results in an initial state with CDW order on the square lattice; see Fig. 6.2(a). The band structure a perfect CDW order is given by $E_\pm(\mathbf{k}) = \pm\sqrt{\epsilon^2_{\mathbf{k}} + g^2\,Q_0^2}$, where $\epsilon_{\mathbf{k}}$ is the energy dispersion of the square-lattice tight-binding Hamiltonian, and Q_0 is the amplitude of the lattice distortion in

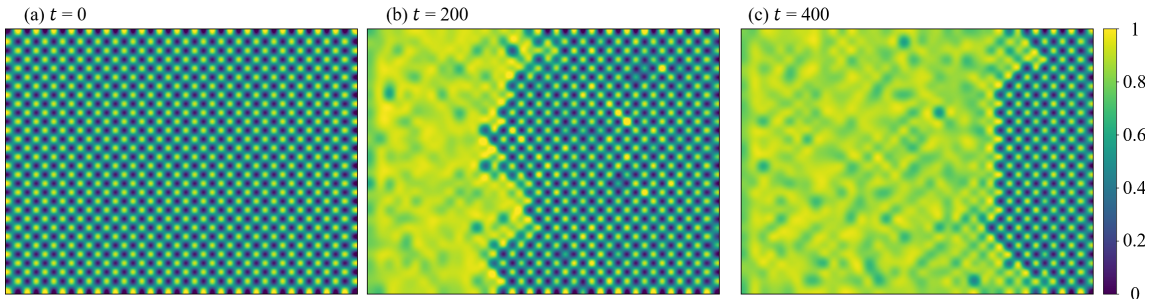

Figure 6.2: NEGF-BD simulations of a driven Holstein model on a 40×30 square lattice. The driven voltage is applied along the longitudinal x direction. Panels (a)–(c) show the snapshots of the system local electron density $n_i = \langle \hat{c}_i^\dagger \hat{c}_i \rangle$ at various simulation times. The following parameters are used: electron-phonon coupling constant $g = 1.5t_{\rm nn}$, the effective spring constant $k_1 = 0.67g$, nearest-neighbor elastic coupling $\kappa = 0.12g$, damping constant $\lambda = 0.2g$, $\Gamma_{\rm lead} = 1.0$, $\Gamma_{\rm bath} = 0.001$, $k_B T = 0.1$, and the bias voltage $eV = 2.5t_{\rm nn}$.

the CDW state. Importantly, an energy gap $\Delta = 2g\,Q_0$ is opened in the spectrum. At half-filling, the valance band $E_-(\mathbf{k})$ is completely filled, and the Holstein model is in a band-insulator state.

Next we turn on the voltage bias V. As illustrated in Fig. 6.1(b), the chemical potential of both the substrate (the right electrode) and the center square lattice is kept at the zero $\epsilon_F = \mu_R = 0$, which is set at the middle of the energy gap, while that of the left electrode is $\mu_L = -eV$. Fig. 6.2 shows an example of the phase transformation of the Holstein system driven by a gating voltage $eV = 2.5$. The color gradient shows the expectation value of the on-site density $\langle \hat{n}_i \rangle$. Other simulation parameters are: electron-phonon coupling constant $g = 1.5t_{\rm nn}$, local displacement Q_i string constant $k_1 = 0.67g$, nearest-neighbor coefficient $\kappa = 0.12g$, damping constant $\lambda = 0.2g$, $\Gamma_{\rm lead} = 1.0$, $\Gamma_{\rm bath} = 0.001$, and the temperature is $k_B T = 0.1$.

As shown in Fig. 6.2, the gating voltage induces an instability of the CDW, which

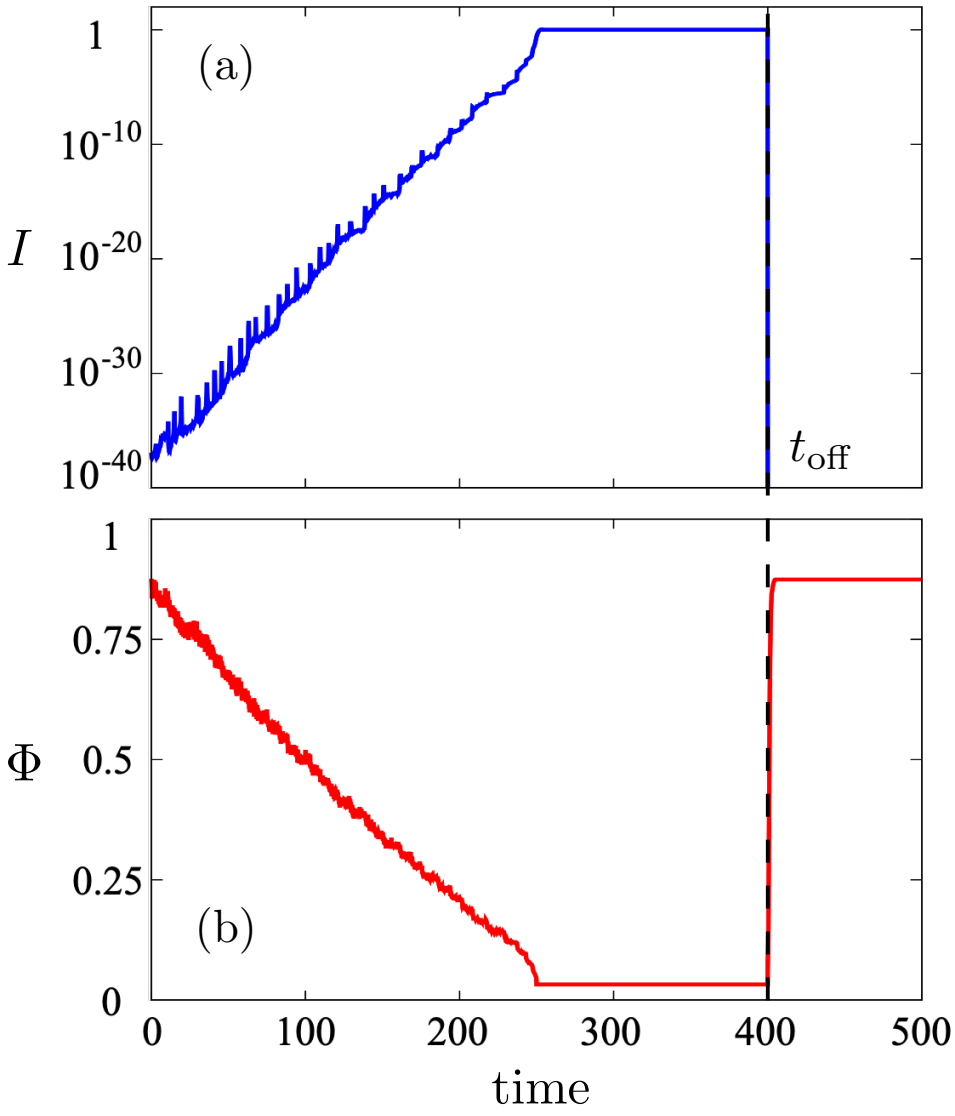

Figure 6.3: Time dependence of (a) the transmission current I and (b) the CDW order parameter Φ at zero temperature. A constant voltage $eV = 2.5t_{\rm nn}$ is applied to the Holstein system during the interval $0 \leq t \leq t_{\rm off} = 400$.

results in the formation of a metallic domain as the electrons are drained from the gating electrode with a lower chemical potential. The resultant CDW-metallic domain wall is then driven to the substrate. To further characterize the phase transformation process, we compute the transmission current of the driven electron state [333–336]

$$I = \int \mathrm{Tr}(\mathbf{\Gamma}_R \, \mathbf{G}^r \, \mathbf{\Gamma}_L \, \mathbf{G}^a)[f_L(\epsilon) - f_R(\epsilon)]d\epsilon, \tag{6.8}$$

where $\mathbf{\Gamma}_{L,R}$ are the diagonal broadening matrices, and $f_{L,R}(\epsilon) = f_{\mathrm{FD}}(\epsilon - \mu_{L,R})$ are the Fermi-Dirac functions. Fig. 6.3(a) shows the transmission current I as a function of time under a bias voltage $eV = 2.5$ at a temperature $T \to 0$. The roughly linear segment in the semi-log plot indicates an exponential growth of the current: $I \sim \exp(c\,t)$, where c is a numerical constant. We note that the phase transformation at such low temperatures is essentially a downhill process, with energy dissipated away both through the effective damping γ of the lattice, and the electron reservoir. Interestingly, as the CDW-metal interface moves across the sample one layer at a time, this discrete procession gives rise to the small spikes on top of the exponential increase in the figure. Fig. 6.3(b) shows the time dependence of the CDW order parameter, which is defined as

$$\Phi = \sum_i \langle \hat{n}_i \rangle \exp(i\mathbf{Q} \cdot \mathbf{r}_i), \tag{6.9}$$

where $\mathbf{Q}_i = (\pi/a, \pi/a)$ is the wave vector characterizing the checkerboard pattern on the square lattice. The almost linear decrease of Φ during the phase transformation indicates an almost linear propagation of the CDW-metal domain wall.

Our simulations also show that the metallic state is stable only in the presence of the bias voltage, which means that the insulator-metal transition is reversible. This is demonstrated in Fig. 6.3 which shows that the CDW order returns to its initial value when the gating voltage is turned off. Interestingly, contrary to the voltage-driven transition which proceeds through intermediate states with propagating domain-wall, the recovery to the CDW state is almost immediate, indicating a bulk instability. It is worth noting that the fact that the CDW order resumes to its initial maximum value in our simulation is a finite-size effect. For large lattices, the reverse transition is likely to result in the formation of multiple CDW domains of opposite signs.

We next investigate the effects of the gate voltage on the phase transformation rate. Fig. 6.4(a) shows the average position ξ of the CDW-metal interface as a function of time at different bias voltage. As expected, a larger gate voltage results in faster domain-wall motion. The average velocity of the interface, obtained through a linear fitting of the domain-wall trajectory $\xi(t)$, is plotted in Fig. 6.4(b) as a function of the bias voltage. Importantly, one can identify three different dynamical regimes from this result. For gate voltage smaller than a threshold $eV_{\rm th} \sim 2.4$, the CDW remains a stable state. In regime-II, a voltage-induced instability results in the formation of CDW-metal interface whose propagation velocity increases with larger voltage. And finally, for V greater than a second critical value, the transition to the metallic phase takes place instantly through a process similar to a dielectric breakdown.

This dynamical phase diagram can be understand from the energy diagram shown

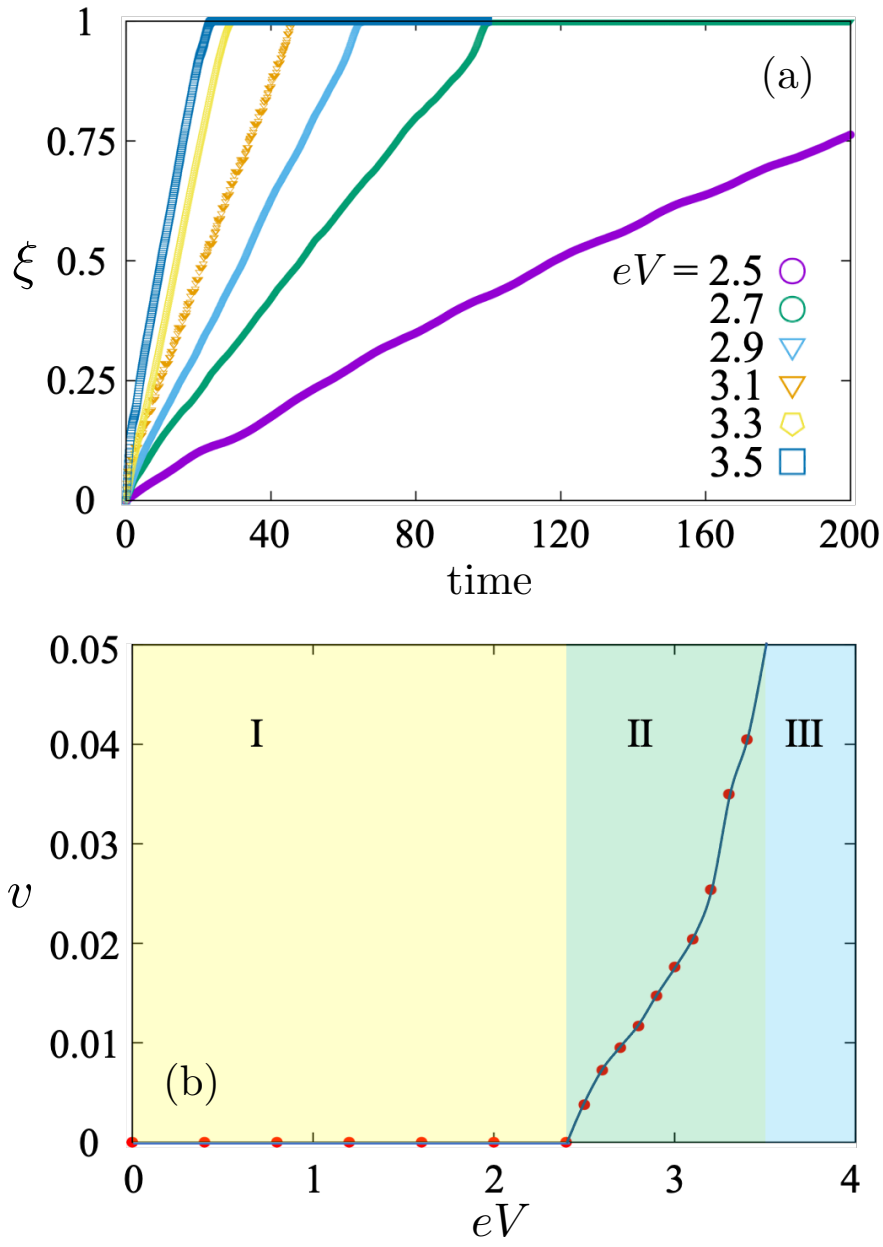

Figure 6.4: (a) average position ξ of the CDW-metal domain-wall as a function of time for different gating voltages at zero temperature. (b) CDW-metal domain-wall velocity v as a function of the gate voltage eV. The velocity is computed from the average slope of the $\xi(t)$ curve in panel (a).

in Fig. 6.1(b), which also provides an explanation of the phase transition mechanisms. For a small bias voltage the chemical potential of the gating electrode μ_L lies in the energy gap of the CDW state and, as a result, is ineffective in inducing the phase transition. As the voltage is increased to a point that μ_L overlaps with the energy levels of the in-gap states that are localized at the left edge, the resonant coupling leads to an instability of the CDW. As electrons are drained from the gate due to a smaller chemical potential, a metallic layer is nucleated near the left edge. The phase transformation then proceeds through the expansion of the metallic domain. This corresponds to the dynamical regime-II in Fig. 6.4(b). As the applied voltage is further increased such that the chemical potential μ_L is lower than the valance band-edge, the gate electrode now directly couples to the energy states in the bulk, thus allowing efficient current flow through the bulk and an instant instability toward the metallic phase.

The domain-wall propagation at $T \to 0$ discussed above is governed by a relaxation dynamics dominated by the dissipation of energy. It is worth noting that the effective potential of the local lattice distortion is similar to a double-well potential, with one minimum at $Q_i = \pm Q_0$ (depending on the sublattice) and the other one at $Q_i \sim 0$ corresponding to the metallic phase. The energy surface of the whole system is more complicated than this simple picture. The layer-by-layer progression the metallic domain indicates that each domain-wall position is a quasi-stable state in the high dimensional configuration space. As more energy is drained from the gate, the decay

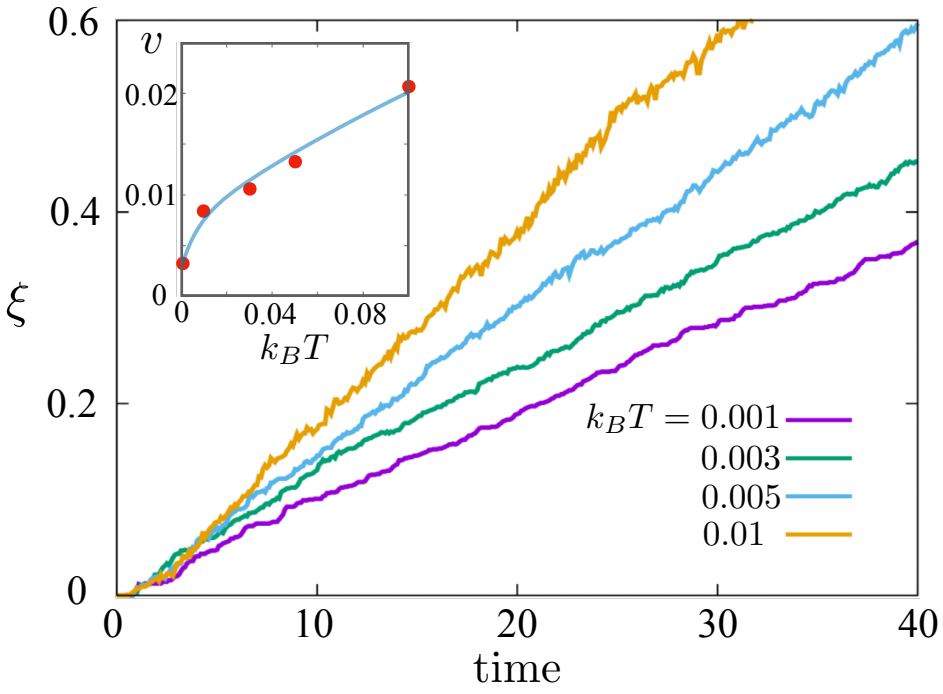

Figure 6.5: Average position ξ of the CDW-metal interface as a function of time for different temperatures with constant driven voltage $eV = 2.5t_{\rm nn}$. The inset shows the CDW-metal domain-wall velocity as a function of temperature.

of a local minimum leads to the advance of the domain-wall by one layer. In the presence of thermal fluctuations, the system can jump from one local minimum to the next one through thermal tunneling, instead of waiting for the decay of the present minimum. It is thus expected that the domain-wall mobility is enhanced at finite temperatures. This is indeed confirmed in our simulations, as shown in Fig. 6.5. Except for the point at very low temperature, we find an approximate linear increase of the domain-wall velocity with temperature.

6.5 Conclusion and Outlook

To sum up, we've executed extensive NEGF-BD simulations to investigate the gating-induced CDW to metal transition in a well-studied electron-phonon system. Besides the familiar electrical breakdown seen at high bias voltages, we reveal a shift marked by meta-stable states featuring moving metal-insulator domain-walls at moderate voltages. This shift originates from an initial CDW instability when a metallic layer forms at the gate electrode. The resulting metal-CDW boundary moves through the system at a nearly steady speed that rises with both voltage bias and temperature. Furthermore, this voltage-driven transition is reversible; the system rapidly returns to the CDW state once the voltage gate is removed.

Our work focuses on the interplay between atomic lattice dynamics and the out-of-equilibrium electrons. Importantly, our study uncovers the instability mechanisms for the CDW states. Nonetheless, the CDW phase transition in real materials, such as the quasi-2D van der Waals materials, is a highly complex process which is beyond

the present work. In particular, by focusing on the solvable Holstein model in the adiabatic limit, our simulations neglect the electron correlation effect, which is likely to be important in compounds such as 1T-TaS_2 that is intensively studied experimentally. Many-body techniques such as Hartree-Fock or Gutzwiller are required to properly model the collective electron behaviors in CDW described by, e.g. the Holstein-Hubbard model [337, 338]. For example, self-consistent Hartree-Fock equation combined with NEGF was used to investigate the complex spin-density wave patterns in voltage-driven Hubbard model [126, 339, 340]. However, a full dynamical modeling of CDW transitions requires further integration with real-space lattice dynamics, which is computationally highly demanding. Machine-learning methods could provide a promising solution for the multi-scale modeling of the complex CDW dynamics.

Chapter 7

Descriptors for Lattice Models

7.1 Introduction

ML is increasingly becoming a powerful paradigm in scientific research and engineering [43–56], particularly in quantum chemistry and materials science [218, 219, 341–347]. In recent years, ML methods have been employed to significantly accelerate computational tasks that are otherwise highly time-consuming, such as first-principles electronic structure calculations based on DFT. Among various ML approaches, deep NNs [348, 349] have proven to be especially versatile and effective, capable of approximating complex high-dimensional functions with remarkable accuracy [280, 281, 350]. This advancement has enabled the development of ML-based interatomic potentials, allowing large-scale MD simulations with the precision of DFT, significantly expanding the scope of computational modeling in materials science [91–97, 125–130].

Traditionally, large-scale simulations of magnetic systems or phase transitions have relied on empirical or effective energy models, such as classical spin Hamiltonians or Ginzburg-Landau energy functionals [70, 72]. While these models incorporate key

symmetries of the original quantum systems, they inherently lack predictive power and fail to account for the intricate interplay between electrons and classical fields during dynamic evolution. In contrast, a quantum approach would require solving the electron Hamiltonian at every time step, analogous to quantum MD simulations where atomic forces are obtained by solving the self-consistent Kohn-Sham equation at each step [87, 351]. However, the computational expense associated with repeated electronic structure calculations severely limits the applicability of this approach to large systems.

Modern ML methods offer a promising solution to this challenge, as demonstrated by recent studies [98, 99] employing ML-based models to simulate quantum LLG dynamics and other complex phenomena in correlated electron systems. These studies have showcased the potential of deep-learning NNs to enable large-scale simulations of systems with strong electron-electron interactions, such as double-exchange models [204, 352–354] or Falicov-Kimball models [59]. In these cases, ML models can emulate the behavior of the electronic subsystem [57, 76] and provide accurate predictions for effective forces or torques acting on classical fields, thereby allowing the exploration of much larger system sizes and longer timescales than would be feasible with conventional methods.

Despite the versatility of ML models, ensuring that the symmetries of the underlying quantum system are preserved remains a crucial challenge. To accurately simulate the dynamics of classical fields coupled to electrons, a proper represen-

tation of the field configuration must be constructed to serve as input to the ML model. This representation, often referred to as a descriptor [104, 182, 184], should be invariant under symmetry transformations of both the classical fields and the lattice [41, 91–93, 104, 104, 105, 175, 181, 181–185, 341, 355, 356]. Similar challenges have been encountered in the context of ML interatomic potentials, where atomic descriptors must be invariant under rotational and permutational symmetries. Over the years, a variety of descriptors have been developed, such as ACSF [91, 181], bond-order parameters [355], and group-theoretical methods based on bispectrum coefficients [92, 104].

In this chapter, we develop a general theory of descriptors for classical fields in condensed matter systems, with a particular focus on lattice models. We explore several approaches, including those generalized from atomic descriptors, to construct faithful representations of local classical fields. The primary focus, however, is on the group-theoretical method, which provides a rigorous representation based on irreducible representations (IRs) of the lattice point group. While this approach is powerful, it is often over-complete and cumbersome to implement in practice. To address these challenges, we propose the concept of reference IRs, which significantly simplify the implementation of bispectrum descriptors. The proposed descriptors are then applied to various lattice models, demonstrating their ability to capture the symmetries and complexity of different classical field configurations.

7.2 Descriptor

The ML energy model provides an effective energy model in terms of the classical fields

$$E\big(\{\mathbf{\Phi}_i\}\big) = \sum_i \varepsilon(\mathcal{C}_i) \tag{7.1}$$

where $\{\mathbf{\Phi}_i\}$ denote a configuration of the emergent classical fields, $\mathcal{C}_i$, defined as $\mathcal{C}_i = \big\{\mathbf{\Phi}_j \,\big|\, R_{ij} = |\mathbf{r}_j - \mathbf{r}_i| \leq R_c\big\}$, denotes the local configuration of the classical variables with a given cutoff radius R_c, and ϵ denotes the local site-energy. This energy model naturally needs to preserve the symmetry of the original Hamiltonian, which includes both the symmetry of the dynamical variables and that of the underlying lattice. Nonetheless, despite the universal approximation capability of ML models, the symmetry of the original electron Hamiltonian is not automatically captured. Since the training of ML model is essentially an optimization process with randomly chosen datasets, the symmetry of the model can only be statistically approximated even with a large amount of training data. As discussed above, a proper representation of the classical fields is required to ensure that the symmetry of the electron Hamiltonian is built into the ML model. A good representation, or descriptor, of the local environment must be invariant with respect to symmetry transformations of the system.

For condensed matter systems defined on a lattice, the ML energy model $\varepsilon(\mathcal{C}_i)$ must be invariant under the discrete transformations of the point group, denoted as

G_L, associated with the center site-i. Moreover, for classical fields with a complex structure, one also needs to take into account the symmetry associated with transformations among the multiple components $\Phi_{i,1}, \Phi_{i,2}, \cdots$ at the same site. We classify the classical fields into two types depending on whether these two symmetries are entangled to each other or not. Examples of these two types are illustrated in Fig. 7.1. For models of the first type, the internal symmetry of the classical variables is coupled to the lattice symmetry. Examples of type-I classical variables include on-site displacement vector fields $\mathbf{u}_i = (u_i^x, u_i^y, u_i^z)$ [357–359], where the transformation of the 3 components of the displacement vector is coupled to the discrete rotations of the point group, such as Chap. 5. Another example is the Jahn-Teller (JT) doublet $\mathbf{Q}_i = (Q_i^{x^2-y^2}, Q_i^{3z^2-r^2})$ characterizing local structural distortion [360, 361]. The only relevant symmetry group for such type-I models is the on-site point group G_L. Under the symmetry operation $\hat{g} \in G_L$, the rearrangement of the classical fields at different lattice sites coincides with the transformation of the various components. Noting that $\Phi_{j,\alpha} = \Phi_\alpha(\mathbf{r}_j)$ with $\alpha = 1, 2, \cdots, M$ being the index of the various components, the transformation of the classical fields is described by

$$\tilde{\Phi}_\alpha\big(O(\hat{g}) \cdot \mathbf{R}_{ij}\big) = \mathcal{M}_{\alpha\beta}(\hat{g})\Phi_\beta(\mathbf{R}_{ij}), \tag{7.2}$$

where $\mathbf{R}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ is the relative position vector of site-j, $\mathcal{M}_{\alpha\beta}(g)$ is the M-dimensional matrix representation of the symmetry operation $\hat{g}$, and $O(\hat{g})$ is the 3-dimensional orthogonal matrix transforming site-i to site-k, i.e. $\mathbf{R}_{ik} = O(\hat{g}) \cdot \mathbf{R}_{ij}$.

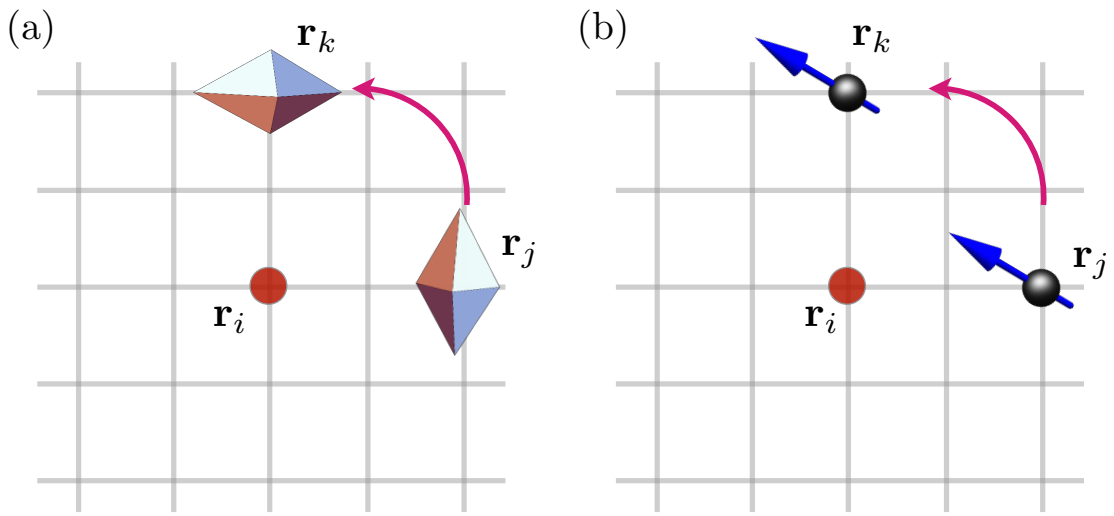

Figure 7.1: Classical fields with different symmetry properties. An example of type-I case is the local JT distortion $\mathbf{Q}_j$ as shown in panel (a). The lattice rotation/reflection is accompanied by a simultaneous transformation of the JT phonons. Panel (b) shows the type-II case exemplified by local spins $\mathbf{S}_j$. The global rotation symmetry of spins is independent of the discrete point-group symmetry of the lattice.

For type-II models, the classical fields are characterized by an independent internal symmetry group, which will be denoted as G_Φ. The most representative example, perhaps, is the models with local classical spins $\mathbf{S}_i$ as illustrated in Fig. 7.1(b). For spins with n-component, the symmetry group of the system is a direct product of the lattice group G_L and the internal symmetry group $G_\Phi = O(n)$ describing the global rotation symmetry of the spins. The most general symmetry operation consists of the lattice rotation/reflection $\hat{g} \in G_L$, and the transformation $\hat{h} \in G_\Phi$:

$$\tilde{\Phi}_\alpha\big(O(\hat{g}) \cdot \mathbf{R}_{ij}\big) = \mathcal{M}_{\alpha\beta}(\hat{h})\Phi_\beta(\mathbf{R}_{ij}), \tag{7.3}$$

Note that $\mathcal{M}_{\alpha\beta}(\hat{h})$ is the matrix representation of the group G_Φ, which is independent of the lattice point group. It is important to note that, for type-II models, the ML potential energy $\varepsilon(\mathcal{C}_i)$ must be invariant under the general combined symmetry transformation $\hat{h} \otimes \hat{g}$.

It is worth noting that while our focus is on the lattice models which are prevalent in condensed matter physics, most of the analysis presented in this work can be generalized to the off-lattice models or disordered systems if the point group G_L is replaced by the continuous 3-dimensional rotation group $O(3)$, of course, assuming the system possesses such a global rotation symmetry. This means that our analysis can also be applied to electron models defined on an amorphous system or an atomic liquid state. In fact, the latter case can be viewed as a molecular dynamics system with an array of classical variables $\mathbf{\Phi}_i$ associated with every atom. A particular

interesting application would be the Gutzwiller MD method where the classical fields $\mathbf{\Phi}_i$ corresponds to the slave-boson amplitudes [362].

Having discussed the general symmetry transformations for the two types of classical fields, we next describe a concise vector representation of the neighborhood $\mathcal{C}_i$ with its center at site-i. We first consider the type-I models; the case of the type-II model will be discussed in Sec. 7.2.2. For convenience, the site-indices of lattice points within $\mathcal{C}_i$ are labeled as j_{r}, where $\mathsf{r} = 1, 2, \cdots, L = |\mathcal{C}_i|$. Essentially, the integer r offers an ordered list of lattice sites in the neighborhood. Under the symmetry operation $\hat{g}$ of the point group, the lattice point j_{r} is mapped to j_{s} if and only if $(\mathbf{r}_{j_{\mathsf{s}}} - \mathbf{r}_i) = O(\hat{g}) \cdot (\mathbf{r}_{j_{\mathsf{r}}} - \mathbf{r}_i)$. Consequently, $\hat{g}$ can be represented by a $L \times L$ permutation matrix $\mathcal{P}$, which means the nonzero matrix elements are $\mathcal{P}_{\mathsf{sr}}(\hat{g}) = 1$ if the two sites j_{r} and j_{s} are related by $\hat{g}$. Next we introduce a vector $\vec{\mathcal{U}}$ whose components are given by the classical fields in the neighborhood:

$$\mathcal{U}_{\mathsf{r},\alpha} = \Phi_\alpha(\mathbf{r}_{j_{\mathsf{r}}}), \tag{7.4}$$

It is easy to see that $\vec{\mathcal{U}}$ offers a vector representation of dimension $L \times M$ for the point group G_L. And the corresponding matrix representation $\mathcal{T}$ of the symmetry operation $\hat{g} \in G_L$ is given by

$$\tilde{\mathcal{U}}_{\mathsf{r},\alpha} = \mathcal{T}_{\mathsf{r}\alpha,\mathsf{s}\beta}(\hat{g})\, \mathcal{U}_{\mathsf{s},\beta} = \mathcal{P}_{\mathsf{rs}}(\hat{g}) \mathcal{M}_{\alpha\beta}(\hat{g})\, \mathcal{U}_{\mathsf{s},\beta}. \tag{7.5}$$

Also importantly, the matrix $\mathcal{T}$ provides an orthogonal matrix representation of the point group, i.e. $\mathcal{T}^\dagger \mathcal{T} = \mathcal{T}\mathcal{T}^\dagger = \mathbb{I}$, where $\mathbb{I}$ is the $L \times M$-dimensional identity

matrix. Next we present two descriptors based on this vector representation of the neighborhood.

7.2.1 Bispectrum Coefficients

In this Section, we present a systematic method for constructing a descriptor based on the group-theoretical method. Specifically, the feature variables are given by the so-called bispectrum coefficients computed from the expansion coefficients of irreducible representations of the point group [103, 363, 364]. The bispectrum coefficients, which are invariant under the symmetry operations of the point group, are in a sense similar to the scalar triple product of three vectors which is invariant under arbitrary rotations. It is also worth noting that similar group-theoretical methods, with important modifications to simplify the implementation, have been proposed as descriptor for ML interatomic potentials in quantum MD simulations [92, 104].

In order to compute the bispectrum coefficients, the first step is to obtain the irreducible representations of the neighborhood. As discussed above, the vector $\vec{\mathcal{U}}$, defined in Eq. (7.4) provides a $L \times M$-dimensional representation of the local environment $\mathcal{C}_i$. This high-dimensional representation can then be decomposed into irreducible representations (IRs) of the point group G_L following the standard procedures [102, 365]. Specifically, we use Γ to label the different IRs in the decomposition, and denote the corresponding basis vector of IR-Γ as

$$\vec{\boldsymbol{\Upsilon}}^{\Gamma} = \left(\vec{\Upsilon}^{\Gamma}_{1}, \vec{\Upsilon}^{\Gamma}_{2}, \cdots, \vec{\Upsilon}^{\Gamma}_{n_{\Gamma}}\right), \tag{7.6}$$

where n_Γ is the dimension of corresponding IR. Note that each "component" $\vec{\Upsilon}^\Gamma_\mu = \{\Upsilon^\Gamma_{\mu;\mathsf{r},\alpha}\}$ is itself a $(L \times M)$-dimensional vector. The neighborhood vector is then decomposed as

$$\mathcal{U}_{\mathsf{r},\alpha} = \sum_\Gamma \sum_{\mu=1}^{n_\Gamma} f^\Gamma_\mu \, \Upsilon^\Gamma_{\mu;\,\mathsf{r},\alpha}. \tag{7.7}$$

The expansion coefficients f^Γ_μ of the IR, play a role similar to the Fourier coefficients for the translation group. Using the orthogonality of the basis vectors of different IRs, the expansion coefficients are given by

$$f^\Gamma_\mu = \vec{\Upsilon}^{\Gamma\dagger}_\mu \cdot \vec{\mathcal{U}} = \sum_{\mathsf{r}=1}^{L} \sum_{\alpha=1}^{M} \Upsilon^{\Gamma *}_{\mu,\,\mathsf{r}\alpha} \, \mathcal{U}_{\mathsf{r}\alpha}. \tag{7.8}$$

For convenience, we can group the expansion coefficients of a given IR into a vector:

$$\boldsymbol{f}^\Gamma = \left(f^\Gamma_1, f^\Gamma_2, \cdots, f^\Gamma_{n_\Gamma}\right). \tag{7.9}$$

In terms of the classical fields, see e.g. Eq. (7.4), the expansion coefficients are

$$f^\Gamma_\mu = \sum_{\mathsf{r}=1}^{L} \sum_{\alpha=1}^{M} \Upsilon^{\Gamma *}_{\mu;\,\mathsf{r}\alpha} \, \Phi_\alpha(\mathbf{r}_{j_\mathsf{r}}). \tag{7.10}$$

Under symmetry operations $\hat{g}$ of the point group G_L, different IRs transform independently of each other. Consequently, the transformation of the vector $\boldsymbol{f}^\Gamma$ of a given IR is described by an $n_\Gamma \times n_\Gamma$ unitary matrix $\boldsymbol{D}^\Gamma$ as

$$\tilde{f}^\Gamma_\mu = \sum_{\mu'} D^\Gamma_{\mu\mu'}(\hat{g}) \, f^\Gamma_{\mu'}, \tag{7.11}$$

or the more concise vector equation: $\tilde{\boldsymbol{f}}^\Gamma = \boldsymbol{D}^\Gamma \cdot \boldsymbol{f}^\Gamma$. From the transformation relation

Eq. (7.5) for the vector $\vec{\mathcal{U}}$, the transformation matrix $\boldsymbol{D}^{\Gamma}$ can be explicitly computed.

$$D^{\Gamma}_{\mu\mu'}(\hat{g}) = \vec{\Upsilon}^{\Gamma\dagger}_{\mu} \cdot \mathcal{T}(\hat{g}) \cdot \vec{\Upsilon}^{\Gamma}_{\mu'}. \tag{7.12}$$

It is worth noting that the transformation matrices of a given IR have been tabulated for most point and double groups. Similar to the ordinary Fourier analysis, we define the power spectrum for a given IR as

$$p^{\Gamma} \equiv \boldsymbol{f}^{\Gamma\dagger} \cdot \boldsymbol{f}^{\Gamma} = \sum_{\mu=1}^{n_{\Gamma}} \left| f^{\Gamma}_{\mu} \right|^2 \tag{7.13}$$

Since the transformation matrices are unitary $\boldsymbol{D}^{\dagger}\boldsymbol{D} = 1$, it is easy to see that the power spectrum is invariant under symmetry operations:

$$\tilde{p}^{\Gamma} = \tilde{\boldsymbol{f}}^{\Gamma\dagger} \cdot \tilde{\boldsymbol{f}}^{\Gamma} = \boldsymbol{f}^{\Gamma\dagger}\boldsymbol{D}^{\Gamma\dagger}\boldsymbol{D}^{\Gamma}\boldsymbol{f}^{\Gamma} = \boldsymbol{f}^{\Gamma\dagger}\boldsymbol{f}^{\Gamma} = p^{\Gamma}, \tag{7.14}$$

This indicates that the amplitude of each IR can be used as the descriptor for the local environment $\mathcal{C}_i$. However, the power spectrum p^{Γ} is not a complete descriptor of the neighborhood function, since it neglects the weight distribution within each IR. Neither does it account for the relative phases between different IRs.

A more complete description, which consists of a larger set of invariants, is given by the bispectrum of the IRs. To this end, we first consider the tensor product of coefficient vectors $\boldsymbol{f}^{\Gamma_1} \otimes \boldsymbol{f}^{\Gamma_2}$, which can be viewed as the expansion coefficients of the tensor-product $\vec{\mathcal{U}} \otimes \vec{\mathcal{U}}$ of the vector representation with a tensor-product basis $\vec{\Upsilon}^{\Gamma_1}_{\mu} \otimes \vec{\Upsilon}^{\Gamma_2}_{\nu}$. Under a symmetry operation, according to Eq. (7.11), the tensor-product

transforms as

$$\begin{aligned} \boldsymbol{f}^{\Gamma_1} \otimes \boldsymbol{f}^{\Gamma_2} &\to \left(\boldsymbol{D}^{\Gamma_1} \cdot \boldsymbol{f}^{\Gamma_1}\right) \otimes \left(\boldsymbol{D}^{\Gamma_2} \cdot \boldsymbol{f}^{\Gamma_2}\right) \\ &= \left(\boldsymbol{D}^{\Gamma_1} \otimes \boldsymbol{D}^{\Gamma_2}\right) \cdot \left(\boldsymbol{f}^{\Gamma_1} \otimes \boldsymbol{f}^{\Gamma_2}\right). \end{aligned} \tag{7.15}$$

As is well established in the representation theory of finite groups, the direct product of two IRs can be decomposed into a direct sum of IRs. This indicates the following decomposition of the direct-product matrices:

$$\boldsymbol{D}^{\Gamma_1} \otimes \boldsymbol{D}^{\Gamma_2} = \left(\boldsymbol{C}^{\Gamma_1,\Gamma_2}\right)^{\dagger} \left[\bigoplus_{\Gamma} \boldsymbol{D}^{\Gamma}\right] \boldsymbol{C}^{\Gamma_1,\Gamma_2}, \tag{7.16}$$

where $\oplus$ means direct sum over the IRs of the direct product. We note that IR of the same dimension and symmetry could appear more than once in the direct sum. The $\boldsymbol{C}^{\Gamma_1,\Gamma_2}$ is a unitary matrix of dimension $n_{\Gamma_1} \times n_{\Gamma_2}$; its matrix elements are known as the Clebsch-Gordan coefficients of the symmetry group under consideration. Explicitly, we have

$$\begin{aligned} & D^{\Gamma_1}_{\mu\mu'}(\hat{g})\, D^{\Gamma_2}_{\nu\nu'}(\hat{g}) \\ &= \sum_{\Gamma} \sum_{\kappa,\kappa'} \left(C^{\Gamma;\Gamma_1,\Gamma_2}_{\kappa,\mu\nu}\right)^{*} D^{\Gamma}_{\kappa\kappa'}(\hat{g})\, C^{\Gamma;\Gamma_1,\Gamma_2}_{\kappa',\mu'\nu'} \end{aligned} \tag{7.17}$$

As mentioned above, the sum over Γ could include multiple IRs of the same transformation properties. To construct the bispectrum coefficients, we first consider the following vector:

$$\boldsymbol{v}^{\Gamma_1,\Gamma_2} = \boldsymbol{C}^{\Gamma_1,\Gamma_2} \cdot \left(\boldsymbol{f}^{\Gamma_1} \otimes \boldsymbol{f}^{\Gamma_2}\right) \tag{7.18}$$

Since the Clebsch-Gordan matrix is essentially a transformation of basis, vector $\boldsymbol{v}$ is thus the expansion coefficients of the irreducible basis for the tensor-product $\vec{\mathcal{U}} \otimes \vec{\mathcal{U}}$. This can also be seen from the transformation of the $\boldsymbol{v}$ vector. Substitute Eq. (7.16) into (7.15), and multiply the resultant expression by the $\boldsymbol{C}^{\Gamma_1,\Gamma_2}$ matrix from the left, we see that under symmetry operation $\hat{g}$, the vector $\boldsymbol{v}^{\Gamma_1,\Gamma_2}$ transforms according to

$$\tilde{\boldsymbol{v}}^{\Gamma_1,\Gamma_2} = \left[\bigoplus_{\Gamma} \boldsymbol{D}^{\Gamma}(\hat{g})\right] \cdot \boldsymbol{v}^{\Gamma_1,\Gamma_2} \tag{7.19}$$

This result thus also indicates we can decompose $\boldsymbol{v}$ into a direct sum of vectors each of which corresponds to an irreducible representation:

$$\boldsymbol{v}^{\Gamma_1,\Gamma_2} = \bigoplus_{\Gamma} \boldsymbol{u}^{\Gamma;\Gamma_1,\Gamma_2} \tag{7.20}$$

Each vector transforms under symmetry operation as

$$\tilde{\boldsymbol{u}}^{\Gamma;\Gamma_1,\Gamma_2} = \boldsymbol{D}^{\Gamma}(\hat{g}) \cdot \boldsymbol{u}^{\Gamma;\Gamma_1,\Gamma_2}. \tag{7.21}$$

From this equation and Eq. (7.11) for the transformation of the vector $\boldsymbol{f}$ belong to the same IR-Γ, it is straightforward to see that the following "inner product" is a scalar invariant under any symmetry operation:

$$b^{\Gamma,\Gamma_1,\Gamma_2} = \boldsymbol{f}^{\Gamma\,\dagger} \cdot \boldsymbol{u}^{\Gamma;\Gamma_1,\Gamma_2}, \tag{7.22}$$

These coefficients are called the bispectrum of the expansion coefficients of the IRs. Using Eq. (7.18) to express the $\boldsymbol{u}$ vectors, we obtain the following explicit formula

for the bispectrum coefficients

$$b^{\Gamma,\Gamma_1,\Gamma_2} = \sum_{\kappa,\mu,\nu} C^{\Gamma;\Gamma_1,\Gamma_2}_{\kappa,\mu\nu} f^{\Gamma *}_{\kappa} f^{\Gamma_1}_{\mu} f^{\Gamma_2}_{\nu}. \tag{7.23}$$

The above expression shows the similarity of the b coefficients with the scalar triple of three O(3) vectors. It should also be noted that the power spectrum p^{Γ} is part of the bispectrum coefficients. In fact, while formally the bispectrum coefficients are built from product of three IR-amplitudes, they can also be used to describe invariants consisting of two IR-coefficients. This corresponds to the case when the decomposition of the direct product representation $\Gamma_1 \otimes \Gamma_2$ includes the trivial one-dimensional representation, denoted as Γ_0 for convenience. By setting the corresponding coefficient to be a constant, e.g. $f_{\Gamma_0} = 1$, we see that the resultant bispectrum coefficient $b^{\Gamma_0,\Gamma_1,\Gamma_2}$ is nonzero only if the two IRs Γ_1 and Γ_2 transform in exactly the same way under symmetry operations, hence have the same dimension. Consequently, we can define the following generalization of power spectrum

$$p^{\Gamma_1,\Gamma_2} = \boldsymbol{f}^{\Gamma_1 \dagger} \cdot \boldsymbol{f}^{\Gamma_2} = \sum_{\mu} f^{\Gamma_1 *}_{\mu} f^{\Gamma_2}_{\mu}. \tag{7.24}$$

The standard power spectrum Eq. (7.13) of a given IR-Γ corresponds to the case $\Gamma_1 = \Gamma_2 = \Gamma$.

Importantly, since the bispectrum coefficients are invariant under symmetry operations of the point group, they serve as proper descriptor to be combined with the ML models. Moreover, it can be shown that the bispectrum provides a faithful representation of the original configuration in the sense that the vector $\mathcal{U}$ can be

rigorously reconstructed from all bispectrum coefficients [103, 363, 364]. For practical applications, however, there are a large number of the bispectrum coefficients for most models and point groups. For example, let $\mathbb{N}$ be the number of IRs from the decomposition of $\vec{\mathcal{U}}$, which is roughly of the order of $\mathbb{N} \sim (L \times M)$, the number of bispectrum is of the order of $\mathbb{N}^3$, which in general is a rather large number. Moreover, as will be demonstrated in explicit examples in Chap. 5, the bispectrum is an over-complete representation with redundant information. Consequently, further simplification is often required for practical implementations.

As an application of the bispectrum method, here we briefly review its application to represent the atomic environment for ML interatomic potentials. The bispectrum method is often combined with the Gaussian kernel potential learning model and the so-called smooth overlap of atomic positions (SOAP) technique, which approximates atoms in the neighborhood by Gaussian functions of a finite width [92, 104]. For MD simulations, the local atomic configuration is described by the charge density $\rho(\mathbf{r})$ with the origin $\mathbf{r} = 0$ corresponding to the center atom. The symmetry group of three-dimensional free space is $G_L = \mathrm{SO}(3)$, and the corresponding irreducible representations are labeled by an integer $\ell = 0, 1, 2, \cdots$, which is essentially the angular momentum quantum numbers [366]. Indeed, the basis function Υ^{Γ}_{μ} for the SO(3) group is simply the spherical harmonics $Y_{\ell,m}$. Choosing a proper radial basis $g_n(r)$, the atomic neighborhood density is expanded as

$$\rho(\mathbf{r}) = \sum_{n=0}^{\infty} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} f_{n\ell m}\, g_n(r) Y_{\ell m}(\theta, \phi), \tag{7.25}$$

Note that there is an additional integer index n for the expansion coefficients due to the radial dependence. The bispectrum coefficients are then labeled by six integers [104]:

$$b^{\ell;\ell_1,\ell_2}_{n;n_1,n_2} = \sum_{m,m_1,m_2} f^*_{n\ell m} C^{\ell;\ell_1,\ell_2}_{m;m_1,m_2} f_{n_1\ell_1 m_1} f_{n_2\ell_2 m_2}, \tag{7.26}$$

where $C^{\ell;\ell_1,\ell_2}_{m;m_1,m_2}$ are Clebsch-Gordan coefficients of the SO(3) group [366]. For a given set of radial indices (n, n_1, n_2), the bispectrum coefficients are nonzero only when $\ell = \ell_1 + \ell_2$ due to conservation of angular momentum. However, there are still an infinite number of the b coefficients, and some cutoff $\ell_{\max}$ has to be introduced for practical implementation. To further simplify the calculation, one can consider only coefficients with $n_1 = n_2 = n$. This, however, implies that rotations of different radial basis are decoupled, thus introducing a spurious symmetry. Nonetheless, some simplifications can be achieved through special designs of the radial basis functions [104].

Instead of dealing with the natural SO(3) group for the three-dimensional space, an alternative approach is to project the atomic environment within a cutoff R_c onto the surface of the four-dimensional sphere S^3 [92, 104]. Specifically, this means that the center-atom is at the north pole, while the cutoff radius, i.e. the 3-sphere specified by $|\mathbf{r}| = R_c$, is mapped to the south pole of the S^3. Next assuming an approximate SO(4) symmetry for the projected atomic density, one can then use the resultant bispectrum coefficients as the descriptor. As the IR of the SO(4) group is again labeled by an integer j, the bispectrum coefficients are indexed by three integers

b^{j,j_1,j_2}. It should be noted that although the projection to S^3 implicitly assumes a spurious SO(4) symmetry, a most crucial advantage of this approach is the absence of the need for radial basis.

7.2.2 Internal Symmetry

As discussed above, the type-II models are characterized by an internal symmetry group G_Φ, independent of the lattice point group, that governs the transformation of the classical fields $\mathbf{\Phi}_i$. The feature variables for the ML models need to be invariant with respect to transformations of both symmetry groups. As the multiple components of the local classical vector $\mathbf{\Phi}_i$ do not transform simultaneously with the lattice symmetry operations, the method described in Sec. 7.2.1 cannot be directly applied to the type-II models.

One solution is to treat each of the M components of the classical fields $\mathbf{\Phi}_i = \{\Phi_{i,\alpha}\}$ ($\alpha = 1, 2, \cdots, M$) as independent. We then view the neighborhood configuration $\vec{\mathcal{U}}_\alpha = (\mathcal{U}_{1,\alpha}, \mathcal{U}_{2,\alpha}, \cdots, \mathcal{U}_{L,\alpha})$ as M independent L-dimensional representations of the neighborhood $\mathcal{C}_i$. Each component is then decomposed into the IRs of the lattice group (c.f. Eq. (7.7) for the type-I case)

$$\mathcal{U}_{\mathsf{r},\alpha} = \sum_\Gamma \sum_{\mu=1}^{n_\Gamma} f^\Gamma_{\mu,\alpha}\, \Upsilon^\Gamma_{\mu;\mathsf{r}}, \tag{7.27}$$

Note the basis function Υ of the IR now only depends on the site-index r. The coefficients of the IRs are similarly obtained based on the orthogonality of the basis

functions

$$f^{\Gamma}_{\mu,\alpha} = \sum_{\mathsf{r}=1}^{L} \Upsilon^{\Gamma *}_{\mu;\mathsf{r}} \mathcal{U}_{\mathsf{r},\alpha} = \sum_{\mathsf{r}=1}^{L} \Upsilon^{\Gamma *}_{\mu;\mathsf{r}} \Phi_{\alpha}(\mathbf{r}_{j_{\mathsf{r}}}). \tag{7.28}$$

For each of the IR Γ in the decomposition (with respect to point group), there are M components indexed by α. As each can be viewed as a M-dimensional representation of the internal symmetry group, it can be decomposed into the IR of G_{Φ} labeled by K:

$$f^{\Gamma}_{\mu,\alpha} = \sum_{\mathsf{K}} \sum_{\mathsf{m}=1}^{n_{\mathsf{K}}} F^{\Gamma,\mathsf{K}}_{\mu,\mathsf{m}} \mathcal{Y}^{\mathsf{K}}_{\mathsf{m},\alpha}. \tag{7.29}$$

Here $\mathcal{Y}^{\mathsf{K}}_{\mathsf{m}}$ is the basis function of the K-th IR whose dimension is n_{K}. Using the orthogonality of the basis functions, the expansion coefficients are given by

$$F^{\Gamma,\mathsf{K}}_{\mu,\mathsf{m}} = \sum_{\alpha=1}^{M} \mathcal{Y}^{\mathsf{K}*}_{\mathsf{m},\alpha} f^{\Gamma}_{\mu,\alpha} = \sum_{\alpha=1}^{M} \sum_{\mathsf{r}=1}^{L} \mathcal{Y}^{\mathsf{K}*}_{\mathsf{m},\alpha} \Upsilon^{\Gamma *}_{\mu;\mathsf{r}} \Phi_{\alpha}(\mathbf{r}_{j_{\mathsf{r}}}). \tag{7.30}$$

Here we have used Eq. (7.28) in the second equality to express $F^{\Gamma,\mathsf{K}}_{\mu,\mathsf{m}}$ in terms of the classical fields. It is worth noting that this mixed expansion coefficients, expressed as a special combination of the classical fields, have well defined transformation properties, indicated by the IR indices Γ and K, under both the point group of the site-symmetry and the internal symmetry group. However, since the two set of symmetry transformations are independent of each other, one cannot obtain simultaneous bispectrum coefficients with respect to both symmetry groups. To proceed, we can first "trace out" the point group indices μ by forming the bispectrum coefficients of the point

group first

$$B^{\Gamma,\Gamma_1,\Gamma_2}_{\mathsf{K}_1,\mathsf{l};\mathsf{K}_2,\mathsf{m};\mathsf{K}_3,\mathsf{n}} = \sum_{\kappa,\mu,\nu} C^{\Gamma;\Gamma_1,\Gamma_2}_{\kappa,\mu\nu} F^{\Gamma,\mathsf{K}_1 *}_{\kappa,\mathsf{l}} F^{\Gamma_1,\mathsf{K}_2}_{\mu,\mathsf{m}} F^{\Gamma_2,\mathsf{K}_3}_{\nu,\mathsf{n}}. \tag{7.31}$$

These coefficients with three indices l, m, n can be viewed as a tensor-product representation $\mathsf{K} \otimes \mathsf{K}_1 \otimes \mathsf{K}_2$ of the internal symmetry group G_Φ. Next we decompose this tensor-product representation into a direct sum of IRs of the group G_Φ. For convenience of the discussion, we denote the coefficients of the IR in the direct sum as $\mathsf{F}^{\mathsf{K}}_{\mathsf{q}}$. Then invariants with respect to the internal symmetry are given by the bispectrum coefficients from the "triple" product of these $\mathsf{F}^{\mathsf{K}}_{\mathsf{q}}$ coefficients. Importantly, these bispectrum coefficients are now invariant with respect to both the lattice and internal symmetry groups. Since the $\mathsf{F}^{\mathsf{K}}_{\mathsf{q}}$ coefficients themselves are already triple product of the field variables, the final invariants in general are composed of 9 classical variables; although some of them can be reduced. Since the number of the coefficients increases even more dramatically with the cutoff radius R_c for type-II models, further approximations are necessary to simplify the implementation of the descriptor.

A second approach, which is physically more intuitive and transparent, is to start from the symmetry of the classical fields and first construct building blocks that are already invariant under the transformations of the internal symmetry group. The group-theoretical method discussed in Sec. 7.2.1 is then applied to these building blocks for the lattice symmetry. To this end, we again note that the classical fields $\mathbf{\Phi}_j = \{\Phi_{j,\alpha}\}$ at every sites in the neighborhood $\mathcal{C}_i$ is obviously an M-dimensional

representation of the internal symmetry group, and can be decomposed into IRs of the G_Φ group:

$$\Phi_{j,\alpha} = \sum_{\mathsf{K}} \sum_{\mathsf{m}=1}^{n_{\mathsf{K}}} \mathsf{f}^{\mathsf{K}}_{j,\mathsf{m}}\, \mathcal{Y}^{\mathsf{K}}_{\mathsf{m},\,\alpha}, \tag{7.32}$$

It is worth noting that the expansion coefficients $\mathsf{f}^{\mathsf{K}}_{j,\mathsf{m}}$ acquires a site index j. Again, using the orthogonality of $\mathcal{Y}$, we have

$$\mathsf{f}^{\mathsf{K}}_{j,\mathsf{m}} = \sum_{\alpha=1}^{M} \mathcal{Y}^{\mathsf{K}\,*}_{\mathsf{m},\,\alpha}\, \Phi_{j,\alpha}. \tag{7.33}$$

If the decomposition in Eq. (7.32) includes the trivial representation K_0 which is by definition a one-dimensional IR, then the coefficients $\mathsf{f}^{\mathsf{K}_0}_j$ are automatically invariant with respect to the internal symmetry group and are part of the building blocks for the lattice group.

Other invariants of the internal symmetry group are provided by the generalized power spectrum Eq. (7.24) and the bispectrum coefficients. The crucial difference here is that these invariants are to be built from different sites, thus also serving as many-body correlation functions. First, we consider the generalized power spectrum obtained from a pair of sites (jk)

$$\mathsf{p}^{\mathsf{K}_1,\mathsf{K}_2}_{jk} = \sum_{\mathsf{m}} \mathsf{f}^{\mathsf{K}_1\,*}_{j,\mathsf{m}}\, \mathsf{f}^{\mathsf{K}_2}_{k,\mathsf{m}}, \tag{7.34}$$

Again, the generalized power spectrum coefficient is nonzero only if the two IRs K_1 and K_2 have the same transformation properties. Similarly, one can build invariants of internal symmetry from a triplet (jkl) of lattice sites based on the bispectrum

coefficients

$$\mathtt{b}_{jkl}^{\mathsf{K},\mathsf{K}_1,\mathsf{K}_2} = \sum_{\mathsf{l},\mathsf{m},\mathsf{n}} \mathtt{C}_{\mathsf{l};\mathsf{m},\mathsf{n}}^{\mathsf{K};\mathsf{K}_1,\mathsf{K}_2}\, \mathtt{f}_{j,\mathsf{l}}^{\mathsf{K}\,*} \mathtt{f}_{k,\mathsf{m}}^{\mathsf{K}_1} \mathtt{f}_{l,\mathsf{n}}^{\mathsf{K}_2}. \tag{7.35}$$

where $\mathtt{C}_{\mathsf{l},\,\mathsf{m},\mathsf{n}}^{\mathsf{K};\mathsf{K}_1,\mathsf{K}_2}$ are the Clebsch-Gordan coefficients of the internal symmetry group G_Φ. Fig. 7.2 shows examples of the atomic pairs (jk) and triplets (jkl) related by the lattice rotation and reflection in the neighborhood of the center site. As mentioned above, these quantities $\mathtt{p}$ and $\mathtt{b}$ also encode the two-body and three-body correlations, respectively, of the neighborhood. Also importantly, they remain unchanged under operations of the internal symmetry group and can be used as building blocks for constructing the invariants of the lattice point group. To this end, we arrange them, including the single-site trivial IR, into a vector of dimension $\mathcal{N}$:

$$\vec{\mathcal{U}} = (\mathcal{U}_1, \mathcal{U}_2, \cdots, \mathcal{U}_{\mathcal{N}}) = \left(\mathtt{f}_j^{\mathsf{K}_0},\, \mathtt{p}_{jk}^{\mathsf{K}_1,\mathsf{K}_2},\, \mathtt{b}_{jkl}^{\mathsf{K},\mathsf{K}_1,\mathsf{K}_2}\right). \tag{7.36}$$

Here we use $\mathcal{U}_J$ to denote the components of this vector, where the index J is used to label either a site j, a pair (jk), or a triplet (jkl). The dimension $\mathcal{N}$ is dominated by the number of atomic pairs and triplets in the neighborhood. For a neighborhood consisting of L sites, these two number scale as L^2 and L^3, respectively. Moreover, one also needs to take into account the number of different IRs. As the total classical degrees of freedom is $L \times M$, the set of all $\mathtt{f}$, $\mathtt{p}$, and $\mathtt{b}$ invariants obviously is an over-complete representation of the neighborhood. Practically, one needs to introduce further constraints in order to reduce this number, for example, by restricting distances between the pairs or triplets to be smaller than another cutoff, or to avoid

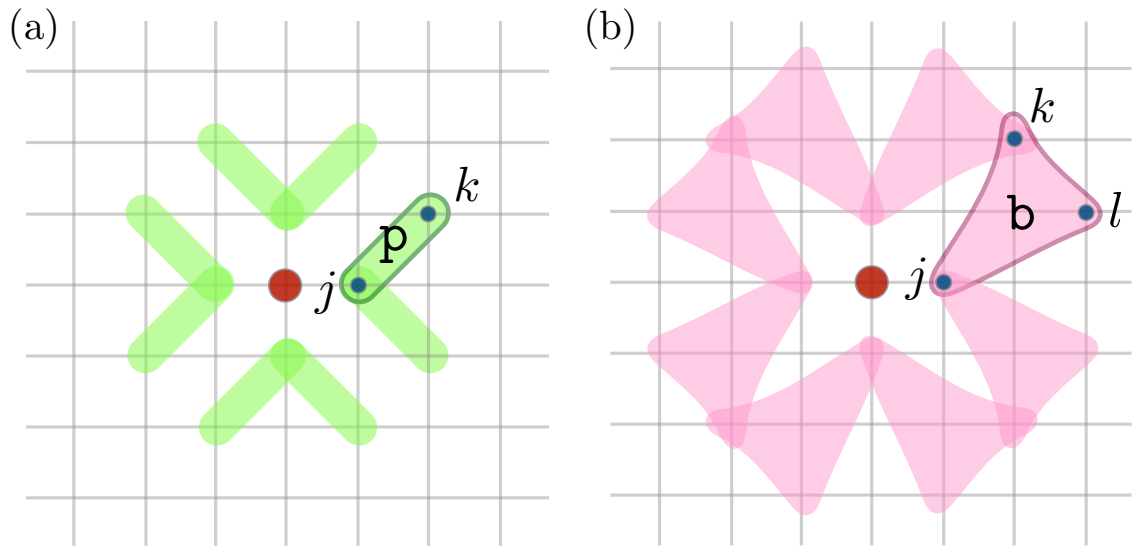

Figure 7.2: Examples showing the atomic pair (jk) and triplet (jkl), which are related by the lattice rotation and reflection symmetries, in the neighborhood of the center site on a square lattice.

too many overlaps of the pairs and triples.

Irrespective of the approximations, by keeping all symmetry related pairs and triples, as shown in Fig. 7.2, in Eq. (7.36), the vector $\vec{\mathcal{U}}$ forms an $\mathcal{N}$-dimensional representation of the lattice point group G_L. We next apply the same group-theoretical method discussed in Sec. 7.2.1 to obtain the bispectrum coefficients of the point group. We again decompose $\vec{\mathcal{U}}$ into the IRs

$$\mathcal{U}_J = \sum_{\Gamma} \sum_{\mu=1}^{n_\Gamma} f_\mu^\Gamma \Upsilon_{\mu;J}^\Gamma. \tag{7.37}$$

where $\Upsilon_{\mu;J}^\Gamma$ are the appropriate basis functions. It is worth noting that the IRs of the single sites, pairs, and triplets are decoupled from each other. The expansion

coefficients are then obtained separately as

$$f_\mu^\Gamma = \begin{cases} \sum_j \ \Upsilon_{\mu;j}^{\Gamma *}\, \mathtt{f}_j^{\mathsf{K}_0} \\ \sum_{(jk)} \ \Upsilon_{\mu;jk}^{\Gamma *}\, \mathtt{p}_{jk}^{\mathsf{K}_1,\mathsf{K}_2} \\ \sum_{(jkl)} \Upsilon_{\mu;jkl}^{\Gamma *}\, \mathtt{b}_{jkl}^{\mathsf{K},\mathsf{K}_1,\mathsf{K}_2} \end{cases} \tag{7.38}$$

Given these IR coefficients, Eqs. (7.23) and (7.24) can then be used to compute the generalized power spectrum and bispectrum coefficients, respectively, which are invariant with respect to both the internal and the lattice symmetry groups of the type-II systems.

7.2.3 Atom-Centered Symmetry Functions

The building blocks introduced in Eqs. (7.34) and (7.35) above also offer the basis for a descriptor which can be viewed as the generalization of the atom-centered symmetry function (ACSF) originally proposed to describe the atomic configurations [91, 127]. Unlike the group-theoretic methods, the ACSF approach is physically more intuitive and relatively simple to implement. On the other hand, it is more difficult to control the errors due to the *ad hoc* parameterizations of the symmetry functions. Nonetheless, ACSF has been successfully applied to the ML interatomic potential for a wide range of materials. We first briefly review the basic features of ACSF using the example of mono-atomic systems. For a given atomic configuration $\{\mathbf{r}_j\}$ in the vicinity of a center atom-i, the fundamental invariants that are invariant under rotations and

reflections of the O(3) group are the distances $R_{ij} = |\mathbf{r}_j - \mathbf{r}_i|$ from the center atom, and the angles $\theta_{ijk} = \arccos[(\mathbf{r}_j - \mathbf{r}_i)\cdot(\mathbf{r}_k - \mathbf{r}_i)/R_{ij}R_{ik}]$. Based on these quantities, two kinds of symmetry functions are introduced. The first type is the two-body (between atoms j and the center atom-i) symmetry function

$$G_2(\{\xi_m\}) = \sum_{j\neq i} F_2(R_{ij}; \{\xi_m\}), \tag{7.39}$$

where $F_2(R; \xi_m)$ is a user-defined function, parameterized by $\{\xi_m\}$ to extract atomic structures at certain distances from the center atom. One popular choice, proposed in the original work [91], is a Gaussian with a soft cutoff at radius R_c

$$F_2(R; \{\xi_m\}) = e^{-(R-\xi_1)^2/\xi_2^2}\, f_c(R). \tag{7.40}$$

Here $f_c(r) = \frac{1}{2}\big[\cos(\frac{\pi r}{R_c}) + 1\big]$ for $R \leq R_c$ and zero otherwise. The two parameters ξ_1 and ξ_2 speficiy the center and width, respectively, of the Gaussian function. The 3-body symmetry functions are defined as

$$G_3(\{\xi_m\}) \;=\; \sum_{j,k\neq i} F_3(R_{ij}, R_{ik}, R_{jk}, \theta_{ijk}; \{\xi_m\}), \tag{7.41}$$

An example of the three-body envelop function characterized by three parameters is [91, 127]

$$\begin{aligned} F_3(R_1, R_2, R_3, \theta; \{\xi_m\}) &= 2^{1-\xi_1}(1 + \xi_2 \cos\theta)^{\xi_1} \\ &\times \exp\big[-(R_1^2 + R_1^2 + R_3^2)/\xi_3^2\big] f_c(R_1) f_c(R_2) f_c(R_3). \end{aligned} \tag{7.42}$$

We note that generalizations to take into account the different atom species have also been made [184]. Moreover, depending on the problems at hand, it might be more convenient to use different F_2 and F_3 functions, and several variants of these functions have been proposed [184].

Next we present a generalization of the ACSF for condensed-matter systems, where each atom is now associated with a dynamical classical field $\mathbf{\Phi}_j$. We emphasize that the formulation presented here can also be used for disordered systems, where the "lattice" point group is replaced by the 3D rotation group SO(3). Moreover, for applications to MD simulation of liquid systems with a dynamical classical fields, the generalized ACSF provides a convenient descriptor for ML energy models for both the atomic dynamics and the classical fields. In order to incorporate the internal symmetry, our approach is to define a set of symmetry functions based on the building blocks in Eq. (7.36). We start with the two-body symmetry functions that include the coefficients of the trivial IR at every sites:

$$G_{2a}(\{\xi_m\}) = \sum_{j\neq i} \mathtt{f}_j^{\mathsf{K}_0} F_2(R_{ij};\{\xi_m\}), \tag{7.43}$$

This is the direct generalization of the original two-body symmetry functions that incorporates the on-site classical fields. Another way to build the 2-body symmetry functions is to use the invariants $\mathtt{p}_{ij}^{\mathsf{K}_1,\mathsf{K}_2}$ between the center site-i and a neighboring site-j:

$$G_{2b}^{\mathsf{K}_1,\mathsf{K}_2}(\{\xi_m\}) = \sum_{j\neq i} \mathtt{p}_{jk}^{\mathsf{K}_1,\mathsf{K}_2} F_2'(R_{ij};\{\xi_m\}), \tag{7.44}$$

The envelope function $F_2'(R)$ is not necessarily the same as the one for G_{2a}. A three-body symmetry function based on single-site invariants is

$$G_{3a}^{\mathsf{K}_1,\mathsf{K}_2}(\{\xi_m\}) = \sum_{jk\neq i} \mathtt{f}_j^{\mathsf{K}_1}\mathtt{f}_k^{\mathsf{K}_2} \times F_3(R_{ij}, R_{ik}, R_{jk}, \theta_{ijk}; \{\xi_m\}), \tag{7.45}$$

The pair-wise invariants can also be combined with the center atom to define a three-body symmetry function:

$$G_{3b}^{\mathsf{K}_1,\mathsf{K}_2}(\{\xi_m\}) = \sum_{jk\neq i} \mathtt{p}_{jk}^{\mathsf{K}_1,\mathsf{K}_2} \times F_3'(R_{ij}, R_{ik}, R_{jk}, \theta_{ijk}; \{\xi_m\}), \tag{7.46}$$

A second type of 3-body symmetry functions is obtained from the invariants $\mathtt{b}_{ijk}^{\mathsf{K},\mathsf{K}_1,\mathsf{K}_2}$ that involves the center atom

$$G_{3c}^{\mathsf{K},\mathsf{K}_1,\mathsf{K}_2}(\{\xi_m\}) = \sum_{jk\neq i} \mathtt{b}_{ijk}^{\mathsf{K},\mathsf{K}_1,\mathsf{K}_2} \times F_3''(R_{ij}, R_{ik}, R_{jk}, \theta_{ijk}; \{\xi_m\}), \tag{7.47}$$

Finally, several four-body symmetry functions can be defined based on the fundamental invariants of the internal symmetry group. For example, combining the triplet (jkl) with the center site, we have

$$G_{4}^{\mathsf{K},\mathsf{K}_1,\mathsf{K}_2}(\{\xi_m\}) = \sum_{jkl\neq i} \mathtt{b}_{jkl}^{\mathsf{K},\mathsf{K}_1,\mathsf{K}_2} \times F_4(R_{ij}, R_{ik}, R_{il}, \cdots; \theta_{ijk}, \theta_{ikl}, \cdots). \tag{7.48}$$

It is worth noting that most of the symmetry functions also depend on the IR indices

K of the internal symmetry group. We also note that since the relative angles θ_{ijk} are pre-defined constants for models on a regular lattice, the dependence of the F functions on these angles is trivial. More importantly, these F functions are used to select the more relevant pairs or triplets to be included in the symmetry functions.

In particular, the symmetry functions can be simplified to a sum over the symmetric-IR for lattice models. Take G_{3b} as an example, we first divide all atomic pairs (jk) in the neighborhood into inequivalent classes such that pairs within the same class are related by the point group symmetry. Moreover, since pairs belong to the same class are related by rotations or reflections that preserve the distance from the center site, they share the same value of the F_3 function; see Fig. 7.2(a) for an example of the symmetry-related pairs on a square lattice. Using π to denote the inequivalent classes of pairs, we then have

$$G_{3b}^{\mathsf{K}_1,\mathsf{K}_2}(\{\xi_m\}) = \sum_{\pi} F_3'(\pi;\{\xi_m\}) \sum_{\hat{g}} \mathsf{p}_{\pi(\hat{g})}^{\mathsf{K}_1,\mathsf{K}_2}. \tag{7.49}$$

Here $\pi(\hat{g})$ denotes atomic pairs (jk) related to a reference pair in the class π by the symmetry operation $\hat{g}$. The sum over $\hat{g}$, which is the symmetric sum of the pair-wise invariants p, corresponds to the 1D trivial IR of the lattice point group. Consequently, the symmetry function G_{3b} is manifestly an invariant of both the internal and lattice symmetry groups.

To briefly conclude this Section, we have formulated a general theory of descriptors for characterizing dynamical classical fields in condensed matter systems, and pre-

sented various different, yet related, approaches for computing the invariant feature variables. The group-theoretical method offers a rigorous and systematic approach to derive a descriptor based on the bispectrum coefficients. Finally, we discuss a descriptor that incorporates the symmetry of the classical fields into the atom-centered symmetry functions. Explicit implementations of these descriptors are demonstrated for well-studied correlated electron systems in the following sections and previous chapters.

7.3 Example: Dynamics of Cooperative Jahn-Teller Coupling

As a second example of the type-I models, we consider a model of itinerant e_g electrons interacting with the vibrational modes of MO_6 octahedra on a square lattice. Since our main interest here is the orbital ordering, we will neglect the electron spin degrees of freedom and the t_{2g} core spins. In $LaMnO_3$, the orbital ordering (accompanied by a JT transition) takes place at $T_{\mathrm{OO}} = 750$ K, which is much higher than the magnetic transition at $T_M = 120$ K, below which an A-type antiferromagnetic order develops. In the paramagnetic phase above T_M, the spin degrees of freedom form a homogeneous, fluctuating background and we only consider the orbital degrees of freedom in this work. Although orbital orders in manganites are three-dimensional, as a first step towards large-scale simulations, in this work we focus on a two-dimensional version of the cooperative JT systems, which exhibits the same $\mathbf{K} = (\pi, \pi)$ C-type order/JT order in the ground state.

The Hamiltonian of the cooperative JT model on a square lattice is given by

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_{\mathrm{K}} + \hat{\mathcal{H}}_{\mathrm{JT}} + \mathcal{E}_{\mathrm{L}}. \tag{7.50}$$

The three terms correspond to the electron kinetic energy, the electron-phonon coupling, and the lattice elastic energy, respectively. The Hamiltonian $\hat{\mathcal{H}}_{\mathrm{K}}$ describes the nearest-neighbor hopping of e_g electrons

$$\hat{\mathcal{H}}_{\mathrm{K}} = \sum_{\gamma=x,y} \sum_{\langle ij\rangle \| \gamma} \sum_{\mu\nu=a,b} \left(t^{\gamma}_{\mu\nu} \hat{c}^{\dagger}_{i\mu} c_{j\nu} + \mathrm{h.c.} \right), \tag{7.51}$$

where $\hat{c}^{\dagger}_{i,\mu}$ and $\hat{c}_{i,\mu}$ represent the creation and annihilation operators, respectively, of an electron with orbital flavor μ at the i-th site, a and b denote the two basis $d_{x^2-y^2}$ and $d_{3z^2-r^2}$ of e_g orbitals, $t^{\gamma}_{\mu\nu}$ denotes the orbital-dependent anisotropic hopping between nearest-neighbor pairs $\langle ij\rangle$ parallel to the direction $\gamma = x, y$ on the square lattice. The following hopping coefficients are used in our model calculations [204,367]: $t^{aa}_{ij} = -\sqrt{3}t^{ab}_{ij} = -\sqrt{3}t^{ba}_{ij} = 3t^{bb}_{ij} = t_{\mathrm{nn}}$ for hopping along the x-direction, and $t^{aa}_{ij} = \sqrt{3}t^{ab}_{ij} = \sqrt{3}t^{ba}_{ij} = 3t^{bb}_{ij} = t_{\mathrm{nn}}$ along the y-direction.

The second term in $\hat{\mathcal{H}}$ describes the JT coupling:

$$\hat{\mathcal{H}}_{\mathrm{JT}} = -\lambda \sum_{i} \left(Q^{A_1}_i \hat{n}_i + \boldsymbol{Q}^{E}_i \cdot \hat{\boldsymbol{\tau}} \right), \tag{7.52}$$

where $Q^{A_1}_i = Q^1_i$ denotes the breathing mode of the MO_6 octahedron at site-i, which couples to the electron number operator

$$\hat{n}_i = \hat{c}^{\dagger}_{ia}\hat{c}_{ia} + \hat{c}^{\dagger}_{ib}\hat{c}_{ib}. \tag{7.53}$$

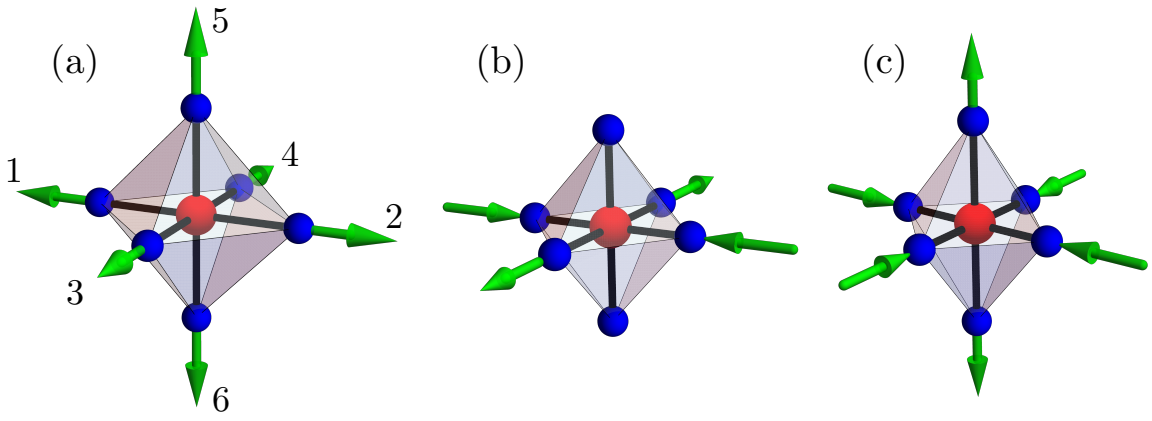

Figure 7.3: Schematic diagram of the vibronic modes for the MnO6 octahedron: (a) the symmetry-preserving breathing mode, (b) and (c) the symmetry-breaking JT modes. In terms of oxygen displacements, the coordinates of these normal modes are: $Q^{A_1} = (-X_1 + X_2 - X_3 + X_4 - X_5 + X_6)/\sqrt{6}$, $Q^x = (-X_1 + X_2 + X_3 - X_4)/2$, and $Q^z = (-X_1 + X_2 - X_3 + X_4 + 2X_5 - 2X_6)/\sqrt{12}$, where X_i denotes the x coordinates of the i-th oxygen, and so on.

The doublet $\boldsymbol{Q}_i^E = (Q_i^x, Q_i^z) = (Q_i^2, Q_i^3)$ describes the JT modes that break the cubic symmetry of the octahedron. Schematic diagrams of these three lattice modes are shown in Fig. 7.3. The JT doublet couples to the pseudo-spin operator $\hat{\boldsymbol{\tau}}_i = (\hat{\tau}_i^x, \hat{\tau}_i^z)$ representing the electron orbital degrees of freedom:

$$\hat{\tau}_i^x = \hat{c}_{ia}^\dagger \hat{c}_{ib} + \hat{c}_{ib}^\dagger \hat{c}_{ia}, \qquad \hat{\tau}_i^z = \hat{c}_{ia}^\dagger \hat{c}_{ia} - \hat{c}_{ib}^\dagger \hat{c}_{ib}. \tag{7.54}$$

The third term in Eq. (7.50) is the classical elastic energy of the lattice distortions

$$\mathcal{E}_{\mathrm{L}} = \frac{K}{2} \sum_i \left[\beta \left(Q_i^{A_1} \right)^2 + \left| \boldsymbol{Q}_i^E \right|^2 \right], \tag{7.55}$$

where K denotes an effective elastic coefficient. The parameter β is defined as $\beta = (\omega_{A_1}/\omega_E)^2$, where ω_{A_1} and ω_{E} are the vibration energies for the breathing and doublet JT modes, respectively, assuming that the reduced masses for these two modes are

the same. Following previous works [367, 368], this parameter is set to $\beta = 2$ in the following calculations.

It is worth noting that the breathing and JT modes of each octahedra are independent of each other in this elastic model. More realistic approach naturally needs to include couplings between neighboring octahedra, for example: $\mathcal{E}'_L = \sum_{ij} K_{ij}^{mn} Q_i^m Q_j^n$, where the indices $m, n = 1, x, z$. The equilibrium structural distortion is thus determined by both such direct elastic interactions as well as the effective electron-mediated interaction through the cooperative JT effect discussed above. In general, as same oxygen atoms are shared by two neighboring octahedra, nearest-neighbor couplings between these Q modes are antiferromagnetic, i.e. $K > 0$, which are compatible with both the C-type orbital order and the meta-stable CDW order to be discussed later. For simplicity, in this work we neglect the direct elastic couplings and focus on the intrinsic cooperative JT mechanism.

The ground state of the cooperative JT model in Eq. (7.50) at half-filling exhibits a C-type orbital ordering, characterized by a wave vector $\mathbf{K} = (\pi, \pi)$, accompanied by a structural distortion of the antiferro-distortive order with a predominate Q^x mode [367,369–372]. A schematic diagram of the orbital/JT order is shown in Fig. 7.4. The orbital order can be described by the expectation values of the pseudo-spins. To this end, minimization of the total energy with respect to the JT distortions yields the relation

$$\langle \hat{\boldsymbol{\tau}}_i \rangle = \frac{K}{\lambda} \boldsymbol{Q}_i^E, \tag{7.56}$$

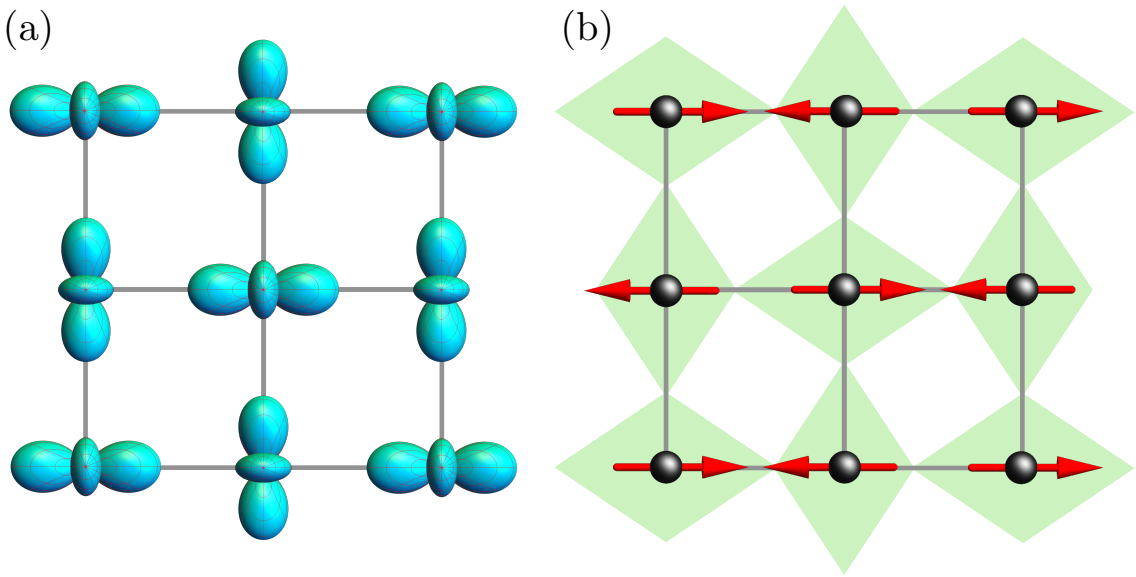

Figure 7.4: Schematic diagram of (a) C-type orbital order, and (b) the concomitant antiferro-distortive JT order. Also shown in panel (b) is the vector representation of the orbital/JT order. The arrows represent either the doublet vector $\boldsymbol{Q}^E$ or the expectation value of the orbital pseudo-spin $\langle\hat{\boldsymbol{\tau}}\rangle$. These two vectors are related to each other via Eq. (7.56) in the ground state.

It is worth noting that the linear JT coupling in Eq. (7.52) of a single octahedron is given by the inner product $(\hat{\boldsymbol{\tau}} \cdot \boldsymbol{Q}^E)$, which implies that the interaction is invariant under simultaneous rotation of the orbital pseudo-spin and JT doublet vector. Inclusion of quadratic JT couplings of the form $QQ\tau$ reduces this O(2) symmetry to a 3-fold degeneracy already at the single-octahedron level. On the other hand, in the cooperative JT scenario, the determination of the lattice distortions needs to include the kinetic energy of itinerant electrons on a lattice. Since the electron hopping in Eq. (7.51) is anisotropic with a strong orbital-dependence, the O(2) symmetry of the doublet vector is reduced to a two-fold mirror symmetry $Q^x \to -Q^x$ for the half-filled cooperative JT system [360]. The resultant minima are found to be along the Q^x

direction $\boldsymbol{Q}_*^E = \pm(Q_*, 0)$, and the checkerboard arrangement of these two symmetry-related minima $\boldsymbol{Q}_*^E$ in the C-type order comes from a dominant electron-mediated nearest-neighbor interaction of antiferro-distortive sign [367].

The Z_2 mirror symmetry discussed above corresponds to a tetragonal lattice symmetry. As shown in Fig. 7.3(b), changing the sign of Q^x sends a tetrahedron elongated in the x-direction to one along the y-direction. The eigenstates of $\hat{\tau}_x$ with eigenvalue $+1$ and -1 are dominated by $d_{3x^2-r^2}$ and $d_{3y^2-r^2}$, respectively. The breaking of the global Z_2 symmetry of the cooperative JT system leads to the C-type orbital order which can be described viewed as a Néel type order of Ising variable Q^x. On symmetry ground, the orbital JT phase transition is expected to belong to the Ising universality class. A detailed thermodynamic study of the orbital/JT phase transition by, e.g. Monte Carlo simulations, remain to be done.

Also of interest is the transition dynamics of the simultaneous JT distortion and orbital ordering. Of particular interest is the coarsening behaviors of orbital domains and whether the resultant domain-growth law falls into well-established universality classes of phase-ordering dynamics. To this end, we consider the dynamical evolution of the JT systems based on the approximation that electron relaxations are much faster than the dynamics of JT distortions. This adiabatic approximation is similar to the Born-Oppenheimer approximation widely used in the *ab initio* molecular dynamics methods [87]. In particular, this means that the electronic contribution to the driving forces are computed from an equilibrium Fermi liquid of the instantaneous lattice

distortion. As both breathing and JT distortions are vibronic normal modes, their dynamics are governed by an effective Newton equation of motion

$$\mu_{A_1}\frac{d^2Q_i^{A_1}}{dt^2}+\gamma_{A_1}\frac{dQ_i^{A_1}}{dt}=-\frac{\partial\mathcal{E}_L}{\partial Q_i^{A_1}}-\frac{\partial\langle\mathcal{H}_{\rm JT}\rangle}{\partial Q_i^{A_1}}+\eta_i^{A_1}(t),$$
$$\mu_E\frac{d^2\boldsymbol{Q}_i^{E}}{dt^2}+\gamma_E\frac{d\boldsymbol{Q}_i^{E}}{dt}=-\frac{\partial\mathcal{E}_L}{\partial \boldsymbol{Q}_i^{E}}-\frac{\partial\langle\mathcal{H}_{\rm JT}\rangle}{\partial \boldsymbol{Q}_i^{E}}+\boldsymbol{\eta}_i^{E}(t). \tag{7.57}$$

Here μ_α is the effective mass of the octahedral normal modes $\alpha = A_1$ and E, the corresponding damping coefficients and Langevin noises are denoted by γ_α and η_α, respectively. For simplicity, we shall assume the same parameters for the breathing and JT modes. In particular, as discussed above, the difference in effective mass can be accounted for by the ratio β of the elastic constants. As in standard Langevin method, the thermal forces are Gaussian random variables with zero mean and variance consistently related to the damping coefficients γ through the dissipation-fluctuation theorem.

The deterministic forces have two contributions, corresponding to the first two terms on the right-hand side: the elastic restoring forces and the electronic forces via JT coupling. The calculation of forces can be simplified using the Hellmann-Feynman theorem: $\partial\langle\hat{\mathcal{H}}\rangle/\partial Q = \langle\partial\hat{\mathcal{H}}/\partial Q\rangle$. Using Eq. (7.52) and (7.55), one obtains the following expressions for the driving forces

$$F_i^{A_1}=-\beta K Q_i^{A_1}+\lambda\langle\hat{n}_i\rangle,$$
$$\boldsymbol{F}_i^{E}=-K\boldsymbol{Q}_i^{E}+\lambda\langle\hat{\boldsymbol{\tau}}_i\rangle. \tag{7.58}$$

The equilibrium condition $\boldsymbol{F}_i^E = 0$ gives the relation in Eq. (7.56). As discussed above, expectation values $\langle \cdots \rangle$ are computed based on an equilibrium electron liquid corresponding to the instantaneous lattice configuration $\{Q_i^{A_1}, \boldsymbol{Q}_i^E\}$. Explicitly, for example, the expectation values of orbital pseudo-spins are given by

$$\langle \hat{\boldsymbol{\tau}}_i \rangle = \frac{1}{Z_e} \mathrm{Tr} \Big[\hat{\boldsymbol{\tau}}_i \, e^{-\beta \hat{\mathcal{H}}_e \left(Q_i^{A_1}, \boldsymbol{Q}_i^E \right)} \Big] \tag{7.59}$$

where $\hat{\mathcal{H}}_e = \hat{\mathcal{H}}_{\mathrm{K}} + \hat{\mathcal{H}}_{\mathrm{JT}}$ is the electron Hamiltonian, $\beta = 1/k_B T$ is the inverse temperature, and $Z_e = \mathrm{Tr} e^{-\beta \hat{\mathcal{H}}_e}$ is the electron partition function. As the electron Hamiltonian $\mathcal{H}_e$ is quadratic in the fermion creation/annihilation operators, it can be solved by the ED in real space. Yet, since the electronic forces have to be computed at every time-step of the Langevin dynamics simulation, the $\mathcal{O}(N^3)$ time complexity of ED can be overwhelmingly time-consuming for large-scale simulations.

7.3.1 Machine Learning Force-Field Models

Here we present a ML framework for computing the electronic forces in cooperative JT systems with a linear-scaling complexity. Fundamentally, as pointed out by W. Kohn, linear-scaling electronic structure methods are possible mainly because of the locality nature or "nearsightedness" principle of many-electron systems [100, 101]. Importantly, modern ML techniques provide an explicit and efficient approach to incorporate the locality principle into the implementation of $O(N)$ methods. Perhaps the most prominent and successful demonstration of this approach is the ML force-field methods developed in quantum chemistry to enable large-scale *ab initio* MD

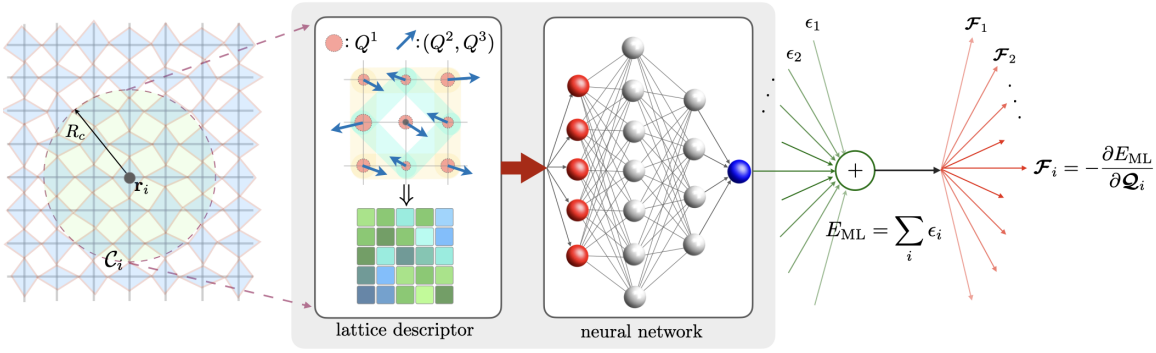

Figure 7.5: Schematic diagram of the ML force-field model for the cooperative JT models. A lattice descriptor transforms the neighborhood distortion configuration $\mathcal{C}_i$ into effective coordinates $\{G_m\}$ which are then fed into a fully connected NN. The output node of the NN corresponds to the local site-energy ϵ_i. The combination of the descriptor and the NN provides an approximation for the universal function $\varepsilon(\cdot)$. The corresponding total potential energy E is obtained from the summation of the local energies. Automatic differentiation is employed to compute the derivatives $\partial E_{\rm ML}/\partial \boldsymbol{\mathcal{Q}}_i$ for the effective forces acting on the breathing and JT modes.

simulations [91–94, 96, 97, 125, 127–129, 175–178]. Similar to the Langevin dynamics for cooperative JT systems described above, the atomic forces in *ab initio* MD are computed by solving, for example, the Kohn-Sham equation, which has to be repeated at every time-step [87]. The central idea behind the linear scalability of ML force-field methods is the divide-and-conquer approach proposed in the pioneering works of Behler and Parrinello [91] and Bartók *et al.* [92]. The ML model is trained to produce a *local* atomic energy from a finite neighborhood. The atomic forces are obtained indirectly from the total energy, which is the sum of all atomic energies.

Similar ML frameworks have recently been developed to enable large-scale dynamical simulations in several condensed-matter lattice systems [58,59,98,99,132,179,180]. In particular, the ML force-field approach was applied to the ferromagnetic Kondo-

lattice or s-d models [98, 99]. The strong-coupling regime of this model corresponds to the DE mechanism which is another important component of colossal magnetoresistance physics. The spin dynamics in the DE system is driven by itinerant electrons, similar to the cooperative JT models. Based on a generalized Behler-Parrinello scheme, a neural network model was trained to predict local effective fields induced by propagating electrons. The ML force-field methods have also been applied to the semiclassical Holstein model on a square lattice [179], a canonical system for studying the physics of electron-phonon coupling and phonon-assisted charge-density wave orders. The scalar Einstein phonons in the standard Holstein model can be viewed as a simplified model for the breathing mode Q^{A_1}. Large-scale dynamical simulations enabled by the ML methods unveiled intriguing anomalous coarsening behavior of charge-density waves in the Holstein model [179].

7.3.2 Behler-Parrinello Machine-Learning Framework

The Langevin dynamics can be viewed as a MD method for the octahedral normal modes. This prompts a generalization of the BP approach for the adiabatic dynamics of JT systems. However, there are two major differences between the two systems. First, both breathing and JT modes are defined on a lattice, in contrast to coordinates of atoms in free space. Second, the symmetry of the JT modes is tied to the symmetry of the underlying lattice, while MD systems are characterized by continuous translational and rotational symmetries. The BP scheme is modified to account for these two important issues; see FIG. 7.5 for a schematic diagram of the ML force-

field model for the JT systems. First, the total energy of the system in the adiabatic approximation is given by the expectation value of the Hamiltonian in Eq. (7.50), which is to be approximated by a ML energy E_{ML}. As in the BP scheme, this system energy is partitioned into local energies, each associated with a lattice site:

$$\langle \hat{\mathcal{H}} \rangle \approx E_{\mathrm{ML}} = \sum_i \epsilon_i = \sum_i \varepsilon(\mathcal{C}_i). \tag{7.60}$$

Here we have invoked the locality assumption and express the site-energy ϵ_i as a function of lattice distortions in a local neighborhood, denoted as $\mathcal{C}_i$. Explicitly, we define the local distortion configuration as

$$\mathcal{C}_i = \left\{ \boldsymbol{\mathcal{Q}}_j \;\middle|\; |\mathbf{r}_j - \mathbf{r}_i| < r_c \right\}, \tag{7.61}$$

where r_c is a cutoff distance which is determined by the locality of the forces, and we have grouped the breathing and JT modes into a three-component vector $\boldsymbol{\mathcal{Q}} = (\mathcal{Q}^1, \mathcal{Q}^2, \mathcal{Q}^3) = (\mathcal{Q}^{A_1}, \mathcal{Q}^x, \mathcal{Q}^z)$ for convenience. It should be noted that the vector notation here does not imply an underlying O(3) rotation symmetry. The complex dependence of the site-energy on local environment is encoded in the function $\varepsilon(\cdot)$ which is universal for a given JT electronic Hamiltonian and electron filling fraction. Importantly, this universal function is to be approximated by a ML model in the BP approach.

As shown in FIG. 7.5, there are two central components in the BP-type ML model: a descriptor and a learning model. The former is to transform the neighborhood configuration into a proper feature variables, while the latter is to approximate the

universal function $\varepsilon(\cdot)$. In this work a feedforward neural network (NN) is employed as the learning model which, according to universal approximation theorem [280, 281], offers the capability of accurately representing complex functions to the desired accuracy. The total energy is given by the sum of all site-energies in the system, which are obtained by applying the same ML model to all lattice sites. The fact that the same ML model is used for all lattice sites simply reflects the translational symmetry of the original Hamiltonian. The effective forces acting on the octahedral normal modes are given by the derivatives of the total energy

$$\boldsymbol{\mathcal{F}}_i = -\frac{\partial E_{\mathrm{ML}}}{\partial \boldsymbol{\mathcal{Q}}_i}, \tag{7.62}$$

which can be efficiently computed using automatic differentiation [211, 373]. This expression also indicates that the ML predicted forces are conservative, which is an appealing feature of the PB-type ML models.

7.3.3 Descriptors for Lattice Distortions

The BP scheme also allows for a systematic approach to incorporate symmetry requirements into ML models through feature engineering. Since the local energy ϵ_i at the output of the NN is a scalar, it is invariant under symmetry operations of the system. However, despite the powerful approximation capability of NNs, such symmetry constraints can only be learnt statistically, sometimes with the help of techniques such as data augmentation. However, the symmetry constraints cannot be exactly implemented based on deep-learning alone. A descriptor is introduced

here to provide a proper representation of the neighborhood configuration in such a way that the representation is itself invariant under transformations of the relevant symmetry group. The resultant feature variables are input to the NN model. As a result, symmetry-related configurations $\mathcal{C}_i$ are described by exactly the same feature variables, which in turn produces exactly the same local energy at the output.

The importance of descriptors in the implementation of ML force field models for quantum MD was also emphasized in the original works of Behler and Parrinello [91]. A set of feature variables called the atom-centered symmetry functions (ACSFs) are introduced to represent local atomic configurations such that the rotation and reflection symmetries are exactly incorporated into the atomic energy function [91, 127]. The building blocks of ACSFs are relative distances and angles of atomic position vectors, which are manifestly invariant under rotations of the SO(3) symmetry group [127]. The ACSF descriptor is physically intuitive, yet to some extent *ad hoc*, approach to ML force-field models for MD systems. Since then many atomic descriptors have been proposed and implemented [92, 96, 104, 175, 181–185].

Since the JT models are defined on a lattice, the relevant symmetry group is reduced from the SO(3) rotation group of free space to the point group associated with the underlying lattice. Moreover, the JT modes are also characterized by well-defined transformation rules of the same point group. To describe the combined symmetry transformations, consider the neighborhood $\mathcal{C}_i$ centered at site-i, and a discrete rotation or reflection $\hat{g}$ of the point group that sends site-j to k, i.e. $\mathbf{R}_{ki} = O(\hat{g}) \cdot \mathbf{R}_{ji}$,

where $O(\hat{g})$ is the orthogonal matrix representation of $\hat{g}$. The transformation of the octahedral distortions is described by

$$\tilde{Q}_k^{(\gamma,m)} = M_{mn}^{(\gamma)}(\hat{g})\, Q_j^{(\gamma,n)}. \tag{7.63}$$

Here γ indicates the irreducible representation (IR) of the vibronic modes, and $M^{(\gamma)}$ is the orthogonal transformation matrix of IR-γ, the indices $m, n = 1, 2, \cdots n_\gamma$ label the different components in this IR. For example, for the double $\boldsymbol{Q}^E$, a matrix corresponding to a 90°-rotation is $M^E(C_{\pi/4}) = \mathrm{diag}(-1, 1)$.

Importantly, a proper representation of the neighborhood $\mathcal{C}_i$ needs to be invariant under these coupled symmetry transformations. A systematic approach to derive invariants of a symmetry group is the group-theoretical bispectrum method [103]. Atomic descriptors based on bispectrum coefficients have been used in conjunction with Gaussian processing learning models for quantum MD simulations [92, 104]. The group-theoretical method has also been employed to develop a general theory of descriptors for electronic lattice models in condensed-matter systems [58,105,131,283]. In particular, a descriptor based on the idea of reference IRs was developed for the Holstein model [179], which is essentially an electron-phonon model with the breathing modes.

Here we outline the group-theoretical derivation of invariant feature variables for the JT models. First, the octahedral distortions as represented by the set $\mathcal{C}_i$ essentially form a high-dimensional representation of the point group. They can then be decom-

posed into fundamental IRs of the point group. This decomposition can be highly simplified as the original representation matrix is automatically block-diagonalized, with each block corresponding to a fixed distance from the center-site. Standard methods can then be applied to the decomposition of each block [102].

For the square-lattice JT system described by the D_4 point-group symmetry, there are three types of blocks with a dimension of either 4 or 8. The derivation of the relevant IRs for each block can be further simplified. This is because while the JT modes $\boldsymbol{Q}^E$ transform as a doublet in the octahedral group O_h, they are reduced to the direct sum of two 1D IRs when restricted to the D_4 group. Explicitly, under symmetry operations of D_4, both Q^{A_1} and Q^z transform as A_1 IR, while Q^x transforms according to IR B_1, acquiring a -1 under reflections about the $y = \pm x$ diagonals. Take the vibronic modes $\{\boldsymbol{\mathcal{Q}}_A, \boldsymbol{\mathcal{Q}}_B, \boldsymbol{\mathcal{Q}}_C, \boldsymbol{\mathcal{Q}}_D\}$ at the four nearest-neighbor sites as an example, the four Q^z modes can be decomposed as $4Q^z = 1A_1 + 1B_1 + 1E$, with the following IR coefficients:

$$\begin{aligned} f^{A_1} &= Q^z_A + Q^z_B + Q^z_C + Q^z_D, \\ f^{B_1} &= Q^z_A - Q^z_B + Q^z_C - Q^z_D, \\ \boldsymbol{f}^E &= (Q^z_A - Q^z_C, Q^z_B - Q^z_D). \end{aligned} \tag{7.64}$$

The decomposition of the breathing Q^{A_1} modes is described by the same formulas. On the other hand, while the four JT Q^x modes are also decomposed as $4Q^x =$

$1A_1 + 1B_1 + 1E$, the B_1 symmetry of Q^x gives rise to different IR coefficients:

$$\begin{aligned} f^{A_1} &= Q^x_A - Q^x_B + Q^x_C - Q^x_D, \\ f^{B_1} &= Q^x_A + Q^x_B + Q^x_C + Q^x_D, \\ \boldsymbol{f}^{E} &= (Q^x_A - Q^x_C, -Q^x_B + Q^x_D). \end{aligned} \tag{7.65}$$

For convenience, we arrange the coefficients of the r-th IR of type Γ in the overall decomposition of $\mathcal{C}_i$ into a vector $\boldsymbol{f}^\Gamma_r = (f^{(\Gamma,r)}_1, f^{(\Gamma,r)}_2, \cdots, f^{(\Gamma,r)}_{n_\Gamma})$. As the sum of the squared coefficients of an IR, $p^\Gamma_r = \left|\boldsymbol{f}^\Gamma_r\right|^2$, is manifestly invariant under the point group, the set of all amplitudes $\{p^\Gamma_r\}$, known as the power spectrum, offers a set of invariant feature variables. The power spectrum is a subset of more general invariant variables called the bispectrum coefficients [103, 104]. A bispectrum coefficient is a product of three IR coefficients and the Clebsch-Gordon coefficients which account for the different transformation properties of the three IRs. Importantly, the relative phases between different IRs are encoded in these coefficients, which provide a faithful invariant representation of the neighborhood.

For most point groups, the dimensions n_Γ of individual IRs are small, which means there is a large multiplicity (indexed by r) for each IR. This in turn results in a large number of possible bispectrum coefficients, often with considerable redundancy. A more economic approach to encode the phase information is the idea of reference IR coefficients $\boldsymbol{f}^\Gamma_{\text{ref}}$, one for each IR type of the point group [99]. These reference IR coefficients are derived using the same decomposition formulas, but based on distor-

tions $\overline{\boldsymbol{\mathcal{Q}}}$ obtained by averaging large blocks of bond and chirality variables in the local neighborhood in order to reduce sensitivity to small variations. Importantly, the reference IR allows one to introduce a "phase" variable for each IR in the decomposition $\exp(\phi_r^\Gamma) \equiv \boldsymbol{f}_r^\Gamma \cdot \boldsymbol{f}_{\rm ref}^\Gamma / |\boldsymbol{f}_r^\Gamma|\,|\boldsymbol{f}_{\rm ref}^\Gamma| = \pm 1$. The relative phase between IRs of the same type can then be inferred from their respective phases relative to the reference. Finally, the relative phases between IRs of different types are provided by the bispectrum coefficients from reference IR alone.

The power spectrum can be combined with the phases to form invariant feature variables $G_r^\Gamma = p_r^\Gamma \exp(i\phi_r^\Gamma)$. These are to be supplemented by the bispectrum coefficients $B_{\rm ref}^{\Gamma 1,\Gamma_2,\Gamma_3}$ obtained from the reference IR. The descriptor can be summarized by the following sequence of representations of the neighborhood

$$\left\{\boldsymbol{\mathcal{Q}}_j\right\} \;\rightarrow\; \left\{\boldsymbol{f}_r^\Gamma\right\} \;\rightarrow\; \left\{G_r^\Gamma, B_{\rm ref}^{\Gamma_1,\Gamma_2,\Gamma_3}\right\}. \tag{7.66}$$

As symmetry-related configurations are represented by exactly the same feature variables, the site-energy ϵ_i at the output of the NN is guaranteed to be the same, thus ensuring that the symmetry is preserved in the ML model.

7.3.4 Implementation Details and Benchmarks

Here we used PyTorch [374] to construct fully connected neural networks with six hidden layers. The number of neurons in successive hidden layers are: $1024 \times 512 \times 256 \times 128 \times 64 \times 64$. With a cutoff radius of $r_c = 7a$, where a is the lattice constant, for defining the size of the neighborhood, the number of neurons at the input layer

is determined by the number of feature variables and is fixed at 450. The ReLU function is used as the activation function between layers [139, 210]. The NN model is trained based on a loss function including the MSE of both the effective field and total energy:

$$L = \mu_F \frac{1}{N} \sum_{i=1}^{N} \left| \boldsymbol{\mathcal{F}}_i^{\mathrm{ED}} - \boldsymbol{\mathcal{F}}_i^{\mathrm{ML}} \right|^2 + \mu_E \left| E_{\mathrm{ED}} - E_{\mathrm{ML}} \right|^2, \tag{7.67}$$

where μ_H and μ_E determine the relative weights of the force and energy constraints in the loss function. Different combinations of these two weights have been experimented. Overall, a better performance was obtained by putting more emphasis on the force accuracy. As shown in Eq. (7.62), the effective forces are obtained from the derivative of the sum of local energies. This can be efficiently done using automatic differentiation in PyTorch [211]. Trainable parameters of the NN are optimized by the Adam stochastic optimizer [212] with a learning rate of 0.001. A 5-fold cross-validation and early stopping regularization are performed to prevent overfitting.

The above ML framework was applied to the JT model at exactly half-filling and for a filling fraction of $f = 0.49$. The nearest-neighbor hopping parameter t_{nn} served as the energy unit. The competition between JT coupling and elastic energy gives a distortion scale of $Q_0 \sim \lambda/K$, indicating an energy scale $E_{\mathrm{JT}} \sim \lambda^2/K$. Comparing this with the electron bandwidth, this suggests a dimensionless coupling constant $\tilde{\lambda} = E_{\mathrm{JT}}/W \sim \lambda^2/Kt_{\mathrm{nn}}$, which is set to 1.35 throughout the simulations. The effective masses μ and the elastic constant determine a the time-scale for the classical vibronic

dynamics: $\tau_0 = \Omega^{-1}$, where the characteristic frequency $\Omega = \sqrt{K/\mu}$. Given this time-scale, an effective damping coefficient $\gamma = 0.25\tau_0^{-1}$ and a time-step $\Delta t = 0.05\tau_0$ were used in our Langevin simulations. The training dataset was obtained from ED calculations of random distortions as well as ED-based Langevin dynamics simulations, both on a 30×30 lattice. It included 1700 random configurations, 1300 intermediate states during the relaxations, and 800 nearly equilibrium states, a total of $\sim 3.4 \times 10^6$ force data.

Although the ML method is originally designed to model the time-consuming electronic structure calculations, the electronic contribution to the forces, i.e. the $\langle \hat{\boldsymbol{\tau}}_i \rangle$ term in Eq. (7.58), exhibits a strong bimodal distribution as the system approaches the ground-state C-type orbital order. The two centers of the bimodal distribution correspond to the $d_{3x^2-r^2}$ and $d_{3y^2-r^2}$ orbitals in the checkerboard pattern. Since the net forces approach zero in such quasi-equilibrium states, the electronic force nearly cancels the classical part. The bimodal distribution can thus be attributed to a dominant leading-order dependence of the electronic force on the on-site JT distortion through the classical force, i.e. $\langle \hat{\boldsymbol{\tau}}_i \rangle = (K/\lambda)\boldsymbol{Q}_i^E + \boldsymbol{h}(\mathcal{C}_i)$; the second term here encodes the weaker yet subtle dependence on the neighborhood distortions. To avoid difficulty due to this bimodal distribution, it is more efficient to remove the dominant on-site classical term and focus the ML training on the second term. On the other hand, Eq. (7.58) shows that this function is simply proportional to the total force: $\boldsymbol{h}(\mathcal{C}_i) = \boldsymbol{F}_i^E/\lambda$. Indeed, our ED-Langevin simulations show that the total

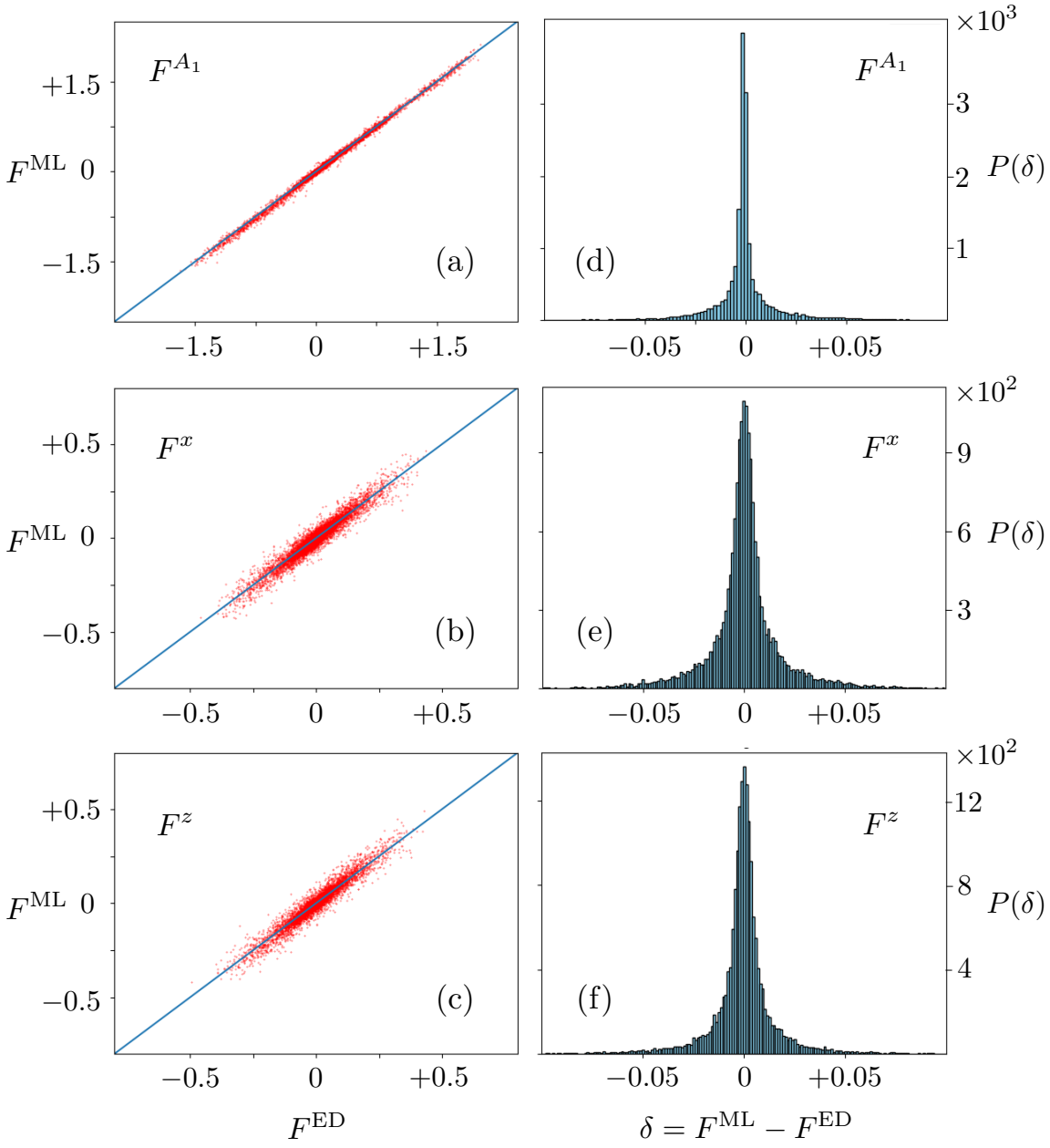

Figure 7.6: Benchmark of the ML force-field models for JT model with a filling fraction $f = 0.49$. Panels (a)–(c) on the left show the ML predicted forces F^{ML} versus the ground truth F^{ED} obtained from exact diagonalization method for the three vibronic modes of the octahedron. These forces are normalized by the electron-phonon coupling constant, which is set to $\lambda = 1.25$. The corresponding histograms of the prediction errors are shown in panels (d)–(f). Similar results are also obtained for the ML models for the half-filling case.

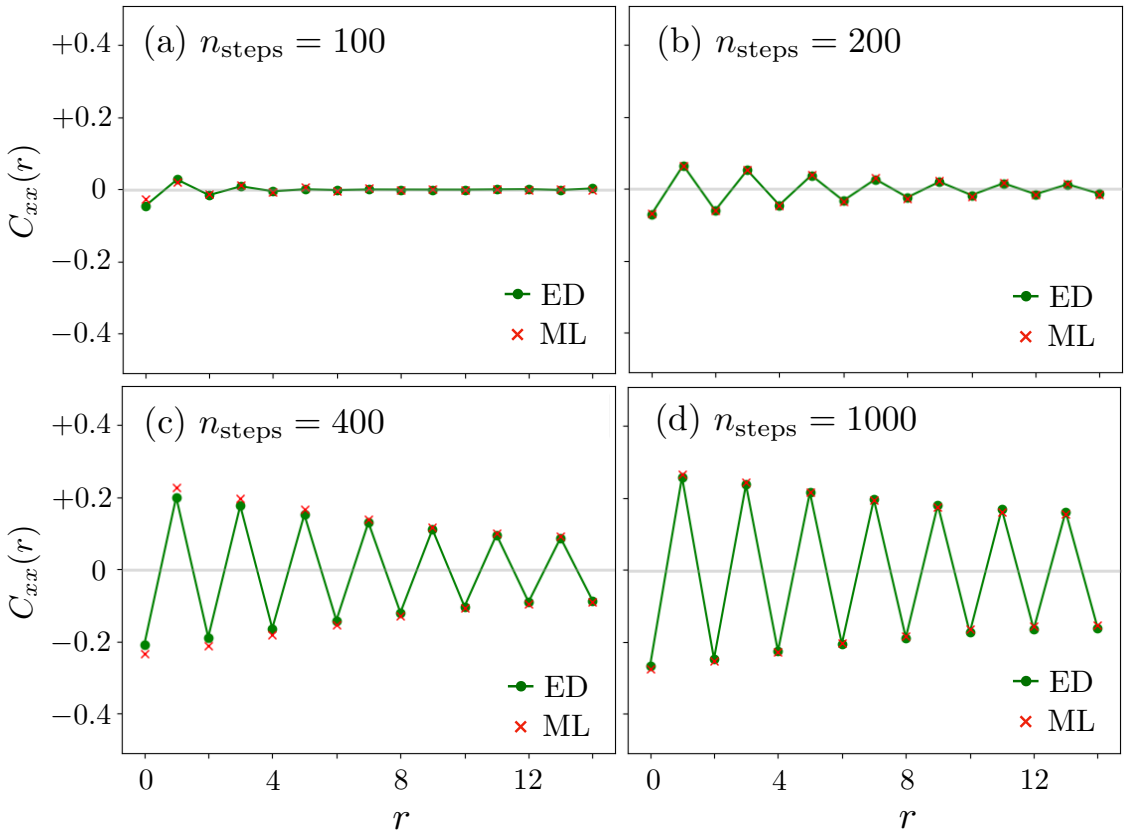

Figure 7.7: Comparison of time-dependent correlation functions $C^{xx}(r,t)$ defined in Eq. (7.68) obtained from Langevin simulations with the ML force-field model and the ED method. The correlation functions were obtained from 40 independent thermal quench simulations on a 30×30 lattice with a filling fraction $f = 0.49$.

forces on the JT modes are well described by a Gaussian-like distribution.

In addition to the generalized BP approach described above, we also implemented a modified ML scheme in which a separate NN is used to predict the forces acting on the breathing modes Q^{A_1}. From the symmetry viewpoint, this is feasible since the breathing mode as well as its corresponding force F^{A_1} are scalars, They are invariant under the point-group symmetry transformations, and the same descriptor can be also used for the corresponding NN model. The fact that no automatic differentiation is

required for the breathing modes also enhances the efficiency. But more importantly, we found that the prediction accuracy of the F^{A_1} forces is significantly improved. With this hybrid approach, rather accurate force predictions were achieved for all three vibronic modes, as shown in FIG. 7.6. The histogram of prediction errors ($\delta = F^{ML} - F^{ED}$), shown on the right panels of FIG. 7.6, are characterized by a small MSE of $\sigma_F \sim 0.005$ for the force predictions.

It is worth noting that both the electron density and the local orbital configuration can be directly obtained from the ML model, thanks to the local and linear coupling between electrons and octahedral modes. As shown in Eq. (7.58), for example, the expectation value of orbital pseudo-spin can be obtained as: $\langle \hat{\boldsymbol{\tau}}_i \rangle = (\hat{\boldsymbol{F}}_i^E + K\boldsymbol{Q}_i^E)/\lambda$, where $\hat{\boldsymbol{F}}_i^E = \boldsymbol{h}(\mathcal{C}_i)$ is the total doublet force predicted by the ML model. Similarly, the on-site electron number can also be obtained from the predicted forces $\hat{F}_i^{A_1}$ for the breathing mode: $\langle \hat{n}_i \rangle = (\hat{F}_i^{A_1} + \beta K Q_i^{A_1})/\lambda$.

Next we integrated the trained ML models into the Langevin dynamics method and conducted thermal quench simulations of the JT model. We initiated the simulations with a random initial state, which was suddenly quenched to a temperature of $T = 0.01$ at $t = 0$. The results from the ML-Langevin simulations were then compared with those from ED simulations. As the ground-state is characterized by a checkerboard pattern of the Q^x JT distortions, we computed the correlation function

$$C^{xx}(r_{ij}) = \langle Q_i^x Q_j^x \rangle \tag{7.68}$$

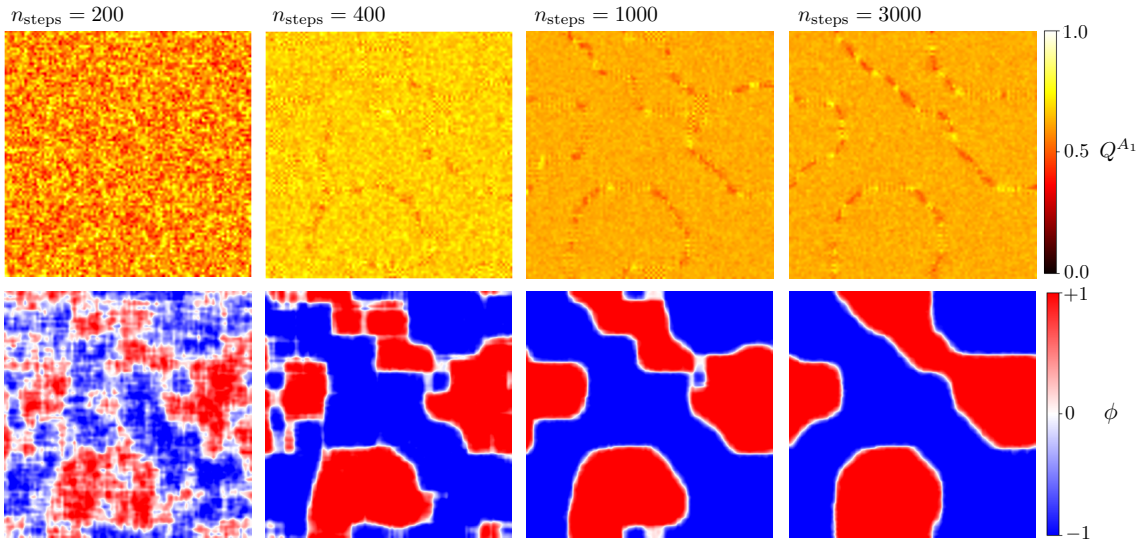

Figure 7.8: Snapshots of local breathing mode Q^{A_1} and orbital Ising order parameter ϕ_i at various time steps after a thermal quench of a half-filled $f = 0.5$ JT model. An initially random configuration is suddenly quenched to a temperature $T = 0.001$ at time $t = 0$ ($n_{\text{step}} = 0$). The ML-Langevin dynamics was used to simulate the relaxation of the system toward equilibrium. The red and blue regions correspond to orbital domains with order parameters $\phi = +1$ and -1, respectively. A time step $\Delta t = 0.05\tau_0$ is used in the simulations.

between the Q^x modes at two octahedra separated by a distance r_{ij} along either the x- or the y-direction. FIG. 7.7 displays the time-dependent correlation functions obtained from both ML and ED Langevin simulations, demonstrating excellent agreement. The relaxation process is characterized by a gradual built-up of staggered correlations of the Q^x distortion, characteristic of the C-type orbital/JT oder. This dynamical benchmark shows that the ML model not only accurately predicts the forces, but also captures the dynamical evolution of the JT system.

7.3.5 Coarsening of Orbital Order in Cooperative Jahn-Teller Model

The ML force-field model is applied to study the large-scale coarsening dynamics of orbital order in the JT model. In addition to the linear scalability of ML method, the efficiency is further enhanced by running the simulations on GPU machines. We performed the thermal quench simulations on a 100×100 lattice where an initial state with random local distortions was suddenly cooled to a low temperature $T = 0.01$ at time $t = 0$. As discussed above, the low-temperature C-type orbital/JT order exhibits a broken Z_2 symmetry, which is physically related to both the sublattice and mirror transformation about the diagonals. For the case of half-filling, this ground state is also characterized by a uniform electron density of $\langle \hat{n}_i \rangle = 1$. The lattice distortions in the ground state can be described by two parameters:

$$Q_i^{A_1} = \eta, \qquad \boldsymbol{Q}_i^E = (\delta, 0) \exp(i\mathbf{K} \cdot \mathbf{r}_i). \tag{7.69}$$

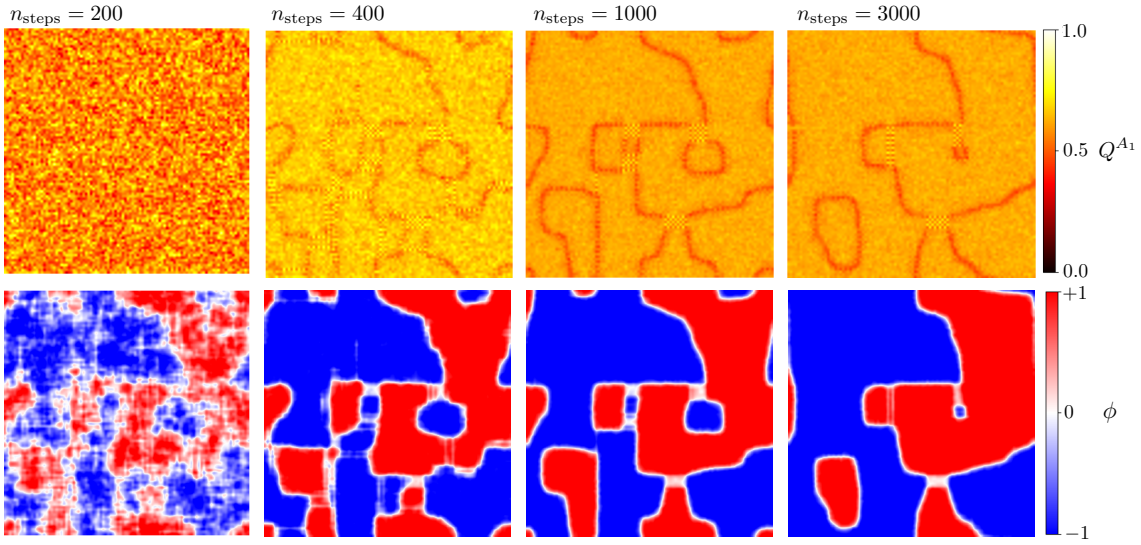

Figure 7.9: Snapshots of local breathing mode Q^{A_1} and orbital Ising order parameter ϕ_i at various time steps after a thermal quench for a JT model with $f = 0.49$ electron filling.

where $\mathbf{K} = (\pi, \pi)$ is the ordering wave vector, and the phase factor $\exp(i\mathbf{K} \cdot \mathbf{r}_i) = \pm 1$ for lattice sites at the two sublattices. The parameter $\eta = \lambda/\beta K$ is the amplitude of the uniform expansion of octahedra, while $\delta = \lambda\langle\hat{\tau}^x\rangle/K$ can be viewed as a global Z_2 order parameter for the staggered JT distortions. To characterize the inhomogeneous states with multiple orbital domains after a thermal quench, we define a scalar order parameter that measures the local staggered JT distortion

$$\phi_i = \left(Q_i^x - \langle Q_{\text{nn}}^x\rangle_i\right) \exp(i\mathbf{K} \cdot \mathbf{r}_i). \tag{7.70}$$

where $\langle Q_{\text{nn}}^x\rangle_i \equiv (Q_{i+\hat{\mathbf{x}}}^x + Q_{i-\hat{\mathbf{x}}}^x + Q_{i+\hat{\mathbf{y}}}^x + Q_{i-\hat{\mathbf{y}}}^x)/4$ denotes the average JT distortion at the four nearest neighbors of site-i. A nonzero ϕ_i thus indicates the presence of a local difference in JT distortion. Indeed, this Ising order parameter is $\phi_i = \delta$ in the long-range staggered JT order described in Eq. (7.69).

FIG. 7.8 show the snapshots of the breathing-mode amplitude $Q_i^{A_1}$ and the local Ising order parameter ϕ_i at different times after a thermal quench of a half-filled JT model. The red and blue regions, corresponding to $\phi_i = +1$ and -1, respectively, are C-type orbital orders related by the Z_2 sublattice symmetry. The two types of orbital domains are separated by interfaces of vanishing ϕ_i, corresponding to the white regions. On the other hand, a rather uniform ordering of Q^{A_1} is quickly developed throughout the system, except at the interfaces of the two orbital domains. The nearly constant amplitude Q^{A_1} corresponds to the uniform charge distribution $\langle \hat{n}_i \rangle \approx 1$ in the ground state as discussed above. At the interfaces that separate Ising domains of opposite signs, there are segments with an electron density both above and below the average value of one electron per site.

The snapshots during the relaxation of a JT system with an electron filling of $f = 0.49$ are shown in FIG. 7.9. Overall, a relaxation behavior similar to that of the half-filled case was obtained. However, there is a major difference regarding the breathing mode and the electron density. In contrast to the milder inhomogeneity for the half-filling case, the amplitude of the Q^{A_1} mode as well as the electron density are significantly reduced at the interfaces, as shown in the top panels of FIG. 7.9. As the C-type orbital/JT order is stabilized in the half-filling limit, this result is consistent with the phase separation scenario where nearly all of the doped holes go into the interfaces between orbital domains.

Another intriguing feature of the Q^{A_1} configurations is the localized checkerboard

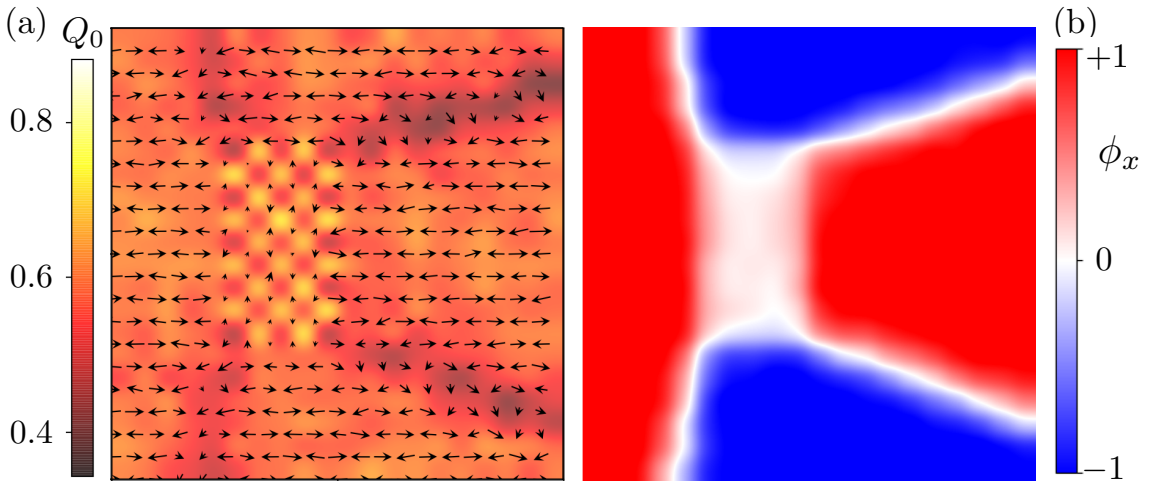

Figure 7.10: Localized checkerboard modulation of breathing mode Q^{A_1} and electron density. An example of such local structures is displayed in terms of (a) the amplitude of breathing mode as well as the $\boldsymbol{Q}^E$ vector and (b) the orbital Ising order parameter.

modulations of the breathing-mode amplitudes. Such localized checkerboard modulations can be seen in both filling fractions, but more pronounced in the $f = 0.49$ case. A close-up view of this localized structure is shown in FIG. 7.10(a). A region of checkerboard modulation of the Q^{A_1} amplitudes is enclosed in orbital domains represented by a Néel-type order of the doublet $\boldsymbol{Q}^E$ vectors predominantly in the Q^x direction. As the electron number $\hat{n}_i$ directly couples to the breathing mode, this checkerboard structure also indicates a concomitant charge density wave (CDW) characterized by the wave vector $\mathbf{K} = (\pi, \pi)$. It is worth noting that these CDW states are essentially the same as the CDW state of the Holstein model at half-filling [375–377]. In terms of the orbital Ising parameter ϕ, shown in FIG. 7.10(b), the checkerboard pattern is accompanied by a vanishing orbital order, similar to the interface that separates different orbital domains. However, the average electron density of the CDW

is found to be exactly one electron per site, which means that the doped holes are not accommodated in such local structures.

It is worth noting that these localized CDW structures are also obtained in the ED-based Langevin simulations on smaller systems, indicating that they are not an artifact of ML models. However, such charge modulation patterns are mostly metastable during the relaxation process. Our ED simulations on a 30×30 lattice found that a state initialized to a homogeneous CDW order is unstable and will decay into the C-type order. Yet, as shown in the snapshots from large-scale ML-Langevin simulations, e.g. FIG. 7.9, such local CDW order persists even at late stage of the phase-ordering where large domains of orbital order have been established. It is likely that the local CDW order is stabilized as an intermediate structure in a multi-domain state during relaxation.

Next we discuss the coarsening of the orbital domains. As discussed above, the C-type orbital order is described by an Ising order parameter. And since the orbital order and JT distortion are not subject to a conservation law, the coarsening of the orbital/JT order is expected to be described by the universality class of the phase ordering of a non-conserved Ising order. It is worth noting that the coarsening dynamics of Ising order has been thoroughly characterized and classified into several super-universal classes which depend on whether the Ising order is conserved and the presence of quenched disorder. For non-conserved Ising order, which is also the case for the C-type orbital order here, the coarsening dynamics is described by a curvature-

driven mechanism summarized in the Allen-Cahn equation [378]. This results in a specific power-law growth $L \sim t^{1/2}$ of Ising domains applicable to two and higher dimensions [70–72]. A natural question is to see whether the coarsening of the orbital domains falls into this Allen-Cahn universality class.

To quantify the coarsening dynamics, we consider the time-dependent correlation length of the orbital/JT order. To this end, we first compute the time-dependent structural factor of the staggered distortion defined as $S(\mathbf{k},t) = |Q^x(\mathbf{k},t)|^2$, where $Q^x(\mathbf{k},t)$ is the Fourier transform the JT Q^x configuration

$$Q^x(\mathbf{k},t) = \frac{1}{N}\sum_i Q_i^x\, e^{i\mathbf{k}\cdot\mathbf{r}_i}. \tag{7.71}$$

In FIG. 7.11 the structure factor is plotted at various time steps after a sudden quench to a low temperature $T = 0.01$. The emergence of the C-type order corresponds to a structure-factor peak at the wave vector $\mathbf{K} = (\pi,\pi)$. However, instead of a delta peak which is characteristic of a long-range order, a diffusive peak was observed even at late times of the phase ordering. The finite width of the diffusive peak is due to the presence of multiple orbital domains of opposite Ising orders. As the system progresses towards equilibrium, the coarsening of these ordered domains results in a stronger and sharper peak at $\mathbf{K}$, as can be seen in FIG. 7.11. The inverse of the width can thus provide a quantitative estimate for the characteristic length scale of ordered domains

$$L^{-1}(t) = \Delta k = \sum_{\mathbf{k}} S(\mathbf{k},t)|\mathbf{k}-\mathbf{K}| \Big/ \sum_{\mathbf{k}} S(\mathbf{k},t), \tag{7.72}$$

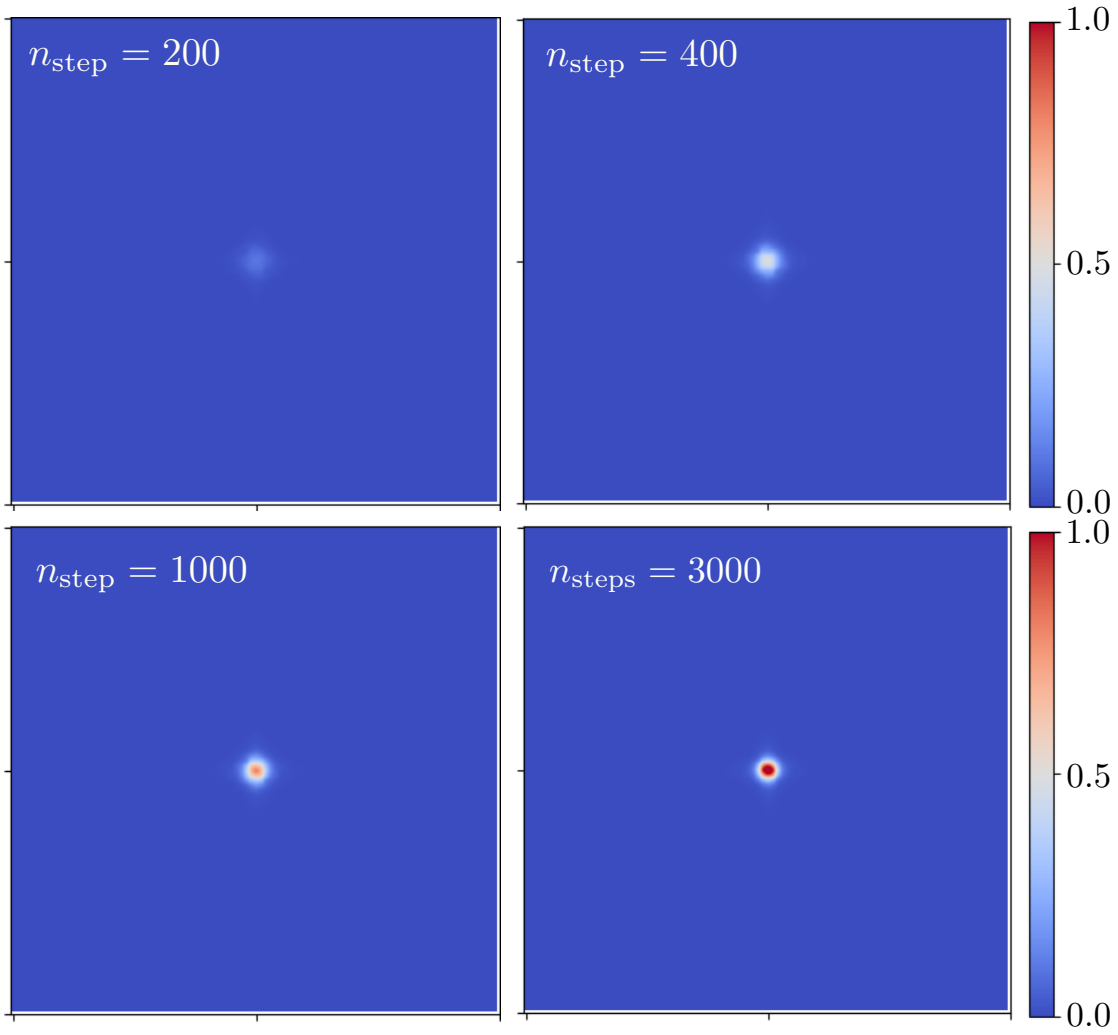

Figure 7.11: Structure factor $S(\mathbf{k}, t)$ of the JT Q^x distortion mode at different time steps after suddenly cooled to a temperature $T = 0.001$, starting from a random state. The system size is 100×100 with filling fraction $f = 0.49$, and averaged over 40 independent configurations.

This characteristic size of orbital domains also provides a measure of the orbital correlation, i.e.

$$C^{xx}(\mathbf{r},t) \sim e^{-|\mathbf{r}|/\xi(t)} e^{i\mathbf{K}\cdot\mathbf{r}}, \tag{7.73}$$

where the correlation length is proportional to the characteristic domain size, $\xi \sim L$.

FIG. 7.12 shows the characteristic length as a function of time for two filling fractions, $f = 0.5$ (one electron per site) and $f = 0.49$, both obtained by averaging over 40 independent ML-Langevin simulations on a 100×100 system. The $L(t)$ curves of the $f = 0.49$ case are also shown for two different temperatures. For all cases, the coarsening of the orbital domains shows a clear two-stage behavior: there is an initial rapid domain growth, which is followed by a much slower coarsening dynamics. This is also consistent with the snapshots shown in both FIG. 7.8 and 7.9. The initial fast stage, as represented by the two $n_{\text{steps}} = 200$ configurations, is characterized by emerging orbital domains accompanied by a featureless random distribution of Q^{A_1} and electron density. However, once the orbital orders are well developed in each domain at $n_{\text{steps}} \gtrsim 400$, or for times $t \gtrsim 20\tau_0$, the coarsening rate is significantly decreased.

By fitting the initial relatively fast growth with the Allen-Cahn law, as shown as dashed lines in FIG. 7.12(a), one can see that the $L(t)$ from simulations starts to deviate from the $\sqrt{t}$ behavior for $t \gtrsim 15$. On the other hand, the snapshots shown in FIG. 7.8 and 7.9 clearly show that there is still coarsening going on even at late times

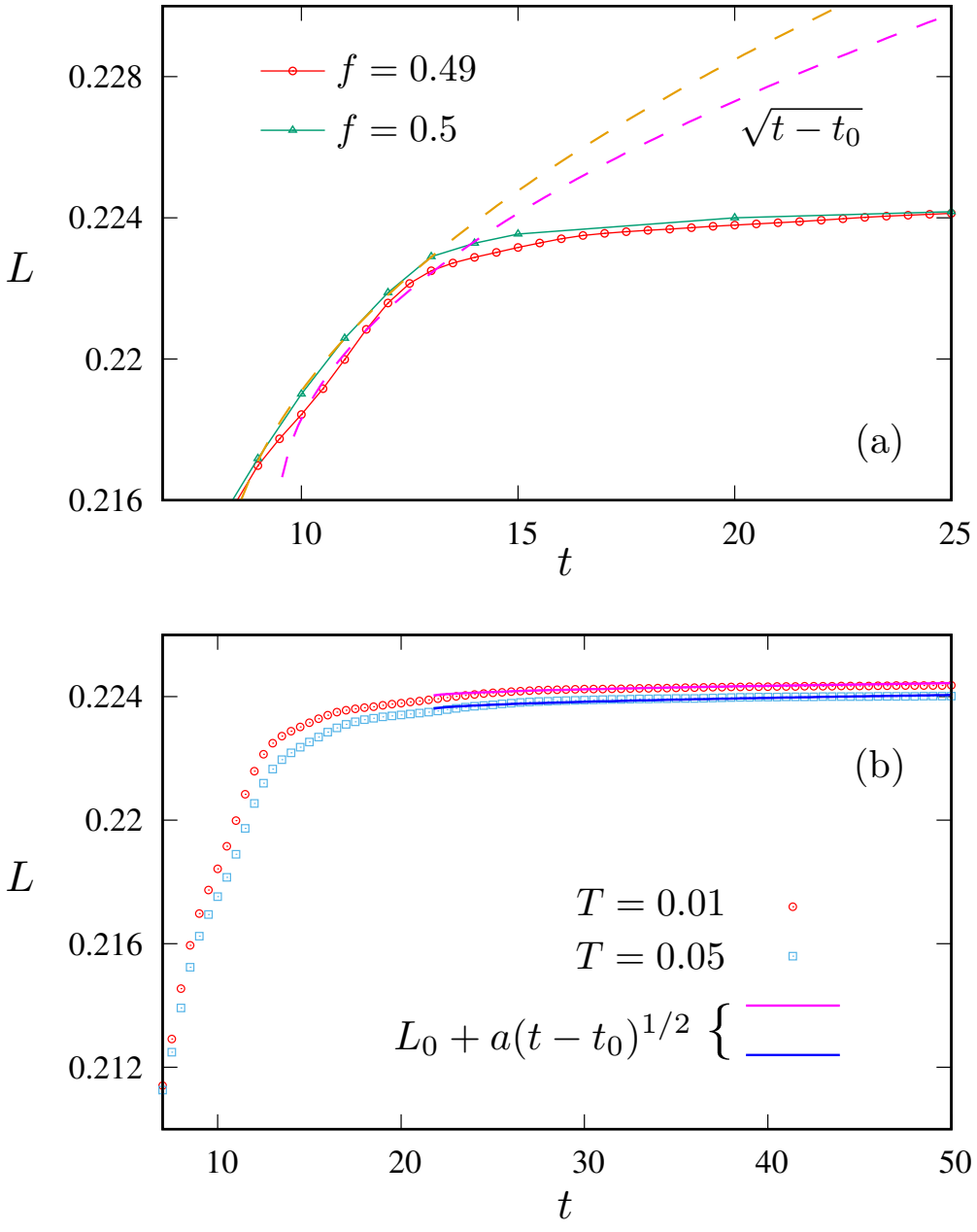

Figure 7.12: The characteristic length $L(t)$ of orbital domains as a function of time, averaging over 40 independent ML-Langevin simulations. Panel (a) shows $L(t)$ curves for the case of half-filling and a filling fraction $f = 0.49$. Panel (b) shows the $L(t)$ curves at two different temperatures for the case of $f = 0.49$ electron filling. The two dashed lines in (a) and the two solid lines in (b) are fit using the Allen-Cahn growth law with a shifted time t_0 and an initial length L_0. The fitted expansion coefficients are $a = 8.1 \times 10^{-5}$ and $a = 9.1 \times 10^{-5}$ for temperature $T = 0.001$ and $T = 0.005$, respectively.

of the relaxation. If we attempt to fit this late-stage coarsening using the equation $L(t) = L_0 + a(t - t_0)^{1/2}$, the extracted expansion coefficient $a \sim 10^{-4}$ is extremely small. This corresponds to a relative domain growth of the order of $\Delta L/L_0 \sim 10^{-3}$ over a time span of 30 oscillation cycles t_0 of the JT modes. This indicates a freezing behavior at the late stage of the phase ordering regardless of whether the phase-ordering can be described by the Allen-Cahn domain growth.

It is worth noting that the Allen-Cahn domain growth law is intimately related to the domain morphologies of the inhomogeneous states. Fundamentally, the Allen-Cahn equation for the domain-wall motion describes a curvature-driven coarsening process. Mathematically, the domain-wall velocity is given by $v = -r\kappa$, where κ is the curvature, and r is a proportional constant. Indeed, this equation implies a power-law domain growth. For inhomogeneous states characterized by a characteristic length L, the domain-wall velocity can be approximated by $v \sim dL/dt$, while the curvature is on average given by $\kappa \sim 1/L$. The Allen-Cahn equation indicates a differential equation $dL/dt \sim -1/L$, which can be readily integrated to give the growth behavior $L \sim t^{1/2}$.

Within the framework of curvature-driven domain growth, the late-stage freezing observed in our simulations could be attributed to the rather straight interfaces that separate orbital Ising domains of opposite signs; see FIG. 7.8 and 7.9. Moreover, the nearly straight interfaces tend to run parallel along the x or y-directions. Since the effective orbital interactions originate from the electrons through a mechanism similar

to RKKY interactions, this interfacial anisotropy can be attributed to the highly directional hopping of e_g electrons. The nearly zero curvature of such straight domain boundaries indicate a vanishing velocity for the domain-wall motion. Consequently, the coarsening is mostly driven by the corner regions of orbital domains, where a finite interfacial curvature remains. Let κ_c be a characteristic curvature of the corner regions, the corresponding length scales of curved interfaces is $\ell \sim 1/\kappa_c$. It should be noted that this curvature κ_c is determined by the competition of the electron and elastic energies of the JT model. When typical domain size L is greater than this length scale, curvature driven domain-wall motion is suppressed by the interfacial stiffness. This indicates a threshold length scale $L_{\rm th} \sim \ell \sim 1/\kappa_c$ such that orbital domains with a linear size $L \gtrsim L_{\rm th}$ start to show freezing behavior. The $L(t)$ curves shown in FIG. 7.12 seem to be consistent with this threshold scenario.

7.4 Summary and Discussion

In this work, we present a numerical framework of utilizing machine learning methods for multi-scale dynamical modeling of condensed matter systems with emergent dynamical classical fields. These classical degrees of freedom could arise from the coupling to lattice dynamics, or magnetic moments of localized d or f electrons. They could also represent the collective electron behaviors, as exemplified by the order-parameter field in a symmetry breaking phase, of interacting electrons. The slow adiabatic dynamics of the emergent classical fields is often dominated by the electrons or quasi-particles, which are assumed to be in quasi-equilibrium of the instantaneous

Hamiltonian parameterized by the classical variables. As in the quantum or *ab initio* molecular dynamics methods, accurate simulation of the dynamical classical fields requires solving the electronic structure problem at every time-step. Motivated by the success of ML-enabled large scale quantum MD simulations, we propose a similar approach for condensed matter systems in which the complex dependence of the local energy on the neighborhood classical field is encoded in a ML energy model.

The two important components of the ML energy model are the descriptor for characterizing the local classical field configuration, and the learning model used to encoded the dependence on the local environment. Several learning models developed in the context of quantum MD can also be used for the effective energy model of the condensed-matter systems. Among the various ML models, the deep-learning NN is perhaps the most versatile and accurate. The descriptor is crucial for properly incorporating symmetry of the system into the ML energy model. The so-called feature variables, which are input to the learning model, must be invariant with respect to symmetry transformations of the electron Hamiltonian. While a large number of descriptors have been proposed for ML-MD methods, the theory of descriptor for classical fields of condensed matter models has yet to be developed.

We discuss common features of the descriptor of classical fields of electronic lattice models, and formulate a general theory by first distinguishing two types of models depending on the absence or presence of an internal symmetry for the classical fields. Several specific approaches to derive a descriptor have been discussed. For example,

the approach, motivated by ML-models for quantum MD simulation, is the generalization of the ACSF which incorporates the internal symmetry of the classical fields.

The majority of our effort focuses on the group theoretical method which offers systematic and controlled approach to build fundamental invariants of the symmetry group. In this approach, the local classical fields, which form a high-dimensional representation of the site-symmetry point group, is first decomposed into the IRs. Fundamental invariants are given by the bispectrum coefficients of three IRs, which are similar to the scalar or triple product of three vectors. To cope with the issue due to the large number of the over-complete bispectrum invariants, we propose a simplification method based on the concept of the reference IRs. Instead of keeping all the bispectrum coefficients, both the amplitude and the relative "phase" of each IR can be faithfully retained via an inner product with the reference IR.

Finally, we demonstrate the implementation of the various descriptors on well-known lattice models including the JT model and the s-d Hamiltonian for itinerant magnets. The classical field in the former case corresponds to local structural distortions. In particular, the scalar field in the Holstein-type models offers the simplest example to illustrate the working of the lattice descriptor. On the other hand, the s-d model characterized by a vector classical field is used to demonstrate the construction of a descriptor with an independent internal symmetry.

Our work laid the foundation for applying ML methods to multi-scale dynamical modeling in condensed matter systems. Contrary to ML-based MD methods which

is an ongoing active research field by itself, the goal here is to model the adiabatic dynamics of classical fields under the influence of quasi-equilibrium electrons. The capability of going beyond empirical methods for large-scale dynamical simulations of such classical fields has numerous implications in condensed matter physics. For example, one particularly important application is the accurate dynamical modeling of topological defects of multi-component classical fields, which are prevalent in condensed matter systems. Notable examples include vortices in superconductivity and skyrmions in itinerant magnetism.

Moreover, complex inhomogeneous electronic states are ubiquitous in correlated electron systems. Not only are these mesoscopic textures of fundamental importance in correlated electron physics, they also play a crucial role in the emergence of novel macroscopic functionalities. For example, complex mixed-phase states are prevalent in colossal magnetoresistant materials and several high-T_c superconductors also exhibit intriguing stripe or checkerboard patterns. Accurate modeling of these complex nanoscale textures is thus of paramount importance in the engineering of these novel material functionalities. However, large-scale simulations of such electronic textures so far are based on empirical or phenomenological models, mostly because of the extreme difficulty for the multi-scale dynamical modeling of such systems. We believe that the ML force field approach along with the proper descriptor outlined in this work will be an indispensable tool to enable large-scale dynamical simulations of complex patterns in correlated electron materials.

Appendix

Appendix A

Details of the Descriptor for displacements on a diamond lattice

The silicon atoms around a centered atom form into four different cases: four points group, six points group, twelve points group, and twenty-four points group. Only former three cases occurred in our case of $2 \times 2 \times 2$-supercell. Four points case displacement vectors (corner points of the cube) can be decomposed into $4 \times 3 = A_1 \bigoplus E \bigoplus T_1 \bigoplus 2T_2$ with the following basis:

$$f^{A_1} = a_x + a_y + a_z - b_x - b_y + b_z + c_x - c_y - c_z - d_x + d_y - d_z$$

$$f_x^E = a_x + a_y - 2a_z - b_x - b_y - 2b_z + c_x - c_y + 2c_z - d_x + d_y + 2d_z$$

$$f_y^E = \sqrt{3}(-a_x + a_y + b_x - b_y - c_x - c_y + d_x + d_y)$$

$$f_x^{T_1} = a_y - a_z + b_y + b_z - c_y + c_z - d_y - d_z$$

$$f_y^{T_1} = a_x - a_z + b_x + b_z - c_x - c_z - d_x + d_z$$

$$f_z^{T_1} = -a_x + a_y + b_x - b_y + c_x + c_y - d_x - d_y$$

$$f_x^{T_2,1} = a_x + b_x + c_x + d_x$$

$$f_y^{T_2,1} = a_y + b_y + c_y + d_y$$

$$f_z^{T_2,1} = a_z + b_z + c_z + d_z$$

$$f_x^{T_2,2} = a_y + a_z + b_y - b_z - c_y - c_z - d_y + d_z$$

$$f_y^{T_2,2} = a_x + a_z + b_x - b_z - c_x + c_z - d_x - d_z$$

$$f_z^{T_2,2} = a_x + a_y - b_x - b_y - c_x + c_y + d_x - d_y$$

where the Cartesian coordinates of those points are $a(0.5, 0.5, 0.5)$, $b(-0.5, -0.5, 0.5)$, $c(0.5, -0.5, -0.5)$, and $d(-0.5, 0.5, -0.5)$, or $\{-a, -b, -c, -d\}$. The six points case displacement vectors (face-centered points of the cube) can be decomposed into $6\times3 = A_1 \bigoplus E \bigoplus 2T_1 \bigoplus 3T_2$ with the following basis:

$$f^{A_1} = a_x + b_y + c_z - d_z - e_y - f_x$$

$$f_x^{E} = a_x + b_y - 2c_z + 2d_z - e_y - f_x$$

$$f_y^{E} = \sqrt{3}(-a_x + b_y - e_y + f_x)$$

$$f_x^{T_1,1} = b_x - c_x - d_x + e_x$$

$$f_y^{T_1,1} = a_y - c_y - d_y + f_y$$

$$f_z^{T_1,1} = -a_z + b_z + e_z - f_z$$

$$f_x^{T_1,2} = -b_z + c_y - d_y + e_z$$

$$f_y^{T_1,2} = -a_z + c_x - d_x + f_z$$

$$f_z^{T_1,2} = a_y - b_x + e_x - f_y$$

$$f_x^{T_2,1} = a_x + f_x$$

$$f_y^{T_2,1} = b_y + e_y$$

$$f_z^{T_2,1} = c_z + d_z$$

$$f_x^{T_2,2} = b_z + c_y - d_y - e_z$$

$$f_y^{T_2,2} = a_z + c_x - d_x - f_z$$

$$f_z^{T_2,2} = a_y + b_x - e_x - f_y$$

$$f_x^{T_2,3} = b_x + c_x + d_x + e_x$$

$$f_y^{T_2,3} = a_y + c_y + d_y + f_y$$

$$f_z^{T_2,3} = a_z + b_z + e_z + f_z$$

where the Cartesian coordinates of those points are $a(2,0,0)$, $b(0,2,0)$, $c(0,0,2)$, $d(0,0,-2)$, $e(0,-2,0)$, and $f(-2,0,0)$. The twelve points case displacement vectors (edge-centered points of the cube) can be decomposed into $12\times3 = 2A_1 \bigoplus A_2 \bigoplus 3E \bigoplus 4T_1 \bigoplus 5T_2$ with the following basis:

$$\begin{aligned} f^{A_1,1} &= a_x + a_y + b_y + b_z + c_x + c_z + d_y - d_z + e_x - e_z - f_y + f_z \\ &\quad - g_x - g_z - h_y - h_z - i_x + i_z + j_x - j_y - k_x + k_y - l_x - l_y \end{aligned}$$

$$f^{A_1,2} = a_z + b_x + c_y - d_x - e_y - f_x + g_y + h_x - i_y - j_z - k_z + l_z$$

$$\begin{aligned} f^{A_2} &= a_x - a_y + b_y - b_z - c_x + c_z + d_y + d_z - e_x - e_z - f_y - f_z \\ &\quad + g_x - g_z - h_y + h_z + i_x + i_z + j_x + j_y - k_x - k_y - l_x + l_y \end{aligned}$$

$$f_x^{E,1} = 2a_x + 2a_y - b_y - b_z - c_x - c_z - d_y + d_z - e_x + e_z + f_y - f_z$$

$$+ g_x + g_z + h_y + h_z + i_x - i_z + 2j_x - 2j_y - 2k_x + 2k_y - 2l_x - 2l_y$$

$$f_y^{E,1} = \sqrt{3}(b_y + b_z - c_x - c_z + d_y - d_z - e_x + e_z$$

$$- f_y + f_z + g_x + g_z - h_y - h_z + i_x - i_z)$$

$$f_x^{E,2} = -b_y + b_z - c_x + c_z - d_y - d_z - e_x - e_z$$

$$+ f_y + f_z + g_x - g_z + h_y - h_z + i_x + i_z$$

$$f_y^{E,2} = \frac{1}{\sqrt{3}}(2a_x - 2a_y - b_y + b_z + c_x - c_z - d_y - d_z + e_x + e_z + f_y + f_z$$

$$- g_x + g_z + h_y - h_z - i_x - i_z + 2j_x + 2j_y - 2k_x - 2k_y - 2l_x + 2l_y)$$

$$f_x^{E,3} = 2a_z - b_x - c_y + d_x + e_y + f_x$$

$$- g_y - h_x + i_y - 2j_z - 2k_z + 2l_z$$

$$f_y^{E,3} = \sqrt{3}(b_x - c_y - d_x + e_y - f_x - g_y + h_x + i_y)$$

$$f_x^{T_1,1} = a_x - c_x - e_x - g_x - i_x + j_x + k_x + l_x$$

$$f_y^{T_1,1} = a_y - b_y - d_y - f_y - h_y + j_y + k_y + l_y$$

$$f_z^{T_1,1} = b_z - c_z + d_z - e_z + f_z - g_z + h_z - i_z$$

$$f_x^{T_1,2} = a_y - c_z + e_z - g_z + i_z - j_y - k_y + l_y$$

$$f_y^{T_1,2} = a_x - b_z + d_z + f_z - h_z - j_x - k_x + l_x$$

$$f_z^{T_1,2} = b_y - c_x - d_y + e_x - f_y - g_x + h_y + i_x$$

$$f_x^{T_1,3} = -b_y + b_z + d_y + d_z - f_y - f_z + h_y - h_z$$

$$f_y^{T_1,3} = -c_x + c_z + e_x + e_z + g_x - g_z - i_x - i_z$$

$$f_z^{T_1,3} = a_x - a_y - j_x - j_y + k_x + k_y - l_x + l_y$$

$$f_x^{T_1,4} = a_z - c_y + e_y + g_y - i_y - j_z + k_z - l_z$$

$$f_y^{T_1,4} = a_z - b_x + d_x - f_x + h_x + j_z - k_z - l_z$$

$$f_z^{T_1,4} = b_x - c_y + d_x - e_y - f_x + g_y - h_x + i_y$$

$$f_x^{T_2,1} = a_x + c_x + e_x + g_x + i_x + j_x + k_x + l_x$$

$$f_y^{T_2,1} = a_y + b_y + d_y + f_y + h_y + j_y + k_y + l_y$$

$$f_z^{T_2,1} = b_z + c_z + d_z + e_z + f_z + g_z + h_z + i_z$$

$$f_x^{T_2,2} = a_y + c_z - e_z + g_z - i_z - j_y - k_y + l_y$$

$$f_y^{T_2,2} = a_x + b_z - d_z - f_z + h_z - j_x - k_x + l_x$$

$$f_z^{T_2,2} = b_y + c_x - d_y - e_x - f_y + g_x + h_y - i_x$$

$$f_x^{T_2,3} = b_y + b_z - d_y + d_z + f_y - f_z - h_y - h_z$$

$$f_y^{T_2,3} = c_x + c_z - e_x + e_z - g_x - g_z + i_x - i_z$$

$$f_z^{T_2,3} = a_x + a_y - j_x + j_y + k_x - k_y - l_x - l_y$$

$$f_x^{T_2,4} = a_z + c_y - e_y - g_y + i_y - j_z + k_z - l_z$$

$$f_y^{T_2,4} = a_z + b_x - d_x + f_x - h_x + j_z - k_z - l_z$$

$$f_z^{T_2,4} = b_x + c_y + d_x + e_y - f_x - g_y - h_x - i_y$$

$$f_x^{T_2,5} = b_x + d_x + f_x + h_x$$

$$f_y^{T_2,5} = c_y + e_y + g_y + i_y$$

$$f_z^{T_2,5} = a_z + j_z + k_z + l_z$$

where the Cartesian coordinates of those points are $a(1,1,0)$, $b(0,1,1)$, $c(1,0,1)$, $d(0,1,-1)$, $e(1,0,-1)$, $f(0,-1,1)$, $g(-1,0,-1)$, $h(0,-1,-1)$, $i(-1,0,1)$, $j(1,-1,0)$, $k(-1,1,0)$, and $l(-1,-1,0)$.

The feature variables which are invariant under symmetry operations of the point-group T_d are given by the following invariants:

$$
\begin{aligned}
G^{A_1} &= f^{A_1} \\
G^{A_2} &= f^{A_2} f^{A_2}_{\mathrm{ref}} \\
G^{E}_1 &= \langle \boldsymbol{f}^{E}, \boldsymbol{f}^{E} \rangle \\
G^{E}_2 &= \langle \boldsymbol{f}^{E}, \boldsymbol{f}^{E}_{\mathrm{ref}} \rangle \\
G^{T_1}_1 &= \langle \boldsymbol{f}^{T_1}, \boldsymbol{f}^{T_1} \rangle \\
G^{T_1}_2 &= \langle \boldsymbol{f}^{T_1}, \boldsymbol{f}^{T_1}_{\mathrm{ref}} \rangle \\
G^{T_1}_3 &= f^{A_2}_{\mathrm{ref}} (f^{T_1}_x f^{T_1}_{\mathrm{ref},y} f^{T_1}_{\mathrm{ref},z} + f^{T_1}_y f^{T_1}_{\mathrm{ref},x} f^{T_1}_{\mathrm{ref},z} + f^{T_1}_z f^{T_1}_{\mathrm{ref},x} f^{T_1}_{\mathrm{ref},y}) \\
G^{T_2}_1 &= \langle \boldsymbol{f}^{T_2}, \boldsymbol{f}^{T_2} \rangle \\
G^{T_2}_2 &= \langle \boldsymbol{f}^{T_2}, \boldsymbol{f}^{T_2}_{\mathrm{ref}} \rangle \\
G^{T_2}_3 &= f^{T_2}_x f^{T_2}_{\mathrm{ref},y} f^{T_2}_{\mathrm{ref},z} + f^{T_2}_y f^{T_2}_{\mathrm{ref},x} f^{T_2}_{\mathrm{ref},z} + f^{T_2}_z f^{T_2}_{\mathrm{ref},x} f^{T_2}_{\mathrm{ref},y}
\end{aligned}
$$

Appendix B

Tests of Gravity with Gravitational-Wave through Machine Learning

B.1 introduction

The rapid advancements in gravitational-wave astronomy have enabled the development of sophisticated and efficient models for gravitational waveforms, particularly in exploring modifications to General Relativity (GR), such as Einstein-dilation Gauss-Bonnet (EdGB) gravity [379]. One promising approach for efficiently predicting gravitational-wave signals is the Deep INference for Gravitational-wave Observations (DINGO) model [380], which employs machine learning techniques to expedite parameter estimation for gravitational-wave signals. This chapter examines how DINGO can be utilized to accurately predict gravitational-wave signals under EdGB modifications.

The DINGO model leverages a neural posterior estimation approach that employs deep learning techniques to predict the posterior distributions of gravitational-wave

events. Traditional Bayesian parameter inference methods, such as Markov Chain Monte Carlo (MCMC), are computationally expensive due to the extensive number of likelihood evaluations required [381]. In contrast, DINGO utilizes neural networks to bypass this computational bottleneck by using simulated data to train a model capable of generating posterior samples in real time once a gravitational-wave detection occurs. This results in a significant reduction in inference time, from several hours or even days to just a few seconds per event.

The primary innovation of DINGO lies in its ability to use neural networks to learn the mapping between gravitational-wave data and source parameters, making it a powerful tool for rapid parameter estimation. This enables real-time data analysis and detailed investigation of gravitational-wave events with exceptional speed and accuracy. DINGO is trained on a comprehensive dataset comprising both simulated gravitational-wave signals and detector noise, allowing the model to generalize effectively across different noise profiles. The main advantage of this approach is a substantial reduction in computational costs without compromising accuracy.

The model architecture of DINGO is based on normalizing flows, a form of deep generative model that allows efficient sampling from complex posterior distributions. By using normalizing flows, the DINGO model can produce highly accurate estimates of gravitational-wave source parameters, and it can do so orders of magnitude faster than conventional techniques. The underlying neural network is trained on a large number of simulated gravitational waveforms, which include variations in detector

noise and system parameters. The process of neural posterior estimation effectively turns the problem of posterior sampling into a problem of supervised learning, where the model is trained to learn the mapping from input gravitational-wave data to the posterior over source parameters.

The implementation of DINGO further incorporates several advanced deep learning techniques, including residual connections and batch normalization, which help stabilize training and improve the model's ability to capture intricate dependencies within the data. By training the model on a diverse set of simulated waveforms, DINGO can accurately account for various features of the gravitational waveforms, including spin precession and eccentricity, which are crucial for realistic gravitational-wave parameter inference.

EdGB gravity is an extension of GR that incorporates additional scalar fields coupled to the Gauss-Bonnet term in the action. This modification introduces higher-order curvature corrections, which become particularly significant in high-curvature regimes, such as those encountered during compact binary mergers. The primary effect of EdGB corrections is the introduction of additional phase terms in the inspiral phase of gravitational waveforms. These corrections are most prominent during the inspiral phase of compact binaries, making gravitational-wave signals a valuable observational tool for probing these modifications to gravity.

B.2 Einstein-dilation Gauss-Bonnet Gravity

B.2.1 Theory

We begin by presenting the action for scalar Gauss-Bonnet (sGB) gravity [382–385]:

$$S = \int d^4x\sqrt{-g}\left[\frac{\mathcal{R}}{16\pi} - \frac{1}{2}(\nabla\phi)^2 + \alpha_{\rm GB} f(\phi)\mathcal{R}^2_{\rm GB}\right] + S_m\,. \tag{B.1}$$

Here g is the determinant for the metric $g_{\mu\nu}$, $\mathcal{R}$ is the Ricci scalar, ϕ is a scalar field, $\alpha_{\rm GB}$ is the coupling constant between the scalar field and the metric, S_m is the matter action, and

$$\mathcal{R}^2_{\rm GB} = R_{\mu\nu\sigma\rho}R^{\mu\nu\sigma\rho} - 4R_{\mu\nu}R^{\mu\nu} + \mathcal{R}^2\,, \tag{B.2}$$

is the GB invariant. $f(\phi)$ is an arbitrary function of the scalar field that determines how it is coupled to the metric. EdGB gravity is realized by choosing $f(\phi) = e^{-\gamma\phi}$ for a constant γ. As shown in [386,387], this theory can be written in a second-order, hyperbolic form that is well-posed for numerical relativity evolution within a range of parameter space.

String theory predicts even higher-order curvature terms in the action that we do not include in the analysis. To justify this and treat the theory as an effective field theory, we work in the small coupling approximation scheme (or reduced-order scheme) where we assume that the GR contribution is dominant and handle EdGB corrections as small perturbations. In particular, we define a dimensionless coupling constant

$$\zeta \equiv \frac{16\pi\alpha^2_{\rm GB}}{L^4}\,, \tag{B.3}$$

where L is the characteristic length of the system and assume $\zeta \ll 1$. This technique has been used to find scalar charges of compact objects [388–390], corrections to the gravitational-wave phase at the inspiral stage [388], and to carry out numerical simulations of binary black hole (BBH) mergers [391].

Let us study the theory within the small coupling approximation scheme in more detail. We perturb field equations in $\alpha_{\rm GB}$ and solve them order by order. Then, $\phi = \mathcal{O}(\alpha_{\rm GB})$ and one can expand $f(\phi)$ in small ϕ as:

$$f(\phi) = f(0) + f'(0)\phi + \mathcal{O}(\phi^2)\,. \tag{B.4}$$

The first term is a constant and this does not change the field equations from the GR ones as the GB invariant is a topological term and can be rewritten as a total derivative. Thus, the leading effect comes from the second term where the scalar field is linearly coupled to the GB invariant. For this reason, we consider the following action in this work:

$$S = \int d^4x\sqrt{-g}\left[\frac{\mathcal{R}}{16\pi} - \frac{1}{2}(\nabla\phi)^2 + \alpha_{\rm GB}\phi\mathcal{R}^2_{\rm GB}\right] + S_m\,, \tag{B.5}$$

where we have absorbed $f'(0)$ into $\alpha_{\rm GB}$. In this theory, BHs can have non-vanishing scalar charges [388, 389] while neutron stars (NS) do not [390] within the small coupling approximation.

B.2.2 Gravitational Waveforms

We next find EdGB corrections to the gravitational waveform phase. Given that most of the signal-to-noise ratios (SNRs) for gravitational-waves from NSBHs and BBHs come from the inspiral portion, we focus on the inspiral stage in our analysis. The leading correction to the phase at the inspiral stage enters at -1 post-Newtonian (PN) order due to the scalar dipole radiation and was derived in [383]. Some of the higher PN corrections were recently derived in [392, 393].

Within the stationary phase approximation [394,395], the waveform in the Fourier space is given by:

$$h(f) = A(f) \exp\left[i\,\Psi(f)\right] , \quad \Psi(f) = \Psi_{\rm GR}(f) + \delta\Psi(f) \,. \tag{B.6}$$

Here $A(f)$ is the amplitude, $\Psi_{\rm GR}$ is the GR phase, and the EdGB correction to the phase $\delta\Psi$ (up to $\mathcal{O}(\alpha_{\rm GB}^2)$) is given in a form

$$\delta\Psi = \sum_i \delta\Psi_{\,i\,\rm PN} = \frac{\alpha_{\rm GB}^2}{M^4} \sum_i c_i\, v^{-5+2i} \,. \tag{B.7}$$

Here $v = (\pi M f)^{1/3}$ is the relative velocity of the binary constituents with gravitational-wave frequency f and the total mass $M = m_1 + m_2$, where m_1 and m_2 are the masses of the primary and secondary objects of the system. The coefficients c_i up to 2PN order can be found in the reference paper [381].

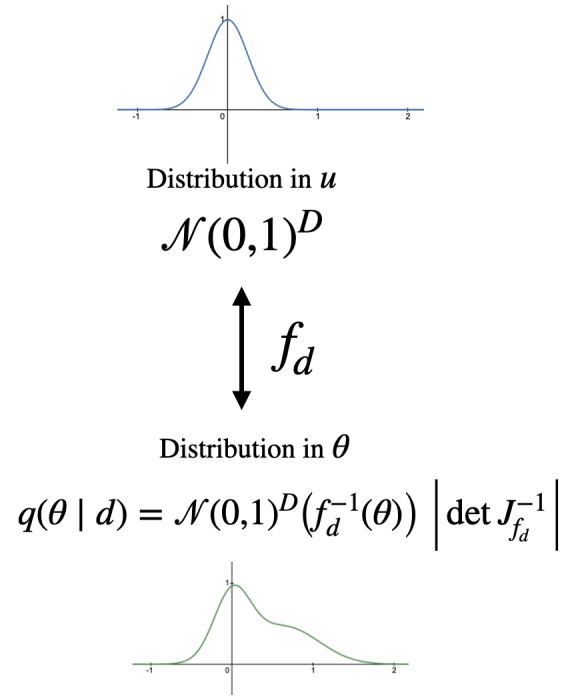

Figure B.1: A diagram showing how normalizing flow works [396].

B.3 Normalizing Flow

Normalizing flows are a type of deep generative model that have proven highly effective in a range of applications involving complex probability distributions. They work by transforming a simple base distribution, such as a standard Gaussian, into a more complicated target distribution through a series of invertible and differentiable transformations, shown in Fig B.1. This framework allows for efficient sampling and density evaluation, making it particularly suitable for tasks requiring posterior inference in high-dimensional parameter spaces.

In the context of gravitational-wave parameter estimation, normalizing flows are used to approximate the posterior distribution of source parameters given the ob-

served data. This is achieved by training the model to learn a sequence of transformations that maps the latent space of base distributions to the target posterior distribution. By leveraging this powerful technique, the DINGO model can perform accurate and rapid inference of gravitational-wave signals, even in scenarios involving modifications to GR like EdGB gravity. Normalizing flows thus provide the foundational generative modeling capabilities that underpin the efficiency and scalability of the DINGO framework.

The neural posterior estimation framework of DINGO can be applied to investigate deviations in gravitational-wave signals from GR due to EdGB modifications. Specifically, the inclusion of EdGB gravity introduces higher PN corrections to the gravitational wave phase, including a – 1PN dipole term and additional higher-order terms. The DINGO model is trained on waveform simulations that incorporate these EdGB terms, thereby enabling it to detect deviations in waveform morphology resulting from EdGB effects.

Incorporating EdGB modifications into the DINGO model involves adjusting the training set to include waveform simulations that reflect the presence of EdGB corrections. These corrections manifest as deviations in the gravitational waveform's phase evolution, particularly during the inspiral phase, which can be used to infer the presence of EdGB contributions. The EdGB coupling constant, denoted as $\sqrt{\alpha_{\rm EdGB}}$, plays a key role in quantifying the strength of these modifications, and its value can be constrained by analyzing the observed gravitational-wave signals.

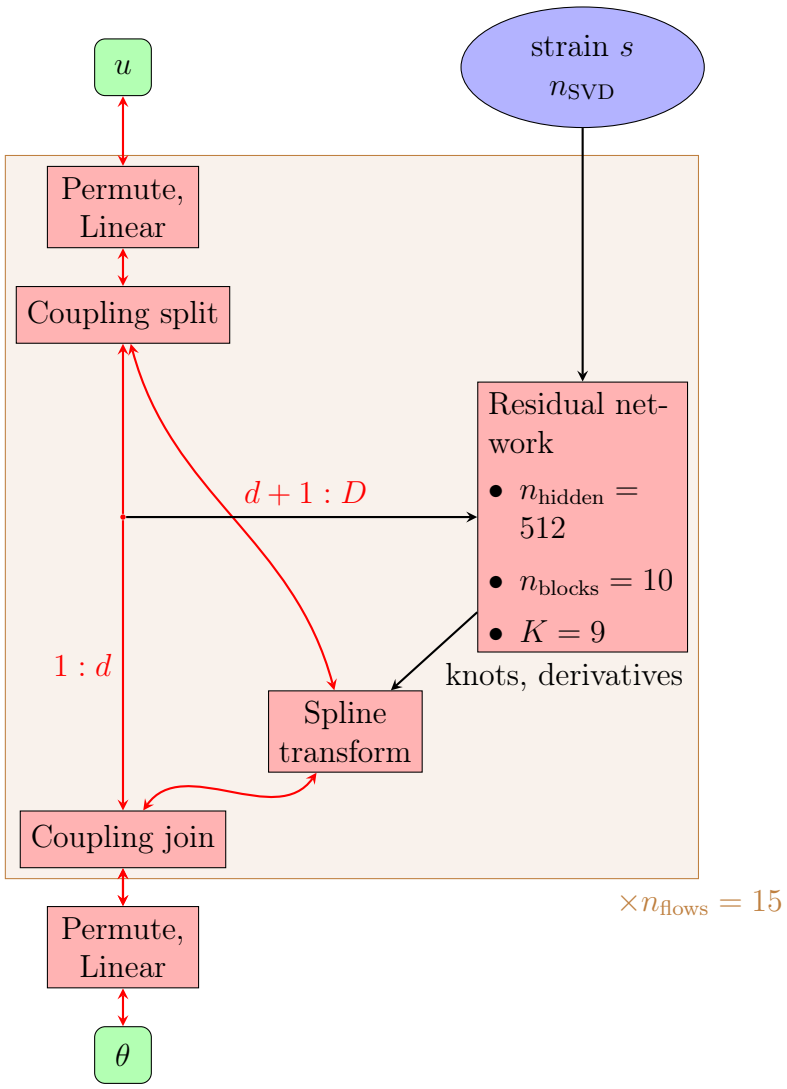

Figure B.2: Overall structure of the normalizing flow [2] from the base space (u) to the parameter space (θ), with optimal hyperparameter choices indicated. Red connections are invertible. The residual network is made up of n_{blocks} residual blocks, each with two fully-connected hidden layers of n_{hidden} units. Prior to each linear transformation [3], we inserted batch normalization layers to speed training [4] and Exponential Linear Units for nonlinearity [5]. Each block is also conditioned on the strain data s.

By using the DINGO model, we expect that posterior distributions for the coupling constant $\sqrt{\alpha_{\rm EdGB}}$ can be efficiently computed for different gravitational-wave events involving NSBH and BBH mergers. Unlike traditional methods, DINGO allows for rapid computations, which are crucial for analyzing large datasets and evaluating EdGB modifications across multiple events simultaneously. This ability to provide real-time estimates of modified gravity parameters is especially valuable for follow-up observations and multi-messenger astronomy, where rapid response times are essential.

Following 500 epochs of training—comprising 200 epochs focused on fiducial ASD noise and an additional 300 epochs encompassing all ASD noise for observing run O1—the loss function for the DINGO model, which utilizes 200 SVD components, has been reduced to 4.18. Importance sampling for GW150914 is depicted in Fig B.3.

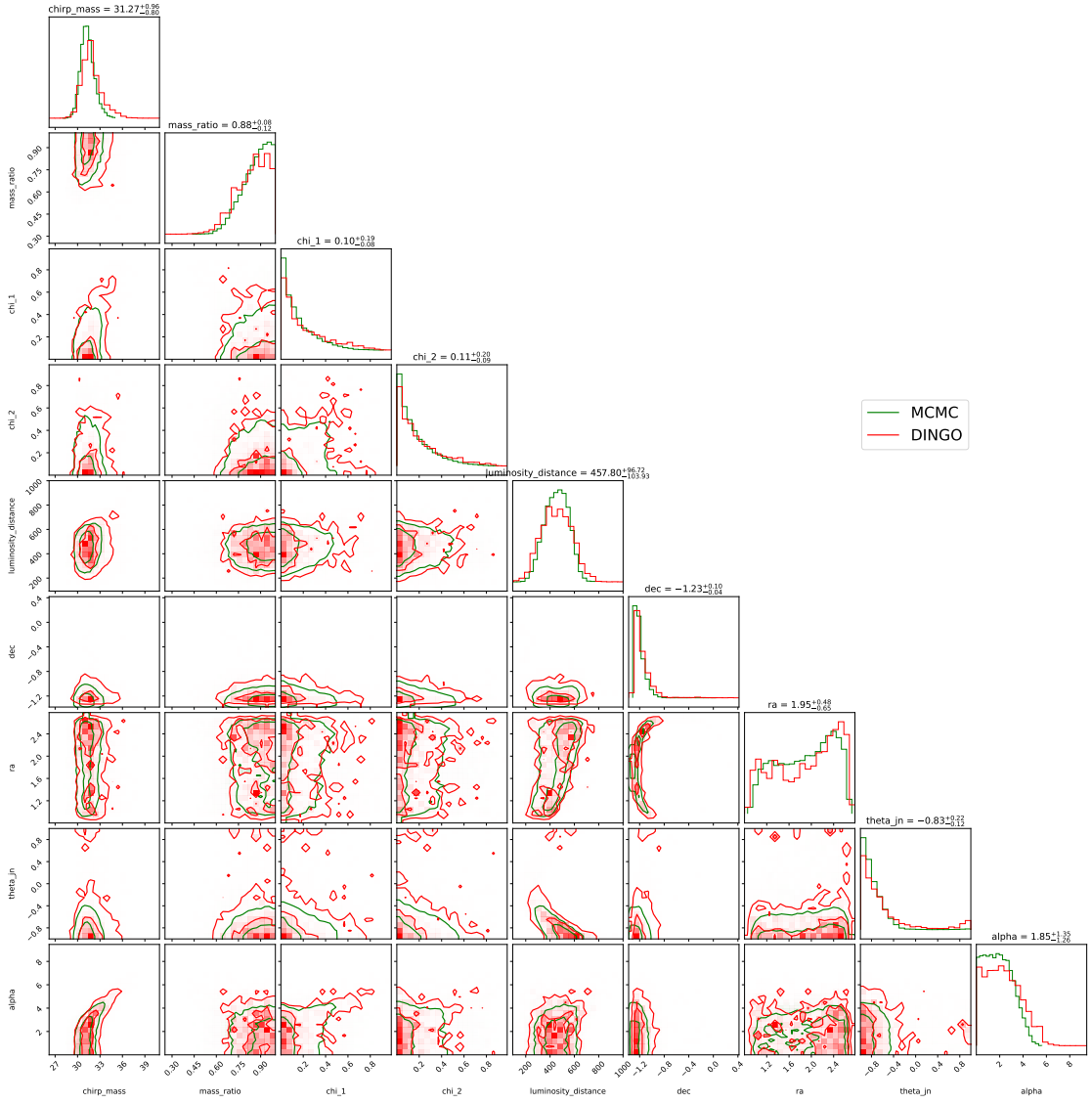

Figure B.3: Marginalized one-dimensional posterior distributions over a subset of parameters, comparing the normalizing flow (red) and bilby dynesty (green). Contours represent 50% and 90% credible regions. Neural network posteriors are constructed from 10^7 samples with 931 effective importance samplings.

Bibliography

[1] V. Alex, S. Finkbeiner, and J. Weber, "Temperature dependence of the indirect energy gap in crystalline silicon," *Journal of Applied Physics*, vol. 79, no. 9, pp. 6943–6946, 05 1996. [Online]. Available: https://doi.org/10.1063/1.362447

[2] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, "Neural spline flows," 2019," 1906.04032. [Online]. Available: https://arxiv.org/abs/1906.04032

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," 2016," 1603.05027. [Online]. Available: https://arxiv.org/abs/1603.05027

[4] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015," 1502.03167. [Online]. Available: https://arxiv.org/abs/1502.03167

[5] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," 2016," 1511.07289. [Online]. Available: https://arxiv.org/abs/1511.07289

[6] M. Uehara, S. Mori, C. Chen, and S. Cheong, "Percolative phase separation underlies colossal magnetoresistance in mixed-valent manganites," *Nature*, vol. 399, no. 6736, pp. 560–563, Jun. 1999.

[7] T. Miao, L. Deng, W. Yang, J. Ni, C. Zheng, J. Etheridge, S. Wang, H. Liu, H. Lin, Y. Yu, Q. Shi, P. Cai, Y. Zhu, T. Yang, X. Zhang, X. Gao, C. Xi, M. Tian, X. Wu, H. Xiang, E. Dagotto, L. Yin, and J. Shen, "Direct experimental evidence of physical origin of electronic phase separation in manganites," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 13, pp. 7090–7094, Mar. 2020.

[8] T. Hanaguri, C. Lupien, Y. Kohsaka, D.-H. Lee, M. Azuma, M. Takano, H. Takagi, and J. C. Davis, "A 'checkerboard' electronic crystal state in lightly hole-doped ca2-xnaxcuo2cl2," *Nature*, vol. 430, no. 7003, pp. 1001–1005, Aug 2004. [Online]. Available: https://doi.org/10.1038/nature02861

[9] M. H. Hamidian, S. D. Edkins, S. H. Joo, A. Kostin, H. Eisaki, S. Uchida, M. J. Lawler, E.-A. Kim, A. P. Mackenzie, K. Fujita, J. Lee, and J. C. S. Davis, "Detection of a cooper-pair density wave in bi2sr2cacu2o8+x," *Nature*, vol. 532, no. 7599, p. 343–347, Apr. 2016. [Online]. Available: http://dx.doi.org/10.1038/nature17411

[10] A. Moreo, S. Yunoki, and E. Dagotto, "Phase separation scenario for manganese oxides and related materials," *Science*, vol. 283, no. 5410, pp. 2034–

2040, 1999, https://www.science.org/doi/pdf/10.1126/science.283.5410.2034. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.283.5410.2034

[11] N. Mathur and P. Littlewood, "Mesoscopic Texture in Manganites," *Physics Today*, vol. 56, no. 1, pp. 25–30, 01 2003, https://pubs.aip.org/physicstoday/article-pdf/56/1/25/16660308/25_1_online.pdf. [Online]. Available: https://doi.org/10.1063/1.1554133

[12] E. Dagotto, T. Hotta, and A. Moreo, "Colossal magnetoresistant materials: the key role of phase separation," *Physics Reports*, vol. 344, no. 1, pp. 1–153, 2001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0370157300001216

[13] V. J. Emery, S. A. Kivelson, and J. M. Tranquada, "Stripe phases in high-temperature superconductors," *Proceedings of the National Academy of Sciences*, vol. 96, no. 16, p. 8814–8817, Aug. 1999. [Online]. Available: http://dx.doi.org/10.1073/pnas.96.16.8814

[14] E. Dagotto, *Nanoscale Phase Separation and Colossal Magnetoresistance: The Physics of Manganites and Related Compounds*, ser. Springer Series in Solid-State Sciences. Springer, 2003. [Online]. Available: https://books.google.com/books?id=eEDipQjwipAC

[15] E. Dagotto, "Complexity in strongly correlated electronic systems," *Science*, vol. 309, no. 5732, p. 257–262, Jul. 2005. [Online]. Available: http://dx.doi.org/10.1126/science.1107559

[16] S. A. Kivelson, E. Fradkin, and V. J. Emery, "Electronic liquid-crystal phases of a doped mott insulator," *Nature*, vol. 393, no. 6685, p. 550–553, Jun. 1998. [Online]. Available: http://dx.doi.org/10.1038/31177

[17] S. A. Kivelson, I. P. Bindloss, E. Fradkin, V. Oganesyan, J. M. Tranquada, A. Kapitulnik, and C. Howald, "How to detect fluctuating stripes in the high-temperature superconductors," *Rev. Mod. Phys.*, vol. 75, pp. 1201–1241, Oct 2003. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.75.1201

[18] E. Fradkin, S. A. Kivelson, and J. M. Tranquada, "Colloquium: Theory of intertwined orders in high temperature superconductors," *Rev. Mod. Phys.*, vol. 87, pp. 457–482, May 2015. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.87.457

[19] A. N. Bogdanov and D. A. Yablonskii, "Thermodynamically stable "vortices" in magnetically ordered crystals: The mixed state of magnets," *Sov. Phys. JETP*, vol. 68, p. 101, 1989.

[20] U. K. Rößler, A. N. Bogdanov, and C. Pfleiderer, "Spontaneous skyrmion ground states in magnetic metals," *Nature*, vol. 442, no. 7104, pp. 797–801,

Aug 2006. [Online]. Available: https://doi.org/10.1038/nature05056

[21] S. Mühlbauer, B. Binz, F. Jonietz, C. Pfleiderer, A. Rosch, A. Neubauer, R. Georgii, and P. Böni, "Skyrmion lattice in a chiral magnet," *Science*, vol. 323, no. 5916, pp. 915–919, 2009. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.1166767

[22] X. Z. Yu, Y. Onose, N. Kanazawa, J. H. Park, J. H. Han, Y. Matsui, N. Nagaosa, and Y. Tokura, "Real-space observation of a two-dimensional skyrmion crystal," *Nature*, vol. 465, no. 7300, pp. 901–904, Jun 2010. [Online]. Available: https://doi.org/10.1038/nature09124

[23] X. Z. Yu, N. Kanazawa, Y. Onose, K. Kimoto, W. Z. Zhang, S. Ishiwata, Y. Matsui, and Y. Tokura, "Near room-temperature formation of a skyrmion crystal in thin-films of the helimagnet fege," *Nature Materials*, vol. 10, no. 2, pp. 106–109, Feb 2011. [Online]. Available: https://doi.org/10.1038/nmat2916

[24] S. Seki, X. Z. Yu, S. Ishiwata, and Y. Tokura, "Observation of skyrmions in a multiferroic material," *Science*, vol. 336, no. 6078, pp. 198–201, Apr 2012. [Online]. Available: https://doi.org/10.1126/science.1214143

[25] N. Nagaosa and Y. Tokura, "Topological properties and dynamics of magnetic skyrmions," *Nature Nanotechnology*, vol. 8, no. 12, pp. 899–911, Dec 2013. [Online]. Available: https://doi.org/10.1038/nnano.2013.243

[26] J. Bonca, S. A. Trugman, and I. Batisti'c, "Holstein polaron," *Phys. Rev. B*, vol. 60, p. 1633, 1999.

[27] D. Golez, J. Bonca, L. Vidmar, and S. A. Trugman, "Relaxation dynamics of the holstein polaron," *Phys. Rev. Lett.*, vol. 109, p. 236402, 2012.

[28] A. S. Mishchenko, N. Nagaosa, G. D. Filippis, A. de Candia, and V. Cataudella, "Mobility of holstein polaron at finite temperature: An unbiased approach," *Phys. Rev. Lett.*, vol. 114, p. 146401, 2015.

[29] Y.-X. Zhang, W.-T. Chiu, N. C. Costa, G. G. Batrouni, and R. T. Scalettar, "Charge order in the holstein model on a honeycomb lattice," *Phys. Rev. Lett.*, vol. 122, p. 077602, 2019.

[30] N. C. Costa, T. Blommel, W.-T. Chiu, G. Batrouni, and R. T. Scalettar, "Phonon dispersion and the competition between pairing and charge order," *Phys. Rev. Lett.*, vol. 120, p. 187003, 2018.

[31] F. Marsiglio, "Pairing and charge-density-wave correlations in the holstein model at half-filling," *Phys. Rev. B*, vol. 42, p. 2416, 1990.

[32] M. Vekic, R. M. Noack, and S. R. White, "Charge-density waves versus superconductivity in the holstein model with next-nearest-neighbor hopping," *Phys. Rev. B*, vol. 46, p. 271, 1992.

[33] H. Zheng and S. Y. Zhu, "Charge-density-wave and superconducting states in the holstein model on a square lattice," *Phys. Rev. B*, vol. 55, p. 3803, 1997.

[34] B. Cohen-Stead, K. Barros, Z. Meng, C. Chen, R. T. Scalettar, and G. G. Batrouni, "Langevin simulations of the half-filled cubic holstein model," *Phys. Rev. B*, vol. 102, p. 161108, 2020.

[35] Y. Oono and S. Puri, "Study of phase-separation dynamics by use of cell dynamical systems. i. modeling," *Phys. Rev. A*, vol. 38, pp. 434–453, Jul 1988. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.38.434

[36] J. D. Gunton, R. Toral, and A. Chakrabarti, "Numerical studies of phase separation in models of binary alloys and polymer blends," *Physica Scripta*, vol. 1990, no. T33, p. 12, jan 1990. [Online]. Available: https://dx.doi.org/10.1088/0031-8949/1990/T33/003

[37] R. Weinkamer, P. Fratzl, H. S. Gupta, O. Penrose, and J. L. Lebowitz, "Using kinetic monte carlo simulations to study phase separation in alloys," *Phase Transitions*, vol. 77, no. 5-7, pp. 433–456, 2004, https://doi.org/10.1080/01411590410001672639. [Online]. Available: https://doi.org/10.1080/01411590410001672639

[38] L.-Q. Chen, "Phase-field models for microstructure evolution," *Annual Review of Materials Research*, vol. 32, no. Volume 32, 2002, pp. 113–140, 2002.

[Online]. Available: https://www.annualreviews.org/content/journals/10.1146/annurev.matsci.32.112001.132041

[39] K. Hansen, G. Montavon, F. Biegler, S. Fazli, M. Rupp, M. Scheffler, O. A. von Lilienfeld, A. Tkatchenko, and K.-R. Müller, "Assessment and validation of machine learning methods for predicting molecular atomization energies," *Journal of Chemical Theory and Computation*, vol. 9, no. 8, pp. 3404–3419, 2013. [Online]. Available: https://doi.org/10.1021/ct400195d

[40] K. T. Schütt, H. Glawe, F. Brockherde, A. Sanna, K. R. Müller, and E. K. U. Gross, "How to represent crystal structures for machine learning: Towards fast prediction of electronic properties," *Phys. Rev. B*, vol. 89, p. 205118, May 2014. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.89.205118

[41] K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. von Lilienfeld, K.-R. Müller, and A. Tkatchenko, "Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space," *The Journal of Physical Chemistry Letters*, vol. 6, no. 12, pp. 2326–2331, 2015. [Online]. Available: https://doi.org/10.1021/acs.jpclett.5b00831

[42] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, "Quantum-chemical insights from deep tensor neural networks," *Nature Communications*, vol. 8, no. 1, p. 13890, Jan 2017. [Online]. Available: https://doi.org/10.1038/ncomms13890

[43] V. Dunjko and H. J. Briegel, "Machine learning & artificial intelligence in the quantum domain: a review of recent progress," *Rep. Prog. Phys.*, vol. 81, p. 074001, 2018.

[44] S. V. Kalinin, B. G. Sumpter, and R. K. Archibald, "Big-deep-smart data in imaging for guiding materials design," *Nat. Mater.*, vol. 14, p. 973, 2015.

[45] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborov'a, "Machine learning and the physical sciences," *Rev. Mod. Phys.*, vol. 91, p. 045002, 2019.

[46] A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao, and T. Wongjirad, "Machine learning at the energy and intensity frontiers of particle physics," *Nature*, vol. 560, p. 41, 2018.

[47] S. D. Sarma, D.-L. Deng, and L.-M. Duan, "Machine learning meets quantum physics," *Phys. Today*, vol. 72, p. 48, 2019.

[48] D. Baron, "Machine learning in astronomy: a practical overview," *arXiv*, 2019, 1904.07248.

[49] D. Morgan and R. Jacobs, "Opportunities and challenges for machine learning in materials science," *Annu. Rev. Mater. Res.*, vol. 50, p. 71, 2020.

[50] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, "Machine learning for molecular and materials science," *Nature*, vol. 559, p. 547, 2018.

[51] M. Meuwly, "Machine learning for chemical reactions," *Chem. Rev.*, vol. 121, p. 10218, 2021.

[52] J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. M"uller, and A. Tkatchenko, "Combining machine learning and computational chemistry for predictive insights into chemical systems," *Chem. Rev.*, vol. 121, p. 9816, 2021.

[53] J. G. Greener, S. M. Kandathil, L. Moffat, and D. T. Jones, "A guide to machine learning for biologists," *Nat. Rev. Mol.*, vol. 23, p. 40, 2022.

[54] E. Bedolla, L. C. Padierna, and R. Castaneda-Priego, "Machine learning for condensed matter physics," *J. Phys.: Condens. Matter*, vol. 33, p. 053001, 2021.

[55] M. W. Libbrecht and W. S. Noble, "Machine learning applications in genetics and genomics," *Nat. Rev. Genetics*, vol. 16, p. 321, 2015.

[56] A. Ourmazd, "Science in the age of machine learning," *Nat. Rev. Phys.*, vol. 2, p. 342, 2020.

[57] L. M. Falicov and J. C. Kimball, "Simple model for semiconductor-metal transitions: Smb_6 and transition-metal oxides," *Phys. Rev. Lett.*, vol. 22, pp. 997–999, May 1969. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.22.997

[58] P. Zhang, S. Zhang, and G.-W. Chern, "Descriptors for machine learning model of generalized force field in condensed matter systems," *arXiv preprint arXiv:2201.00798*, 2022.

[59] S. Zhang, P. Zhang, and G.-W. Chern, "Anomalous phase separation in a correlated electron system: Machine-learning–enabled large-scale kinetic monte carlo simulations," *Proceedings of the National Academy of Sciences*, vol. 119, no. 18, p. e2119957119, 2022.

[60] P. B. Visscher, "Phase separation instability in the hubbard model," *Phys. Rev. B*, vol. 10, pp. 943–945, Aug 1974. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.10.943

[61] Schulz, H.J., "Domain walls in a doped antiferromagnet," *J. Phys. France*, vol. 50, no. 18, pp. 2833–2849, 1989. [Online]. Available: https://doi.org/10.1051/jphys:0198900500180283300

[62] V. J. Emery, S. A. Kivelson, and H. Q. Lin, "Phase separation in the t-j model," *Phys. Rev. Lett.*, vol. 64, pp. 475–478, Jan 1990. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.64.475

[63] J. Zaanen and O. Gunnarsson, "Charged magnetic domain lines and the magnetism of high-T_c oxides," *Phys. Rev. B*, vol. 40, pp. 7391–7394, Oct 1989. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.40.7391

[64] M. Kato, K. Machida, H. Nakanishi, and M. Fujita, "Soliton Lattice Modulation of Incommensurate Spin Density Wave in Two Dimensional Hubbard Model —A Mean Field Study—," *Journal of the Physical Society of Japan*, vol. 59, no. 3, p. 1047, Mar. 1990.

[65] L. Gehlhoff, "Phase separation in the one-band hubbard model," *Journal of Physics: Condensed Matter*, vol. 8, no. 16, p. 2851, apr 1996. [Online]. Available: https://dx.doi.org/10.1088/0953-8984/8/16/014

[66] S. R. White and D. J. Scalapino, "Phase separation and stripe formation in the two-dimensional $t - j$ model: A comparison of numerical results," *Phys. Rev. B*, vol. 61, pp. 6320–6326, Mar 2000. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.61.6320

[67] C.-H. Yee and L. Balents, "Phase separation in doped mott insulators," *Phys. Rev. X*, vol. 5, p. 021007, Apr 2015. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevX.5.021007

[68] J. D. Gunton, M. San Miguel, and P. S. Sahni, "The dynamics of first order phase transitions," *Phase Transitions and Critical Phenomena*, vol. 8, pp. 269–446, 1983.

[69] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics*. Cambridge University Press, 2010.

[70] A. Bray, "Theory of phase-ordering kinetics," *Advances in Physics*, vol. 43, no. 3, p. 357–459, Jun. 1994. [Online]. Available: http://dx.doi.org/10.1080/00018739400101505

[71] S. Puri, "Kinetics of phase transitions," *Phase Transitions*, vol. 77, no. 5-7, pp. 407–431, 2004, https://doi.org/10.1080/01411590410001672648. [Online]. Available: https://doi.org/10.1080/01411590410001672648

[72] A. Onuki, *Phase Transition Dynamics*. Cambridge University Press, 2002.

[73] P. C. Hohenberg and B. I. Halperin, "Theory of dynamic critical phenomena," *Rev. Mod. Phys.*, vol. 49, pp. 435–479, Jul 1977. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.49.435

[74] H. Furukawa, "A dynamic scaling assumption for phase separation," *Advances in Physics*, vol. 34, no. 6, pp. 703–750, 1985, https://doi.org/10.1080/00018738500101841. [Online]. Available: https://doi.org/10.1080/00018738500101841

[75] J. Hubbard, "Electron correlations in narrow energy bands," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 276, no. 1365, pp. 238–257, 1963. [Online]. Available: http://www.jstor.org/stable/2414761

[76] J. K. Freericks and V. Zlatić, "Exact dynamical mean-field theory of the falicov-kimball model," *Rev. Mod. Phys.*, vol. 75, pp. 1333–1382, Oct 2003. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.75.1333

[77] T. Kennedy and E. H. Lieb, "An itinerant electron model with crystalline or magnetic long range order," *Physica A: Statistical Mechanics and its Applications*, vol. 138, no. 1, pp. 320–358, 1986. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0378437186901883

[78] J. K. Freericks and R. Lemański, "Segregation and charge-density-wave order in the spinless falicov-kimball model," *Physical Review B*, vol. 61, no. 20, p. 13438–13444, May 2000. [Online]. Available: http://dx.doi.org/10.1103/PhysRevB.61.13438

[79] J. K. Freericks, E. H. Lieb, and D. Ueltschi, "Phase separation due to quantum mechanical correlations," *Physical Review Letters*, vol. 88, no. 10, Feb. 2002. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.88.106401

[80] G. I. Watson and R. Lemanski, "The ground-state phase diagram of the two-dimensional falicov-kimball model," *Journal of Physics: Condensed Matter*, vol. 7, no. 49, p. 9521, dec 1995. [Online]. Available: https://dx.doi.org/10.1088/0953-8984/7/49/016

[81] R. Lemański, J. K. Freericks, and G. Banach, "Stripe phases in the two-dimensional falicov-kimball model," *Phys. Rev. Lett.*, vol. 89, p. 196403,

Oct 2002. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.89.196403

[82] M.-T. Tran, "Inhomogeneous phases in the falicov-kimball model: Dynamical mean-field approximation," *Phys. Rev. B*, vol. 73, p. 205110, May 2006. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.73.205110

[83] M. M. Maśka and K. Czajka, "Pattern formation in the falicov–kimball model," *physica status solidi (b)*, vol. 242, no. 2, pp. 479–483, 2005, https://onlinelibrary.wiley.com/doi/pdf/10.1002/pssb.200460067. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/pssb.200460067

[84] M. M. Maśka and K. Czajka, "Thermodynamics of the two-dimensional falicov-kimball model: A classical monte carlo study," *Phys. Rev. B*, vol. 74, p. 035109, Jul 2006. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.74.035109

[85] A. E. Antipov, Y. Javanmard, P. Ribeiro, and S. Kirchner, "Interaction-tuned anderson versus mott localization," *Phys. Rev. Lett.*, vol. 117, p. 146601, Sep 2016. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.117.146601

[86] G. Barkema, "Monte carlo simulations of domain growth," *Kinetics of Phase Transitions*, 03 2009.

[87] D. Marx and J. Hutter, *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge University Press, 2009.

[88] P. Fratzl and R. Weinkamer, *Phase Separation in Binary Alloys - Modeling Approaches*. Vienna: Springer Vienna, 2005, pp. 57–116. [Online]. Available: https://doi.org/10.1007/3-211-27404-9_2

[89] A. B. Bortz, M. H. Kalos, J. L. Lebowitz, and M. A. Zendejas, "Time evolution of a quenched binary alloy: Computer simulation of a two-dimensional model system," *Phys. Rev. B*, vol. 10, pp. 535–541, Jul 1974. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.10.535

[90] A. Weiße, G. Wellein, A. Alvermann, and H. Fehske, "The kernel polynomial method," *Rev. Mod. Phys.*, vol. 78, pp. 275–306, Mar 2006. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.78.275

[91] J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," *Phys. Rev. Lett.*, vol. 98, p. 146401, Apr 2007. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.98.146401

[92] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, "Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons," *Phys. Rev. Lett.*, vol. 104, p. 136403, Apr 2010. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.104.136403

[93] L. Zhang, J. Han, H. Wang, R. Car, and W. E, "Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics," *Phys. Rev. Lett.*, vol. 120, p. 143001, Apr 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.120.143001

[94] H. Suwa, J. S. Smith, N. Lubbers, C. D. Batista, G.-W. Chern, and K. Barros, "Machine learning for molecular dynamics with strongly correlated electrons," *Phys. Rev. B*, vol. 99, p. 161107, Apr 2019. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.99.161107

[95] F. Noé, A. Tkatchenko, K.-R. Müller, and C. Clementi, "Machine learning for molecular simulation," *Annual Review of Physical Chemistry*, vol. 71, no. Volume 71, 2020, pp. 361–390, 2020. [Online]. Available: https://www.annualreviews.org/content/journals/10.1146/annurev-physchem-042018-052331

[96] Z. Li, J. R. Kermode, and A. De Vita, "Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces," *Phys. Rev. Lett.*, vol. 114, p. 096405, Mar 2015. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.114.096405

[97] J. S. Smith, O. Isayev, and A. E. Roitberg, "Ani-1: an extensible neural network potential with dft accuracy at force field computational

cost," *Chem. Sci.*, vol. 8, pp. 3192–3203, 2017. [Online]. Available: http://dx.doi.org/10.1039/C6SC05720A

[98] P. Zhang, P. Saha, and G.-W. Chern, "Machine learning dynamics of phase separation in correlated electron magnets," 2020," 2006.04205. [Online]. Available: https://arxiv.org/abs/2006.04205

[99] P. Zhang and G.-W. Chern, "Arrested phase separation in double-exchange models: Large-scale simulation enabled by machine learning," *Phys. Rev. Lett.*, vol. 127, p. 146401, Sep 2021. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.127.146401

[100] W. Kohn, "Density functional and density matrix method scaling linearly with the number of atoms," *Phys. Rev. Lett.*, vol. 76, pp. 3168–3171, Apr 1996. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.76.3168

[101] E. Prodan and W. Kohn, "Nearsightedness of electronic matter," *Proceedings of the National Academy of Sciences*, vol. 102, no. 33, pp. 11 635–11 638, 2005, https://www.pnas.org/doi/pdf/10.1073/pnas.0505436102. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.0505436102

[102] M. Hamermesh, *Group Theory and Its Application to Physical Problems*, ser. Addison Wesley Series in Physics. Dover Publications, 1989. [Online]. Available: https://books.google.com/books?id=c0o9_wlCzgcC

[103] R. Kondor, "A complete set of rotationally and translationally invariant features for images," *CoRR*, vol. abs/cs/0701127, 2007, cs/0701127. [Online]. Available: http://arxiv.org/abs/cs/0701127

[104] A. P. Bartók, R. Kondor, and G. Csányi, "On representing chemical environments," *Phys. Rev. B*, vol. 87, p. 184115, May 2013. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.87.184115

[105] J. Ma, P. Zhang, Y. Tan, A. W. Ghosh, and G.-W. Chern, "Machine learning electron correlation in a disordered medium," *Phys. Rev. B*, vol. 99, p. 085118, Feb 2019. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.99.085118

[106] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf

[107] J. Schmidhuber, "Deep learning in neural networks: An overview," *CoRR*, vol. abs/1404.7828, 2014, 1404.7828. [Online]. Available: http://arxiv.org/abs/1404.7828

[108] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015. [Online]. Available: http://arxiv.org/abs/1412.6980

[109] I. Lifshitz and V. Slyozov, "The kinetics of precipitation from supersaturated solid solutions," *Journal of Physics and Chemistry of Solids*, vol. 19, no. 1, pp. 35–50, 1961. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0022369761900543

[110] C. Wagner, "Theorie der alterung von niederschlägen durch umlösen (ostwald-reifung)," *Zeitschrift für Elektrochemie, Berichte der Bunsengesellschaft für physikalische Chemie*, vol. 65, no. 7-8, pp. 581–591, 1961, https://onlinelibrary.wiley.com/doi/pdf/10.1002/bbpc.19610650704. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/bbpc.19610650704

[111] J. W. Cahn and J. E. Hilliard, "Free Energy of a Nonuniform System. I. Interfacial Free Energy," *The Journal of Chemical Physics*, vol. 28, no. 2, pp. 258–267, 02 1958, https://pubs.aip.org/aip/jcp/article-

pdf/28/2/258/18813541/258_1_online.pdf. [Online]. Available: https://doi.org/10.1063/1.1744102

[112] D. A. Huse, "Corrections to late-stage behavior in spinodal decomposition: Lifshitz-slyozov scaling and monte carlo simulations," *Phys. Rev. B*, vol. 34, pp. 7845–7850, Dec 1986. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.34.7845

[113] S. M. Allen and J. W. Cahn, "A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening," *Acta Metallurgica*, vol. 27, no. 6, pp. 1085–1095, 1979. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0001616079901962

[114] J. D. Shore, M. Holzer, and J. P. Sethna, "Logarithmically slow domain growth in nonrandomly frustrated systems: Ising models with competing interactions," *Physical Review B*, vol. 46, no. 18, p. 11376–11404, Nov. 1992. [Online]. Available: http://dx.doi.org/10.1103/PhysRevB.46.11376

[115] M. R. Evans, "Anomalous coarsening and glassy dynamics," *Journal of Physics: Condensed Matter*, vol. 14, no. 7, p. 1397–1422, Feb. 2002. [Online]. Available: http://dx.doi.org/10.1088/0953-8984/14/7/302

[116] H. Tanaka, "Viscoelastic phase separation," *Journal of Physics: Condensed Matter*, vol. 12, p. R207, 03 2000.

[117] M. Zannetti, "Aging in domain growth," 2014," 1412.4670. [Online]. Available: https://arxiv.org/abs/1412.4670

[118] F. Corberi, "Coarsening in inhomogeneous systems," *Comptes Rendus. Physique*, vol. 16, no. 3, p. 332–342, Apr. 2015. [Online]. Available: http://dx.doi.org/10.1016/j.crhy.2015.03.019

[119] D. Loss and P. M. Goldbart, "Persistent currents from berry's phase in mesoscopic systems," *Phys. Rev. B*, vol. 45, pp. 13 544–13 561, Jun 1992. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.45.13544

[120] J. Ye, Y. B. Kim, A. J. Millis, B. I. Shraiman, P. Majumdar, and Z. Tešanović, "Berry phase theory of the anomalous hall effect: Application to colossal magnetoresistance manganites," *Phys. Rev. Lett.*, vol. 83, pp. 3737–3740, Nov 1999. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.83.3737

[121] G. Tatara and H. Kawamura, "Chirality-driven anomalous hall effect in weak coupling regime," *Journal of the Physical Society of Japan*, vol. 71, no. 11, pp. 2613–2616, 2002. [Online]. Available: https://doi.org/10.1143/JPSJ.71.2613

[122] M. V. Berry, "Quantal phase factors accompanying adiabatic changes," *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, vol. 392, no. 1802, pp. 45–57, 1984.

[123] R. F. L. Evans, W. J. Fan, P. Chureemart, T. A. Ostler, M. O. A. Ellis, and R. W. Chantrell, "Atomistic spin model simulations of magnetic nanomaterials," *Journal of Physics: Condensed Matter*, vol. 26, no. 10, p. 103202, feb 2014.

[124] A. Vansteenkiste, J. Leliaert, M. Dvornik, M. Helsen, F. Garcia-Sanchez, and B. Van Waeyenberge, "The design and verification of MuMax3," *AIP Advances*, vol. 4, no. 10, 10 2014, 107133.

[125] V. Botu, R. Batra, J. Chapman, and R. Ramprasad, "Machine learning force fields: construction, validation, and outlook," *The Journal of Physical Chemistry C*, vol. 121, no. 1, pp. 511–522, 2017.

[126] Y. Li, H. Li, F. C. Pickard IV, B. Narayanan, F. G. Sen, M. K. Chan, S. K. Sankaranarayanan, B. R. Brooks, and B. Roux, "Machine learning force field parameters from ab initio data," *Journal of chemical theory and computation*, vol. 13, no. 9, pp. 4492–4503, 2017.

[127] J. Behler, "Perspective: Machine learning potentials for atomistic simulations," *The Journal of chemical physics*, vol. 145, no. 17, p. 170901, 2016.

[128] V. L. Deringer, M. A. Caro, and G. Csányi, "Machine learning interatomic potentials as emerging tools for materials science," *Advanced Materials*, vol. 31, no. 46, p. 1902765, 2019.

[129] R. T. McGibbon, A. G. Taube, A. G. Donchev, K. Siva, F. Hernández, C. Hargus, K.-H. Law, J. L. Klepeis, and D. E. Shaw, "Improving the accuracy of

møller-plesset perturbation theory with neural networks," *The Journal of chemical physics*, vol. 147, no. 16, p. 161725, 2017.

[130] T. Mueller, A. Hernandez, and C. Wang, "Machine learning for interatomic potential models," *The Journal of chemical physics*, vol. 152, no. 5, p. 050902, 2020.

[131] Y.-H. Liu, S. Zhang, P. Zhang, T.-K. Lee, and G.-W. Chern, "Machine learning predictions for local electronic properties of disordered correlated electron systems," *Physical Review B*, vol. 106, no. 3, p. 035131, 2022.

[132] P. Zhang and G.-W. Chern, "Machine learning nonequilibrium electron forces for spin dynamics of itinerant magnets," *npj Computational Materials*, vol. 9, no. 1, p. 32, 2023.

[133] M. A. Brännvall, D. Gambino, R. Armiento, and B. Alling, "Machine learning approach for longitudinal spin fluctuation effects in bcc fe at t c and under earth-core conditions," *Physical Review B*, vol. 105, no. 14, p. 144417, 2022.

[134] I. Novikov, B. Grabowski, F. Körmann, and A. Shapeev, "Magnetic moment tensor potentials for collinear spin-polarized materials reproduce different magnetic states of bcc fe," *npj Computational Materials*, vol. 8, no. 1, p. 13, 2022.

[135] W. F. Brown Jr, "Thermal fluctuations of a single-domain particle," *Physical review*, vol. 130, no. 5, p. 1677, 1963.

[136] M. A. Ruderman and C. Kittel, "Indirect exchange coupling of nuclear magnetic moments by conduction electrons," *Physical Review*, vol. 96, no. 1, p. 99, 1954.

[137] T. Kasuya, "A theory of metallic ferro-and antiferromagnetism on zener's model," *Progress of theoretical physics*, vol. 16, no. 1, pp. 45–57, 1956.

[138] K. Yosida, "Magnetic properties of cu-mn alloys," *Physical Review*, vol. 106, no. 5, p. 893, 1957.

[139] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.

[140] R. Ozawa, S. Hayami, and Y. Motome, "Zero-field skyrmions with a high topological number in itinerant magnets," *Phys. Rev. Lett.*, vol. 118, p. 147205, Apr 2017. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.118.147205

[141] Z. Wang, G.-W. Chern, C. D. Batista, and K. Barros, "Gradient-based stochastic estimation of the density matrix," *The Journal of Chemical Physics*, vol. 148, no. 9, 03 2018. [Online]. Available: https://doi.org/10.1063/1.5017741

[142] J. H. Mentink, M. V. Tretyakov, A. Fasolino, M. I. Katsnelson, and T. Rasing, "Stable and fast semi-implicit integration of the stochastic landau–lifshitz equation," *Journal of Physics: Condensed Matter*, vol. 22, no. 17, p. 176001, apr 2010. [Online]. Available: https://dx.doi.org/10.1088/0953-8984/22/17/176001

[143] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017," 1412.6980.

[144] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Z. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," *CoRR*, vol. abs/1912.01703, 2019, 1912.01703. [Online]. Available: http://arxiv.org/abs/1912.01703

[145] R. Ozawa, S. Hayami, K. Barros, G.-W. Chern, Y. Motome, and C. D. Batista, "Vortex crystals with chiral stripes in itinerant magnets," *Journal of the Physical Society of Japan*, vol. 85, no. 10, p. 103703, 2016. [Online]. Available: https://doi.org/10.7566/JPSJ.85.103703

[146] S. Hayami and Y. Motome, "Topological spin crystals by itinerant frustration," *Journal of Physics: Condensed Matter*, vol. 33, no. 44, p. 443001, aug 2021. [Online]. Available: https://dx.doi.org/10.1088/1361-648X/ac1a30

[147] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021.

[148] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: going beyond euclidean data," *IEEE Signal Processing*

Magazine, vol. 34, no. 4, pp. 18–42, 2017.

[149] S. Chen, E. Dobriban, and J. H. Lee, "A group-theoretic framework for data augmentation," *The Journal of Machine Learning Research*, vol. 21, no. 1, pp. 9885–9955, 2020.

[150] P. C. Nguyen, Y.-T. Nguyen, J. B. Choi, P. K. Seshadri, H. Udaykumar, and S. Baek, "PARC: Physics-aware recurrent convolutional neural networks to assimilate meso scale reactive mechanics of energetic materials," *Science Advances*, vol. 9, no. 17, p. eadd6868, 4 2023.

[151] P. C. Nguyen, J. B. Choi, H. Udaykumar, and S. Baek, "Challenges and opportunities for machine learning in multiscale computational modeling," *ASME Journal of Computing and Information Science in Engineering*, 4 2023.

[152] P. C. Nguyen, Y.-T. Nguyen, P. K. Seshadri, J. B. Choi, H. Udaykumar, and S. Baek, "A physics-aware deep learning model for energy localization in multiscale shock-to-detonation simulations of heterogeneous energetic materials," *Propellants, Explosives, Pyrotechnics*, vol. 48, no. 4, p. e202200268, 3 2023.

[153] A. László, "Unification mechanism for gauge and spacetime symmetries," *Journal of Physics A: Mathematical and Theoretical*, vol. 50, no. 11, p. 115401, 2017.

[154] T. Dao, A. Gu, A. Ratner, V. Smith, C. De Sa, and C. Ré, "A kernel theory of modern data augmentation," in *International Conference on Machine Learning*. PMLR, 2019, pp. 1528–1537.

[155] R. Wang, R. Walters, and R. Yu, "Data augmentation vs. equivariant networks: A theory of generalization on dynamics forecasting," *arXiv preprint arXiv:2206.09450*, 2022.

[156] W. Luo, Y. Li, R. Urtasun, and R. S. Zemel, "Understanding the effective receptive field in deep convolutional neural networks," *CoRR*, vol. abs/1701.04128, 2017, 1701.04128. [Online]. Available: http://arxiv.org/abs/1701.04128

[157] M. Geiger and T. Smidt, "e3nn: Euclidean neural networks," 2022," 2207.09453.

[158] M. Mezard, G. Parisi, and M. Virasoro, *Spin Glass Theory And Beyond: An Introduction To The Replica Method And Its Applications*, ser. World Scientific Lecture Notes In Physics. World Scientific Publishing Company, 1987. [Online]. Available: https://books.google.com/books?id=DwY8DQAAQBAJ

[159] K. Binder and A. P. Young, "Spin glasses: Experimental facts, theoretical concepts, and open questions," *Rev. Mod. Phys.*, vol. 58, pp. 801–976, Oct 1986. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.58.801

[160] J. A. Mydosh, *Spin Glasses: An Experimental Introduction*. CRC Press, 1993.

[161] J. A. Mydosh, "Spin glasses: redux: an updated experimental/materials survey," *Reports on Progress in Physics*, vol. 78, no. 5, p. 052501, apr 2015. [Online]. Available: https://dx.doi.org/10.1088/0034-4885/78/5/052501

[162] J. C. Dyre, "Colloquium: The glass transition and elastic models of glass-forming liquids," *Rev. Mod. Phys.*, vol. 78, pp. 953–972, Sep 2006. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.78.953

[163] P. Charbonneau, J. Kurchan, G. Parisi, P. Urbani, and F. Zamponi, "Glass and jamming transitions: From exact results to finite-dimensional descriptions," *Annual Review of Condensed Matter Physics*, vol. 8, no. Volume 8, 2017, pp. 265–288, 2017. [Online]. Available: https://www.annualreviews.org/content/journals/10.1146/annurev-conmatphys-031016-025334

[164] S. F. Edwards and P. W. Anderson, "Theory of spin glasses," *Journal of Physics F: Metal Physics*, vol. 5, no. 5, p. 965, may 1975. [Online]. Available: https://dx.doi.org/10.1088/0305-4608/5/5/017

[165] D. Sherrington and S. Kirkpatrick, "Solvable model of a spin-glass," *Phys. Rev. Lett.*, vol. 35, pp. 1792–1796, Dec 1975. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.35.1792

[166] P. W. A. D. J. Thouless and R. G. Palmer, "Solution of 'solvable model of a spin glass'," *The Philosophical Magazine: A Journal of Theoretical Experimental and Applied Physics*, vol. 35, no. 3, pp. 593–

601, 1977, https://doi.org/10.1080/14786437708235992. [Online]. Available: https://doi.org/10.1080/14786437708235992

[167] G. Parisi, "Infinite number of order parameters for spin-glasses," *Phys. Rev. Lett.*, vol. 43, pp. 1754–1756, Dec 1979. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.43.1754

[168] G. Parisi, "Order parameter for spin-glasses," *Phys. Rev. Lett.*, vol. 50, pp. 1946–1948, Jun 1983. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.50.1946

[169] M. Mézard, G. Parisi, N. Sourlas, G. Toulouse, and M. Virasoro, "Nature of the spin-glass phase," *Phys. Rev. Lett.*, vol. 52, pp. 1156–1159, Mar 1984. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.52.1156

[170] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction*. Oxford University Press, 07 2001. https://academic.oup.com/book/5185/book-pdf/54038185/acprof-9780198509400.pdf. [Online]. Available: https://doi.org/10.1093/acprof:oso/9780198509417.001.0001

[171] D. L. Stein and C. M. Newman, *Spin Glasses and Complexity*. Princeton University Press, 2013. [Online]. Available: http://www.jstor.org/stable/j.ctt12f4hf

[172] P. Charbonneau, E. Marinari, M. Mézard, G. Parisi, F. Ricci-Tersenghi, G. Sicuro, and F. Zamponi, *Spin Glass Theory and Far Beyond*. WORLD SCIENTIFIC, 2023. https://www.worldscientific.com/doi/pdf/10.1142/13341. [Online]. Available: https://www.worldscientific.com/doi/abs/10.1142/13341

[173] V. Cannella and J. A. Mydosh, "Magnetic ordering in gold-iron alloys," *Phys. Rev. B*, vol. 6, pp. 4220–4237, Dec 1972. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.6.4220

[174] F. J. Lamelas, S. A. Werner, S. M. Shapiro, and J. A. Mydosh, "Intrinsic spin-density-wave magnetism in cu-mn alloys," *Phys. Rev. B*, vol. 51, pp. 621–624, Jan 1995. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.51.621

[175] A. V. Shapeev, "Moment tensor potentials: A class of systematically improvable interatomic potentials," *Multiscale Modeling & Simulation*, vol. 14, no. 3, pp. 1153–1173, 2016, https://doi.org/10.1137/15M1054183. [Online]. Available: https://doi.org/10.1137/15M1054183

[176] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, "Machine learning of accurate energy-conserving molecular force fields," *Science Advances*, vol. 3, no. 5, p. e1603015, 2017, https://www.science.org/doi/pdf/10.1126/sciadv.1603015. [Online]. Available: https://www.science.org/doi/abs/10.1126/sciadv.1603015

[177] S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko, "Towards exact molecular dynamics simulations with machine-learned force fields," *Nature Communications*, vol. 9, no. 1, p. 3887, Sep 2018. [Online]. Available: https://doi.org/10.1038/s41467-018-06169-2

[178] H. E. Sauceda, M. Gastegger, S. Chmiela, K.-R. Müller, and A. Tkatchenko, "Molecular force fields with gradient-domain machine learning (GDML): Comparison and synergies with classical force fields," *The Journal of Chemical Physics*, vol. 153, no. 12, p. 124109, 09 2020, https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0023005/14754092/124109_1_online.pdf. [Online]. Available: https://doi.org/10.1063/5.0023005

[179] C. Cheng, S. Zhang, and G.-W. Chern, "Machine learning for phase ordering dynamics of charge density waves," *Phys. Rev. B*, vol. 108, p. 014301, Jul 2023. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.108.014301

[180] X. Cheng, S. Zhang, P. C. H. Nguyen, S. Azarfar, G.-W. Chern, and S. S. Baek, "Convolutional neural networks for large-scale dynamical modeling of itinerant magnets," *Phys. Rev. Res.*, vol. 5, p. 033188, Sep 2023. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevResearch.5.033188

[181] J. Behler, "Atom-centered symmetry functions for constructing high-dimensional neural network potentials," *The Journal of Chemical Physics*,

vol. 134, no. 7, p. 074106, 02 2011, https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.3553717/15435271/074106_1_online.pdf. [Online]. Available: https://doi.org/10.1063/1.3553717

[182] L. M. Ghiringhelli, J. Vybiral, S. V. Levchenko, C. Draxl, and M. Scheffler, "Big data of materials science: Critical role of the descriptor," *Phys. Rev. Lett.*, vol. 114, p. 105503, Mar 2015. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.114.105503

[183] R. Drautz, "Atomic cluster expansion for accurate and transferable interatomic potentials," *Phys. Rev. B*, vol. 99, p. 014104, Jan 2019. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.99.014104

[184] L. Himanen, M. O. Jäger, E. V. Morooka, F. Federici Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke, and A. S. Foster, "Dscribe: Library of descriptors for machine learning in materials science," *Computer Physics Communications*, vol. 247, p. 106949, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010465519303042

[185] H. Huo and M. Rupp, "Unified representation of molecules and crystals for machine learning," *Machine Learning: Science and Technology*, vol. 3, no. 4, p. 045017, nov 2022. [Online]. Available: https://dx.doi.org/10.1088/2632-2153/aca005

[186] B. Skubic, O. E. Peil, J. Hellsvik, P. Nordblad, L. Nordström, and O. Eriksson, "Atomistic spin dynamics of the cu-mn spin-glass alloy," *Phys. Rev. B*, vol. 79, p. 024411, Jan 2009. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.79.024411

[187] J. Hellsvik, "Atomistic spin dynamics simulations on mn-doped gaas and cumn," *Journal of Physics: Conference Series*, vol. 200, no. 7, p. 072040, jan 2010. [Online]. Available: https://dx.doi.org/10.1088/1742-6596/200/7/072040

[188] O. E. Peil, A. V. Ruban, and B. Johansson, "Detailed ab initio calculations of the structure and magnetic state of a metallic spin glass," *New Journal of Physics*, vol. 10, no. 8, p. 083026, aug 2008. [Online]. Available: https://dx.doi.org/10.1088/1367-2630/10/8/083026

[189] M. F. Ling, J. B. Staunton, and D. D. Johnson, "A 'first-principles' theory for magnetic correlations and atomic short-range order in paramagnetic alloys. ii. application to cumn," *Journal of Physics: Condensed Matter*, vol. 6, no. 30, p. 6001, jul 1994. [Online]. Available: https://dx.doi.org/10.1088/0953-8984/6/30/017

[190] A. V. Ruban, S. Shallcross, S. I. Simak, and H. L. Skriver, "Atomic and magnetic configurational energetics by the generalized perturbation method," *Phys. Rev. B*, vol. 70, p. 125115, Sep 2004. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.70.125115

[191] P. W. Anderson, "Localized magnetic states in metals," *Phys. Rev.*, vol. 124, pp. 41–53, Oct 1961. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.124.41

[192] R. Drautz, "Atomic cluster expansion of scalar, vectorial, and tensorial properties including magnetism and charge transfer," *Phys. Rev. B*, vol. 102, p. 024104, Jul 2020. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.102.024104

[193] M. Eckhoff and J. Behler, "High-dimensional neural network potentials for magnetic systems using spin-dependent atom-centered symmetry functions," *npj Computational Materials*, vol. 7, no. 1, p. 170, Oct 2021. [Online]. Available: https://doi.org/10.1038/s41524-021-00636-z

[194] M. Domina, M. Cobelli, and S. Sanvito, "Spectral neighbor representation for vector fields: Machine learning potentials including spin," *Phys. Rev. B*, vol. 105, p. 214439, Jun 2022. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.105.214439

[195] J. B. J. Chapman and P.-W. Ma, "A machine-learned spin-lattice potential for dynamic simulations of defective magnetic iron," *Scientific Reports*, vol. 12, no. 1, p. 22451, Dec 2022. [Online]. Available: https://doi.org/10.1038/s41598-022-25682-5

[196] J. Tersoff, "New empirical approach for the structure and energy of covalent systems," *Phys. Rev. B*, vol. 37, pp. 6991–7000, Apr 1988. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.37.6991

[197] D. W. Brenner, "Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films," *Phys. Rev. B*, vol. 42, pp. 9458–9471, Nov 1990. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.42.9458

[198] D. G. Pettifor and I. I. Oleinik, "Analytic bond-order potentials beyond tersoff-brenner. i. theory," *Phys. Rev. B*, vol. 59, pp. 8487–8499, Apr 1999. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.59.8487

[199] Y. Akagi, M. Udagawa, and Y. Motome, "Hidden multiple-spin interactions as an origin of spin scalar chiral order in frustrated kondo lattice models," *Phys. Rev. Lett.*, vol. 108, p. 096401, Feb 2012. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.108.096401

[200] S. Hayami, R. Ozawa, and Y. Motome, "Effective bilinear-biquadratic model for noncoplanar ordering in itinerant magnets," *Phys. Rev. B*, vol. 95, p. 224424, Jun 2017. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.95.224424

[201] N. Furukawa, "Transport properties of the kondo lattice model in the limit s=∞ and d=∞," *Journal of the Physical Society of Japan*, vol. 63, no. 9, pp.

3214–3217, 1994, https://doi.org/10.1143/JPSJ.63.3214. [Online]. Available: https://doi.org/10.1143/JPSJ.63.3214

[202] E. Dagotto, S. Yunoki, A. L. Malvezzi, A. Moreo, J. Hu, S. Capponi, D. Poilblanc, and N. Furukawa, "Ferromagnetic kondo model for manganites: Phase diagram, charge segregation, and influence of quantum localized spins," *Phys. Rev. B*, vol. 58, pp. 6414–6427, Sep 1998. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.58.6414

[203] D. Pekker, S. Mukhopadhyay, N. Trivedi, and P. M. Goldbart, "Double-exchange model for noninteracting electron spins coupled to a lattice of classical spins: Phase diagram at zero temperature," *Phys. Rev. B*, vol. 72, p. 075118, Aug 2005. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.72.075118

[204] S. Yunoki, J. Hu, A. L. Malvezzi, A. Moreo, N. Furukawa, and E. Dagotto, "Phase separation in electronic models for manganites," *Phys. Rev. Lett.*, vol. 80, pp. 845–848, Jan 1998. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.80.845

[205] W. Y. Ching and D. L. Huber, "Numerical studies of energy levels and eigenfunction localization in dilute three-dimensional systems with exponential interactions," *Phys. Rev. B*, vol. 25, pp. 1096–1100, Jan 1982. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.25.1096

[206] D. E. Logan and M. D. Winn, "The density of states of a spatially disordered tight-binding model," *Journal of Physics C: Solid State Physics*, vol. 21, no. 34, p. 5773, dec 1988. [Online]. Available: https://dx.doi.org/10.1088/0022-3719/21/34/013

[207] M. D. Winn and D. E. Logan, "A soluble theory for the density of states of a spatially disordered system," *Journal of Physics: Condensed Matter*, vol. 1, no. 9, p. 1753, mar 1989. [Online]. Available: https://dx.doi.org/10.1088/0953-8984/1/9/018

[208] I. J. Bush, D. E. Logan, P. A. Madden, and M. D. Winn, "The density of states of a spatially disordered system: theory compared with simulation," *Journal of Physics: Condensed Matter*, vol. 1, no. 14, p. 2551, apr 1989. [Online]. Available: https://dx.doi.org/10.1088/0953-8984/1/14/011

[209] D. J. Priour, "Electronic states in one-, two-, and three-dimensional highly amorphous materials: A tight-binding treatment," *Phys. Rev. B*, vol. 85, p. 014209, Jan 2012. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.85.014209

[210] J. T. Barron, "Continuously differentiable exponential linear units," 2017," 1704.07483. [Online]. Available: https://arxiv.org/abs/1704.07483

[211] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch,"

2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:40027675

[212] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017," 1412.6980. [Online]. Available: https://arxiv.org/abs/1412.6980

[213] G.-W. Chern, K. Barros, Z. Wang, H. Suwa, and C. D. Batista, "Semiclassical dynamics of spin density waves," *Phys. Rev. B*, vol. 97, p. 035120, Jan 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.97.035120

[214] D. S. Fisher and D. A. Huse, "Nonequilibrium dynamics of spin glasses," *Phys. Rev. B*, vol. 38, pp. 373–385, Jul 1988. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.38.373

[215] D. A. Huse, "Monte carlo simulation study of domain growth in an ising spin glass," *Phys. Rev. B*, vol. 43, pp. 8673–8675, Apr 1991. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.43.8673

[216] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation," *APL Materials*, vol. 1, no. 1, p. 011002, 07 2013. [Online]. Available: https://doi.org/10.1063/1.4812323

[217] A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk, "Scaling deep learning for materials discovery,"

Nature, vol. 624, no. 7990, pp. 80–85, Dec. 2023. [Online]. Available: https://doi.org/10.1038/s41586-023-06735-9

[218] J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, and K. Burke, "Finding density functionals with machine learning," *Phys. Rev. Lett.*, vol. 108, p. 253002, Jun 2012. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.108.253002

[219] F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller, "Bypassing the kohn-sham equations with machine learning," *Nature Communications*, vol. 8, no. 1, p. 872, Oct 2017. [Online]. Available: https://doi.org/10.1038/s41467-017-00839-3

[220] R. Nagai, R. Akashi, and O. Sugino, "Completing density functional theory by machine learning hidden messages from molecules," *npj Computational Materials*, vol. 6, no. 1, p. 43, May 2020. [Online]. Available: https://doi.org/10.1038/s41524-020-0310-0

[221] L. Li, S. Hoyer, R. Pederson, R. Sun, E. D. Cubuk, P. Riley, and K. Burke, "Kohn-sham equations as regularizer: Building prior knowledge into machine-learned physics," *Phys. Rev. Lett.*, vol. 126, p. 036401, Jan 2021. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.126.036401

[222] J. Kirkpatrick, B. McMorrow, D. H. P. Turban, A. L. Gaunt, J. S. Spencer, A. G. D. G. Matthews, A. Obika, L. Thiry, M. Fortunato, D. Pfau, L. R. Castellanos,

S. Petersen, A. W. R. Nelson, P. Kohli, P. Mori-Sánchez, D. Hassabis, and A. J. Cohen, "Pushing the frontiers of density functionals by solving the fractional electron problem," *Science*, vol. 374, no. 6573, pp. 1385–1389, 2021. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.abj6511

[223] R. Pederson, B. Kalita, and K. Burke, "Machine learning and density functional theory," *Nature Reviews Physics*, vol. 4, no. 6, pp. 357–358, Jun 2022. [Online]. Available: https://doi.org/10.1038/s42254-022-00470-2

[224] A. Seko, A. Togo, H. Hayashi, K. Tsuda, L. Chaput, and I. Tanaka, "Prediction of low-thermal-conductivity compounds with first-principles anharmonic lattice-dynamics calculations and bayesian optimization," *Phys. Rev. Lett.*, vol. 115, p. 205901, Nov 2015. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.115.205901

[225] F. A. Faber, A. Lindmaa, O. A. von Lilienfeld, and R. Armiento, "Machine learning energies of 2 million elpasolite (ABC_2D_6) crystals," *Phys. Rev. Lett.*, vol. 117, p. 135502, Sep 2016. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.117.135502

[226] D. Xue, P. V. Balachandran, J. Hogden, J. Theiler, D. Xue, and T. Lookman, "Accelerated search for materials with targeted properties by adaptive design," *Nature Communications*, vol. 7, no. 1, p. 11241, 2016. [Online]. Available: https://doi.org/10.1038/ncomms11241

[227] S. Curtarolo, G. L. W. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy, "The high-throughput highway to computational materials design," *Nature Materials*, vol. 12, no. 3, pp. 191–201, 2013. [Online]. Available: https://doi.org/10.1038/nmat3568

[228] F. Giustino, "Electron-phonon interactions from first principles," *Rev. Mod. Phys.*, vol. 89, p. 015003, Feb 2017. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.89.015003

[229] H. Shang and J. Yang, "The electron-phonon renormalization in the electronic structure calculation: Fundamentals, current status, and challenges," *The Journal of Chemical Physics*, vol. 158, no. 13, p. 130901, 04 2023. [Online]. Available: https://doi.org/10.1063/5.0140724

[230] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi, "Phonons and related crystal properties from density-functional perturbation theory," *Rev. Mod. Phys.*, vol. 73, pp. 515–562, Jul 2001. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.73.515

[231] F. Giustino, M. L. Cohen, and S. G. Louie, "Electron-phonon interaction using wannier functions," *Phys. Rev. B*, vol. 76, p. 165108, Oct 2007. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.76.165108

[232] M. Bernardi, "First-principles dynamics of electrons and phonons," *The European Physical Journal B*, vol. 89, no. 11, p. 239, Nov 2016. [Online].

Available: https://doi.org/10.1140/epjb/e2016-70399-4

[233] H. Y. Fan, "Temperature dependence of the energy gap in semiconductors," *Phys. Rev.*, vol. 82, pp. 900–905, Jun 1951. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.82.900

[234] P. B. Allen and V. Heine, "Theory of the temperature dependence of electronic band structures," *Journal of Physics C: Solid State Physics*, vol. 9, no. 12, p. 2305, jun 1976. [Online]. Available: https://dx.doi.org/10.1088/0022-3719/9/12/013

[235] P. B. Allen and M. Cardona, "Theory of the temperature dependence of the direct gap of germanium," *Phys. Rev. B*, vol. 23, pp. 1495–1505, Feb 1981. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.23.1495

[236] M. Cardona and M. L. W. Thewalt, "Isotope effects on the optical spectra of semiconductors," *Rev. Mod. Phys.*, vol. 77, pp. 1173–1224, Nov 2005. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.77.1173

[237] F. Giustino, S. G. Louie, and M. L. Cohen, "Electron-phonon renormalization of the direct band gap of diamond," *Phys. Rev. Lett.*, vol. 105, p. 265501, Dec 2010. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.105.265501

[238] H. Kawai, K. Yamashita, E. Cannuccia, and A. Marini, "Electron-electron and electron-phonon correlation effects on the finite-temperature electronic and

optical properties of zinc-blende gan," *Phys. Rev. B*, vol. 89, p. 085202, Feb 2014. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.89.085202

[239] S. Poncé, G. Antonius, Y. Gillet, P. Boulanger, J. Laflamme Janssen, A. Marini, M. Côté, and X. Gonze, "Temperature dependence of electronic eigenenergies in the adiabatic harmonic approximation," *Phys. Rev. B*, vol. 90, p. 214304, Dec 2014. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.90.214304

[240] L. Hedin, "New method for calculating the one-particle green's function with application to the electron-gas problem," *Phys. Rev.*, vol. 139, pp. A796–A823, Aug 1965. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.139.A796

[241] F. Aryasetiawan and O. Gunnarsson, "The gw method," *Reports on Progress in Physics*, vol. 61, no. 3, p. 237, mar 1998. [Online]. Available: https://dx.doi.org/10.1088/0034-4885/61/3/002

[242] G. Onida, L. Reining, and A. Rubio, "Electronic excitations: density-functional versus many-body green's-function approaches," *Rev. Mod. Phys.*, vol. 74, pp. 601–659, Jun 2002. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.74.601

[243] M. T. Yin and M. L. Cohen, "Microscopic theory of the phase transformation and lattice dynamics of si," *Phys. Rev. Lett.*, vol. 45, pp. 1004–1007, Sep 1980.

[Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.45.1004

[244] P. K. Lam and M. L. Cohen, "Ab initio calculation of phonon frequencies of al," *Phys. Rev. B*, vol. 25, pp. 6139–6145, May 1982. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.25.6139

[245] M. M. Dacorogna, M. L. Cohen, and P. K. Lam, "Self-consistent calculation of the q dependence of the electron-phonon coupling in aluminum," *Phys. Rev. Lett.*, vol. 55, pp. 837–840, Aug 1985. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.55.837

[246] A. Fleszar and R. Resta, "Dielectric matrices in semiconductors: A direct approach," *Phys. Rev. B*, vol. 31, pp. 5305–5310, Apr 1985. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.31.5305

[247] P. K. Lam, M. M. Dacorogna, and M. L. Cohen, "Self-consistent calculation of electron-phonon couplings," *Phys. Rev. B*, vol. 34, pp. 5065–5069, Oct 1986. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.34.5065

[248] R. B. Capaz, C. D. Spataru, P. Tangney, M. L. Cohen, and S. G. Louie, "Temperature dependence of the band gap of semiconducting carbon nanotubes," *Phys. Rev. Lett.*, vol. 94, p. 036801, Jan 2005. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.94.036801

[249] B. Monserrat, "Electron–phonon coupling from finite differences," *Journal of Physics: Condensed Matter*, vol. 30, no. 8, p. 083001, feb 2018. [Online]. Available: https://dx.doi.org/10.1088/1361-648X/aaa737

[250] J. H. Lloyd-Williams and B. Monserrat, "Lattice dynamics and electron-phonon coupling calculations using nondiagonal supercells," *Phys. Rev. B*, vol. 92, p. 184301, Nov 2015. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.92.184301

[251] C. E. Patrick and F. Giustino, "Quantum nuclear dynamics in the photophysics of diamondoids," *Nature Communications*, vol. 4, no. 1, p. 2006, Jun. 2013. [Online]. Available: https://doi.org/10.1038/ncomms3006

[252] M. Zacharias, C. E. Patrick, and F. Giustino, "Stochastic approach to phonon-assisted optical absorption," *Phys. Rev. Lett.*, vol. 115, p. 177401, Oct 2015. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.115.177401

[253] B. Monserrat, N. D. Drummond, C. J. Pickard, and R. J. Needs, "Electron-phonon coupling and the metallization of solid helium at terapascal pressures," *Phys. Rev. Lett.*, vol. 112, p. 055504, Feb 2014. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.112.055504

[254] B. Monserrat, E. A. Engel, and R. J. Needs, "Giant electron-phonon interactions in molecular crystals and the importance of nonquadratic

coupling," *Phys. Rev. B*, vol. 92, p. 140302, Oct 2015. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.92.140302

[255] B. J. Alder and T. E. Wainwright, "Studies in Molecular Dynamics. I. General Method," *The Journal of Chemical Physics*, vol. 31, no. 2, pp. 459–466, 08 1959. [Online]. Available: https://doi.org/10.1063/1.1730376

[256] R. Car and M. Parrinello, "Unified approach for molecular dynamics and density-functional theory," *Phys. Rev. Lett.*, vol. 55, pp. 2471–2474, Nov 1985. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.55.2471

[257] D. Pan, Q. Wan, and G. Galli, "The refractive index and electronic gap of water and ice increase with increasing pressure," *Nature Communications*, vol. 5, no. 1, p. 3919, May 2014. [Online]. Available: https://doi.org/10.1038/ncomms4919

[258] J. Cao and G. A. Voth, "The formulation of quantum statistical mechanics based on the Feynman path centroid density. I. Equilibrium properties," *The Journal of Chemical Physics*, vol. 100, no. 7, pp. 5093–5105, 04 1994. [Online]. Available: https://doi.org/10.1063/1.467175

[259] J. Cao and G. A. Voth, "The formulation of quantum statistical mechanics based on the Feynman path centroid density. II. Dynamical properties," *The Journal of Chemical Physics*, vol. 100, no. 7, pp. 5106–5117, 04 1994. [Online]. Available: https://doi.org/10.1063/1.467176

[260] R. Ramírez, C. P. Herrero, and E. R. Hernández, "Path-integral molecular dynamics simulation of diamond," *Phys. Rev. B*, vol. 73, p. 245202, Jun 2006. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.73.245202

[261] M. A. Morales, J. M. McMahon, C. Pierleoni, and D. M. Ceperley, "Towards a predictive first-principles description of solid molecular hydrogen with density functional theory," *Phys. Rev. B*, vol. 87, p. 184107, May 2013. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.87.184107

[262] B. Monserrat, "Vibrational averages along thermal lines," *Phys. Rev. B*, vol. 93, p. 014302, Jan 2016. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.93.014302

[263] B. Monserrat, "Correlation effects on electron-phonon coupling in semiconductors: Many-body theory along thermal lines," *Phys. Rev. B*, vol. 93, p. 100301, Mar 2016. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.93.100301

[264] M. Zacharias and F. Giustino, "One-shot calculation of temperature-dependent optical spectra and phonon-induced band-gap renormalization," *Phys. Rev. B*, vol. 94, p. 075125, Aug 2016. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.94.075125

[265] J. Jung, J. I. Yoon, H. K. Park, J. Y. Kim, and H. S. Kim, "An efficient machine learning approach to establish structure-property linkages,"

Computational Materials Science, vol. 156, pp. 17–25, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0927025618306335

[266] Z. Fan, J. Ding, and E. Ma, "Machine learning bridges local static structure with multiple properties in metallic glasses," *Materials Today*, vol. 40, pp. 48–62, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1369702120302108

[267] P. Hundi and R. Shahsavari, "Deep learning to speed up the development of structure–property relations for hexagonal boron nitride and graphene," *Small*, vol. 15, no. 19, p. 1900656, 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/smll.201900656

[268] H. Zhang, Z. Guo, H. Hu, G. Zhou, Q. Liu, Y. Xu, Q. Qian, and D. Dai, "A novel structure-property relationship model based on machine learning," *Modelling and Simulation in Materials Science and Engineering*, vol. 28, no. 3, p. 035002, feb 2020. [Online]. Available: https://dx.doi.org/10.1088/1361-651X/ab6bb7

[269] R. Abdusalamov, P. Pandit, B. Milow, M. Itskov, and A. Rege, "Machine learning-based structure–property predictions in silica aerogels," *Soft Matter*, vol. 17, pp. 7350–7358, 2021. [Online]. Available: http://dx.doi.org/10.1039/D1SM00307K

[270] Z. Li, G. Antonius, M. Wu, F. H. da Jornada, and S. G. Louie, "Electron-phonon coupling from ab initio linear-response theory within the

gw method: Correlation-enhanced interactions and superconductivity in $ba_{1-x}k_xbio_3$," *Phys. Rev. Lett.*, vol. 122, p. 186402, May 2019. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.122.186402

[271] C. P. Koçer, K. Haule, G. L. Pascut, and B. Monserrat, "Efficient lattice dynamics calculations for correlated materials with DFT + DMFT," *Phys. Rev. B*, vol. 102, p. 245104, Dec 2020. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.102.245104

[272] F. E. Williams, "Theoretical low temperature spectra of the thallium activated potassium chloride phosphor," *Phys. Rev.*, vol. 82, pp. 281–282, Apr 1951. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.82.281.2

[273] M. Lax, "The Franck-Condon Principle and Its Application to Crystals," *The Journal of Chemical Physics*, vol. 20, no. 11, pp. 1752–1760, 11 1952. [Online]. Available: https://doi.org/10.1063/1.1700283

[274] S. Baroni, P. Giannozzi, and A. Testa, "Green's-function approach to linear response in solids," *Phys. Rev. Lett.*, vol. 58, pp. 1861–1864, May 1987. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.58.1861

[275] P. Giannozzi, S. de Gironcoli, P. Pavone, and S. Baroni, "Ab initio calculation of phonon dispersions in semiconductors," *Phys. Rev. B*, vol. 43, pp. 7231–7242, Mar 1991. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.43.7231

[276] X. Gonze, "First-principles responses of solids to atomic displacements and homogeneous electric fields: Implementation of a conjugate-gradient algorithm," *Phys. Rev. B*, vol. 55, pp. 10 337–10 354, Apr 1997. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.55.10337

[277] F. Karsai, M. Engel, E. Flage-Larsen, and G. Kresse, "Electron–phonon coupling in semiconductors within the gw approximation," *New Journal of Physics*, vol. 20, no. 12, p. 123008, dec 2018. [Online]. Available: https://dx.doi.org/10.1088/1367-2630/aaf53f

[278] G. Antonius, S. Poncé, E. Lantagne-Hurtubise, G. Auclair, X. Gonze, and M. Côté, "Dynamical and anharmonic effects on the electron-phonon coupling and the zero-point renormalization of the electronic structure," *Phys. Rev. B*, vol. 92, p. 085137, Aug 2015. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.92.085137

[279] A. Kundu and G. Galli, "Quantum vibronic effects on the excitation energies of the nitrogen-vacancy center in diamond," *The Journal of Physical Chemistry Letters*, vol. 15, no. 3, pp. 802–810, Jan 2024. [Online]. Available: https://doi.org/10.1021/acs.jpclett.3c03269

[280] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals and Systems*, vol. 2, no. 4, pp. 303–314, Dec 1989. [Online]. Available: https://doi.org/10.1007/BF02551274

[281] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0893608089900208

[282] T. Xie and J. C. Grossman, "Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties," *Phys. Rev. Lett.*, vol. 120, p. 145301, Apr 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.120.145301

[283] Z. Tian, S. Zhang, and G.-W. Chern, "Machine learning for structure-property mapping of ising models: Scalability and limitations," *Phys. Rev. E*, vol. 108, p. 065304, Dec 2023. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.108.065304

[284] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, and et al, "Quantum espresso: a modular and open-source software project for quantum simulations of materials," *Journal of Physics: Condensed Matter*, vol. 21, no. 39, p. 395502 (19pp), 2009. [Online]. Available: http://www.quantum-espresso.org

[285] J. P. Perdew and A. Zunger, "Self-interaction correction to density-functional approximations for many-electron systems," *Phys. Rev. B*, vol. 23, pp.

5048–5079, May 1981. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.23.5048

[286] S. Wei and M. Y. Chou, "Phonon dispersions of silicon and germanium from first-principles calculations," *Phys. Rev. B*, vol. 50, pp. 2221–2226, Jul 1994. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.50.2221

[287] A. F. Agarap, "Deep learning using rectified linear units (relu)," 2019," 1803.08375. [Online]. Available: https://arxiv.org/abs/1803.08375

[288] A. L. Maas, "Rectifier nonlinearities improve neural network acoustic models," 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:16489696

[289] D. Hendrycks and K. Gimpel, "Gaussian error linear units (gelus)," 2023," 1606.08415. [Online]. Available: https://arxiv.org/abs/1606.08415

[290] P. Souvatzis, O. Eriksson, M. Katsnelson, and S. Rudin, "The self-consistent ab initio lattice dynamical method," *Computational Materials Science*, vol. 44, no. 3, pp. 888–894, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0927025608003121

[291] M. Zacharias, M. Scheffler, and C. Carbogno, "Fully anharmonic nonperturbative theory of vibronically renormalized electronic band structures," *Phys. Rev. B*, vol. 102, p. 045126, Jul 2020. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.102.045126

[292] N. F. Mott, *Metal-Insulator Transitions.* New York: Taylor & Francis, 1997.

[293] M. Imada, A. Fujimori, and Y. Tokura, "Metal-insulator transitions," *Rev. Mod. Phys.*, vol. 70, p. 1039, 1998.

[294] V. Dobrosavljevic, *Introduction to Metal-Insulator Transitions.* Oxford: Oxford University Press, 2012.

[295] M. Nakano, K. Shibuya, D. Okuyama, T. Hatano, S. Ono, M. Kawasaki, Y. Iwasa, and Y. Tokura, "Collective bulk carrier delocalization driven by electrostatic surface charge accumulation," *Nature*, vol. 487, p. 459, 2012.

[296] N. Shukla, A. V. Thathachary, A. Agrawal, H. Paik, A. Aziz, D. G. Schlom, S. K. Gupta, R. Engel-Herbert, and S. Datta, "A steep-slope transistor based on abrupt electronic phase transition," *Nat. Commun.*, vol. 6, p. 7812, 2015.

[297] J. Song, J. Woo, S. Lee, and A. Prakash, *IEEE Electron Device Lett.*, vol. 37, p. 932, 2016.

[298] G. Gr"uner, "The dynamics of charge-density waves," *Rev. Mod. Phys.*, vol. 60, p. 1129, 1988.

[299] R. E. Thorne, "Charge-density-wave conductors," *Phys. Today*, vol. 49, p. 42, 1996.

[300] A. A. Balandin, S. V. Zaitsev-Zotov, and G. Gr"uner, "Charge-density-wave quantum materials and devices – new developments and future prospects," *Appl. Phys. Lett.*, vol. 119, p. 170401, 2021.

[301] S. Hellmann, T. Rohwer, M. Kall"ane, K. Hanff, C. Sohrt, A. Stange, A. Carr, M. M. Murnane, H. C. Kapteyn, L. Kipp, M. Bauer, and K. Rossnagel, "Time-domain classification of charge-density-wave insulators," *Nat. Commun.*, vol. 3, p. 1069, 2012.

[302] M. Porer, U. Leierseder, J.-M. M'enard, H. Dachraoui, L. Mouchliadis, I. E. Perakis, U. Heinzmann, J. Demsar, K. Rossnagel, and R. Huber, "Non-thermal separation of electronic and structural orders in a persisting charge density wave," *Nat. Mater.*, vol. 13, p. 857, 2014.

[303] L. Stojchevska, I. Vaskivskyi, T. Mertelj, P. Kusar, D. Svetin, S. Brazovskii, and D. Mihailovic, "Ultrafast switching to a stable hidden quantum state in an electronic crystal," *Science*, vol. 344, p. 177, 2014.

[304] Y. Yu, F. Yang, X. Lu, Y. Yan, Y.-H. Cho, L. Ma, X. Niu, S. Kim, Y.-W. Son, D. Feng, S. Li, S.-W. Cheong, X. Chen, and Y. Zhang, "Gate-tunable phase transitions in thin flakes of 1t-tas_2," *Nat. Nanotech.*, vol. 10, p. 270, 2015.

[305] R. Samnakay, D. Wickramaratne, T. R. Pope, R. K. Lake, T. T. Salguero, and A. A. Balandin, *Nano Lett.*, vol. 15, p. 2965, 2015.

[306] X. Xi, L. Zhao, Z. Wang, H. Berger, L. Forr'o, J. Shan, and K. F. Mak, "Strongly enhanced charge-density-wave order in monolayer nbse$_2$," *Nat. Nanotechnol.*, vol. 10, p. 765, 2015.

[307] I. Vaskivskyi, J. Gospodaric, S. Brazovskii, D. Svetin, P. Sutar, E. Goreshnik, I. A. Mihailovic, T. Mertelj, and D. Mihailovic, "Controlling the metal-to-insulator relaxation of the metastable hidden quantum state in 1t-tas$_2$," *Sci. Adv.*, vol. 1, p. e1500168, 2015.

[308] S. Vogelgesang, G. Storeck, J. G. Horstmann, T. Diekmann, M. Sivis, S. Schramm, K. Rossnagel, S. Sch"afer, and C. Ropers, "Phase ordering of charge density waves traced by ultrafast low-energy electron diffraction," *Nat. Phys.*, vol. 14, p. 184, 2018.

[309] D. Cho, S. Cheon, K.-S. Kim, S.-H. Lee, Y.-H. Cho, S.-W. Cheong, and H. W. Yeom, "Nanoscale manipulation of the mott insulating state coupled to charge order in 1t-tas$_2$," *Nat. Commun.*, vol. 7, p. 10453, 2016.

[310] M. J. Hollander, Y. Liu, W.-J. Lu, L.-J. Li, Y.-P. Sun, J. A. Robinson, and S. Datta, "Electrically driven reversible insulator-metal phase transition in 1t-tas$_2$," *Nano Lett.*, vol. 15, p. 1861, 2015.

[311] I. Vaskivskyi, I. A. Mihailovic, S. Brazovskii, J. Gospodaric, T. Mertelj, D. Svetin, P. Sutar, and D. Mihailovic, "Fast electronic resistance switching

involving hidden charge density wave states," *Nat. Commun.*, vol. 7, p. 11442, 2016.

[312] B. Grisafe, R. Zhao, R. K. Ghosh, J. A. Robinson, and S. Datta, "Electrically triggered insulator-to-metal phase transition in two-dimensional (2d) heterostructures," *Appl. Phys. Lett.*, vol. 113, p. 142101, 2018.

[313] Y. Li, K.-A. N. Duerloo, K. Wauson, and E. J. Reed, "Structural semiconductor-to-semimetal phase transition in two-dimensional materials induced by electrostatic gating," *Nat. Commun.*, vol. 7, p. 10671, 2016.

[314] A. K. Geremew, S. Rumyantsev, F. Kargar, B. Debnath, A. Nosek, M. A. Bloodgood, M. Bockrath, T. Salguero, R. K. Lake, and A. A. Balandin, "Bias-voltage driven switching of the charge-density-wave and normal metallic phases in 1t-tas_2 thin-film devices," *ACS Nano*, vol. 13, p. 7231, 2019.

[315] S. Zheng, F. Liu, C. Zhu, Z. Liu, and H. J. Fan, "Room-temperature electrically driven phase transition of two-dimensional 1t-tas_2 layers," *Nanoscale*, vol. 9, p. 2436, 2017.

[316] G. Liu, B. Debnath, T. R. Pope, T. T. Salguero, R. K. Lake, and A. A. Balandin, "A charge-density-wave oscillator based on an integrated tantalum disulfide-boron nitride-graphene device operating at room temperature," *Nat. Nanotechnol.*, vol. 11, p. 845, 2016.

[317] A. G. Khitun, A. K. Geremew, and A. A. Balandin, "Transistor-less logic circuits implemented with 2-d charge density wave devices," *IEEE Electron Device Lett.*, vol. 39, p. 1449, 2018.

[318] R. Waser and M. Aono, "Nanoionics-based resistive switching memories," *Nat. Mater.*, vol. 6, p. 833, 2007.

[319] A. Sawa, "Resistive switching in transition metal oxides," *Mater. Today*, vol. 11, p. 28, 2008.

[320] R. Waser, R. Dittmann, G. Staikov, and K. Szot, "Redox-based resistive switching memories - nanoionic mechanisms, prospects, and challenges," *Adv. Mater.*, vol. 21, p. 2632, 2009.

[321] K. M. Kim, D. S. Jeong, and C. S. Hwang, "Nanofilamentary resistive switching in binary oxide system; a review on the present status and outlook," *Nanotechnology*, vol. 22, p. 254002, 2011.

[322] D. S. Jeong, R. Thomas, R. S. Katiyar, J. F. Scott, H. Kohlstedt, A. Petraru, and C. S. Hwang, "Emerging memories: resistive switching mechanisms and current status," *Rep. Prog. Phys.*, vol. 75, p. 076502, 2012.

[323] J. S. Lee, S. Lee, and T. W. Noh, "Resistive switching phenomena: A review of statistical physics approaches," *Appl. Phys. Rev.*, vol. 2, p. 031303, 2015.

[324] E. Janod, J. Tranchant, B. Corraze, M. Querre, P. Stoliar, M. Rozenberg, T. Cren, D. Roditchev, V. T. Phuoc, M.-P. Besland, and L. Cario, "Resistive switching in mott insulators and correlated systems," *Adv. Funct. Mater.*, vol. 25, p. 6287, 2015.

[325] T. Holstein, "Studies of polaron motion: Part i. the molecular-crystal model," *Ann. Phys.*, vol. 8, p. 325, 1959.

[326] A. Goetz, S. Beyl, M. Hohenadler, and F. F. Assaad, "Langevin dynamics simulations of the two-dimensional su-schrieffer-heeger model," *arXiv*, 2021, 2102.08899.

[327] K. Michielsen and H. D. Raedt, "Quantum molecular dynamics study of the su-schrieffer-heeger model," *Z. Phys. B*, vol. 103, p. 391, 1997.

[328] M. D. Ventra and S. T. Pantelides, "Hellmann-feynman theorem and the definition of forces in quantum time-dependent and transport problems," *Phys. Rev. B*, vol. 61, p. 16207, 2000.

[329] T. N. Todorov, "Time-dependent tight-binding," *J. Phys.: Condens. Matter*, vol. 13, p. 10125, 2001.

[330] J.-T. L"u, M. Brandbyge, P. Hedegard, T. N. Todorov, and D. Dundas, "Current-induced atomic dynamics, instabilities, and raman signals: Quasiclassical langevin equation approach," *Phys. Rev. B*, vol. 85, p. 245444, 2012.

[331] T. N. Todorov, D. Dundas, and E. J. McEniry, "Nonconservative generalized current-induced forces," *Phys. Rev. B*, vol. 81, p. 075416, 2010.

[332] D. Dundas, E. J. McEniry, and T. N. Todorov, "Current-driven atomic waterwheels," *Nat. Nanotech.*, vol. 4, p. 99, 2009.

[333] Y. Meir and N. S. Wingreen, "Landauer formula for the current through an interacting electron region," *Phys. Rev. Lett.*, vol. 68, p. 2512, 1992.

[334] A.-P. Jauho, N. S. Wingreen, and Y. Meir, "Time-dependent transport in interacting and noninteracting resonant-tunneling systems," *Phys. Rev. B*, vol. 50, p. 5528, 1994.

[335] H. Haug and A.-P. Jauho, *Quantum Kinetics in Transport and Optics of Semiconductors*, ser. Springer Series in Solid-State Sciences. Berlin: Springer-Verlag, 2008, vol. 123.

[336] M. D. Ventra, *Electrical Transport in Nanoscale Systems*. Cambridge: Cambridge University Press, 2008.

[337] J. Zhong and H.-B. Sch"uttler, "Polaronic anharmonicity in the holstein-hubbard model," *Phys. Rev. Lett.*, vol. 69, p. 1600, 1992.

[338] M. Weber and M. Hohenadler, "Two-dimensional holstein-hubbard model: Critical temperature, ising universality, and bipolaron liquid," *Phys. Rev. B*, vol. 98, p. 085405, 2018.

[339] P. Ribeiro, A. E. Antipov, and A. N. Rubtsov, "Nonequilibrium breakdown of a correlated insulator through pattern formation," *Phys. Rev. B*, vol. 93, p. 144305, 2016.

[340] A. Dutta and P. Majumdar, "Spatial behavior in a mott insulator near the voltage-driven resistive transition," *Phys. Rev. B*, vol. 101, p. 245155, 2020.

[341] M. Rupp, A. Tkatchenko, K.-R. M"uller, and O. A. von Lilienfeld, "Fast and accurate modeling of molecular atomization energies with machine learning," *Phys. Rev. Lett.*, vol. 108, p. 058301, 2012.

[342] K. T. Sch"utt, M. Gastegger, A. Tkatchenko, K.-R. M"uller, and R. J. Maurer, "Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions," *Nat. Commun.*, vol. 10, p. 5024, 2019.

[343] S. Wang, K. Fan, N. Luo, Y. Cao, F. Wu, C. Zhang, K. A. Heller, and L. You, "Massive computational acceleration by using neural networks to emulate mechanism-based biological models," *Nat. Commun.*, vol. 10, p. 4354, 2019.

[344] M. Tsubaki and T. Mizoguchi, "Quantum deep field: Data-driven wave function, electron density generation, and atomization energy prediction and extrapolation with machine learning," *Phys. Rev. Lett.*, vol. 125, p. 206401, 2020.

[345] M. B"urkle, U. Perera, F. Gimbert, H. Nakamura, M. Kawata, and Y. Asai, "Deep-learning approach to first-principles transport simulations," *Phys. Rev. Lett.*, vol. 126, p. 177701, 2021.

[346] L. Huang and L. Wang, "Accelerated monte carlo simulations with restricted boltzmann machines," *Phys. Rev. B*, vol. 95, p. 035105, 2017.

[347] J. Liu, Y. Qi, Z. Y. Meng, and L. Fu, "Self-learning monte carlo method," *Phys. Rev. B*, vol. 95, p. 041101(R), 2017.

[348] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, p. 85, 2015.

[349] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, p. 436, 2015.

[350] A. Barron, "Universal approximation bounds for superpositions of a sigmoidal function," *IEEE Trans. Inf. Theory*, vol. 39, p. 930, 1993.

[351] R. Iftimie, P. Minary, and M. E. Tuckerman, "Ab initio molecular dynamics: concepts, recent developments, and future trends," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 19, pp. 6654–6659, May 2005.

[352] C. Zener, "Interaction between the *d*-shells in the transition metals. ii. ferromagnetic compounds of manganese with perovskite structure,"

Phys. Rev., vol. 82, pp. 403–405, May 1951. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.82.403

[353] P. W. Anderson and H. Hasegawa, "Considerations on double exchange," *Phys. Rev.*, vol. 100, pp. 675–681, Oct 1955. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.100.675

[354] P. G. de Gennes, "Effects of double exchange in magnetic crystals," *Phys. Rev.*, vol. 118, pp. 141–154, Apr 1960. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.118.141

[355] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, "Bond-orientational order in liquids and glasses," *Phys. Rev. B*, vol. 28, pp. 784–805, Jul 1983. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.28.784

[356] F. Faber, A. Lindmaa, O. A. von Lilienfeld, and R. Armiento, "Crystal structure representations for machine learning models of formation energies," 2015," 1503.07406. [Online]. Available: https://arxiv.org/abs/1503.07406

[357] S. Mazumdar and S. N. Dixit, "Coulomb effects on one-dimensional peierls instability: The peierls-hubbard model," *Phys. Rev. Lett.*, vol. 51, pp. 292–295, Jul 1983. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.51.292

[358] J. E. Hirsch, "Effect of coulomb interactions on the peierls instability," *Phys. Rev. Lett.*, vol. 51, pp. 296–299, Jul 1983. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.51.296

[359] W. P. Su, J. R. Schrieffer, and A. J. Heeger, "Soliton excitations in polyacetylene," *Phys. Rev. B*, vol. 22, pp. 2099–2111, Aug 1980. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.22.2099

[360] Z. Popovic and S. Satpathy, "Cooperative jahn-teller coupling in the manganites," *Phys. Rev. Lett.*, vol. 84, pp. 1603–1606, Feb 2000. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.84.1603

[361] C. Şen, G. Alvarez, H. Aliaga, and E. Dagotto, "Colossal magnetoresistance observed in monte carlo simulations of the one- and two-orbital models for manganites," *Phys. Rev. B*, vol. 73, p. 224441, Jun 2006. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.73.224441

[362] G.-W. Chern, K. Barros, C. D. Batista, J. D. Kress, and G. Kotliar, "Mott transition in a metallic liquid: Gutzwiller molecular dynamics simulations," *Phys. Rev. Lett.*, vol. 118, p. 226401, Jun 2017. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.118.226401

[363] R. Kakarala, "A group-theoretic approach to the triple correlation," in *[1993 Proceedings] IEEE Signal Processing Workshop on Higher-Order Statistics*, 1993, pp. 28–32.

[364] R. Kakarala, "Completeness of bispectrum on compact groups," 2009. [Online]. Available: https://api.semanticscholar.org/CorpusID:18425284

[365] T. Inui, Y. Tanabe, and Y. Onodera, *Group Theory and Its Applications in Physics*, ser. Springer Series in Solid-State Sciences. Springer, 1990.

[366] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii, *Quantum Theory of Angular Momentum.* Singapore: World Scientific, 1988.

[367] T. Hotta, S. Yunoki, M. Mayr, and E. Dagotto, "A-type antiferromagnetic and c-type orbital-ordered states in $lamno_3$ using cooperative jahn-teller phonons," *Phys. Rev. B*, vol. 60, pp. R15 009–R15 012, Dec 1999. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.60.R15009

[368] M. N. Iliev, M. V. Abrashev, H.-G. Lee, V. N. Popov, Y. Y. Sun, C. Thomsen, R. L. Meng, and C. W. Chu, "Raman spectroscopy of orthorhombic perovskitelike $ymno_3$ and $lamno_3$," *Phys. Rev. B*, vol. 57, pp. 2872–2877, Feb 1998. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.57.2872

[369] S. Yunoki, A. Moreo, and E. Dagotto, "Phase separation induced by orbital degrees of freedom in models for manganites with jahn-teller phonons," *Phys. Rev. Lett.*, vol. 81, pp. 5612–5615, Dec 1998. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.81.5612

[370] T. Hotta, M. Moraghebi, A. Feiguin, A. Moreo, S. Yunoki, and E. Dagotto, "Unveiling new magnetic phases of undoped and doped manganites,"

Phys. Rev. Lett., vol. 90, p. 247203, Jun 2003. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.90.247203

[371] J. Salafranca and L. Brey, "Phase diagram and incommensurate phases in undoped manganites," *Phys. Rev. B*, vol. 73, p. 024422, Jan 2006. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.73.024422

[372] B. R. K. Nanda and S. Satpathy, "Magnetic and orbital order in lamno$_3$ under uniaxial strain: A model study," *Phys. Rev. B*, vol. 81, p. 174423, May 2010. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.81.174423

[373] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, "Automatic differentiation in machine learning: a survey," *Journal of Machine Learning Research*, vol. 18, no. 153, pp. 1–43, 2018. [Online]. Available: http://jmlr.org/papers/v18/17-468.html

[374] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019.

[Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf

[375] R. M. Noack, D. J. Scalapino, and R. T. Scalettar, "Charge-density-wave and pairing susceptibilities in a two-dimensional electron-phonon model," *Phys. Rev. Lett.*, vol. 66, pp. 778–781, Feb 1991. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.66.778

[376] M. Hohenadler and G. G. Batrouni, "Dominant charge density wave correlations in the holstein model on the half-filled square lattice," *Phys. Rev. B*, vol. 100, p. 165114, Oct 2019. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.100.165114

[377] I. Esterlis, S. A. Kivelson, and D. J. Scalapino, "Pseudogap crossover in the electron-phonon system," *Phys. Rev. B*, vol. 99, p. 174516, May 2019. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.99.174516

[378] S. Allen and J. Cahn, "Ground state structures in ordered binary alloys with second neighbor interactions," *Acta Metallurgica*, vol. 20, no. 3, pp. 423–433, 1972. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0001616072900375

[379] Z. Carson, B. C. Seymour, and K. Yagi, "Future prospects for probing scalar–tensor theories with gravitational waves from mixed binaries," *Classical*

and Quantum Gravity, vol. 37, no. 6, p. 065008, Feb 2020. [Online]. Available: http://dx.doi.org/10.1088/1361-6382/ab6a1f

[380] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, "Real-time gravitational wave science with neural posterior estimation," *Physical Review Letters*, vol. 127, no. 24, Dec. 2021. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.127.241103

[381] Z. Lyu, N. Jiang, and K. Yagi, "Constraints on Einstein-dilation-Gauss-Bonnet gravity from black hole-neutron star gravitational wave events," *Phys. Rev. D*, vol. 105, no. 6, p. 064001, 2022, 2201.02543, [Erratum: Phys.Rev.D 106, 069901 (2022), Erratum: Phys.Rev.D 106, 069901 (2022)].

[382] S. Nojiri, S. D. Odintsov, and M. Sasaki, "Gauss-Bonnet dark energy," *Phys. Rev. D*, vol. 71, p. 123509, 2005, hep-th/0504052.

[383] K. Yagi, "A New constraint on scalar Gauss-Bonnet gravity and a possible explanation for the excess of the orbital decay rate in a low-mass X-ray binary," *Phys. Rev.*, vol. D86, p. 081504, 2012, 1204.4524.

[384] G. Antoniou, A. Bakopoulos, and P. Kanti, "Black-Hole Solutions with Scalar Hair in Einstein-Scalar-Gauss-Bonnet Theories," *Phys. Rev. D*, vol. 97, no. 8, p. 084037, 2018, 1711.07431.

[385] G. Antoniou, A. Bakopoulos, and P. Kanti, "Evasion of no-hair theorems and novel black-hole solutions in gauss-bonnet theories," *Physical Review Letters*,

vol. 120, no. 13, Mar 2018. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.120.131102

[386] W. E. East and J. L. Ripley, "Dynamics of spontaneous black hole scalarization and mergers in einstein-scalar-gauss-bonnet gravity," *Physical Review Letters*, vol. 127, no. 10, Sep 2021. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.127.101102

[387] W. E. East and J. L. Ripley, "Evolution of einstein-scalar-gauss-bonnet gravity using a modified harmonic formulation," *Physical Review D*, vol. 103, no. 4, Feb 2021. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.103.044040

[388] K. Yagi, L. C. Stein, N. Yunes, and T. Tanaka, "Post-Newtonian, Quasi-Circular Binary Inspirals in Quadratic Modified Gravity," *Phys. Rev. D*, vol. 85, p. 064022, 2012, 1110.5950, [Erratum: Phys.Rev.D 93, 029902 (2016)].

[389] K. Yagi, L. C. Stein, and N. Yunes, "Challenging the Presence of Scalar Charge and Dipolar Radiation in Binary Pulsars," *Phys. Rev.*, vol. D93, no. 2, p. 024010, 2016, 1510.02152.

[390] E. Berti, K. Yagi, and N. Yunes, "Extreme Gravity Tests with Gravitational Waves from Compact Binary Coalescences: (I) Inspiral-Merger," *Gen. Rel. Grav.*, vol. 50, no. 4, p. 46, 2018, 1801.03208.

[391] M. Okounkova, "Numerical relativity simulation of GW150914 in Einstein dilaton Gauss-Bonnet gravity," *Phys. Rev. D*, vol. 102, no. 8, p. 084046, 2020, 2001.03571.

[392] B. Shiralilou, T. Hinderer, S. Nissanke, N. Ortiz, and H. Witek, "Post-newtonian gravitational and scalar waves in scalar-gauss-bonnet gravity," 2021," 2105.13972.

[393] S. E. Perkins, R. Nair, H. O. Silva, and N. Yunes, "Improved gravitational-wave constraints on higher-order curvature theories of gravity," *Physical Review D*, vol. 104, no. 2, Jul 2021. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.104.024060

[394] C. Cutler and E. E. Flanagan, "Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral waveform?" *Phys. Rev. D*, vol. 49, pp. 2658–2697, Mar 1994. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.49.2658

[395] A. Buonanno, B. R. Iyer, E. Ochsner, Y. Pan, and B. S. Sathyaprakash, "Comparison of post-newtonian templates for compact binary inspiral signals in gravitational-wave detectors," *Physical Review D*, vol. 80, no. 8, Oct 2009. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.80.084043

[396] N. Jiang, "Testing general relativity with gravitational waves from compact binaries," Ph.D. dissertation, University of Virginia, 2023.