Machine learning predictions for local electronic properties of disordered correlated electron systems

Yi-Hsuan Liu [1,2] Sheng Zhang [3] Puhan Zhang,[3] Ting-Kuo Lee [1,2,4] and Gia-Wei Chern[3]

[1]*Department of Physics, National Tsing Hua University, Hsinchu 30013, Taiwan*
[2]*Institute of Physics, Academia Sinica, Nankang 11529, Taiwan*
[3]*Department of Physics, University of Virginia, Charlottesville, Virginia 22904, USA*
[4]*Department of Physics, National Sun Yat-sen University, Kaohsiun 80424, Taiwan*

We present a scalable machine learning (ML) model to predict local electronic properties such as on-site electron number and double occupation for disordered correlated electron systems. Our approach is based on the locality principle, or the nearsightedness nature, of many-electron systems, which means local electronic properties depend mainly on the immediate environment. A ML model is developed to encode this complex dependence of local quantities on the neighborhood. We demonstrate our approach using the square-lattice Anderson-Hubbard model, which is a paradigmatic system for studying the interplay between Mott transition and Anderson localization. We develop a lattice descriptor based on the group-theoretical method to represent the on-site random potentials within a finite region. The resultant feature variables are used as input to a multilayer fully connected neural network, which is trained from data sets of variational Monte Carlo (VMC) simulations on small systems. We show that the ML predictions agree reasonably well with the VMC data. Our work underscores the promising potential of ML methods for multiscale modeling of correlated electron systems.

I. INTRODUCTION

The growing field of machine learning (ML) is rapidly revolutionizing the scientific research. In materials science and condensed-matter physics, the ML methods have opened up many research possibilities which are beyond conventional approaches. In particular, the introduction of ML techniques has reinvigorated the field of multiscale modeling of complex materials. A fundamental issue in multiscale simulations is the trade-off between efficiency and accuracy of the numerical methods. In particular, an accurate treatment of complex quantum materials often requires time-consuming calculations, which significantly limit the accessible system sizes and time scales. Recent advances in supervised learning are able to bridge this gap by providing an efficient, yet accurate, model to approximate the outcomes of complicated quantum calculations, thus enabling large-scale simulations. A supervised ML model is essentially a complex high-dimensional function with numerous tunable parameters, whose optimal values can be determined from a large number of training data sets. Among the various ML models, deep neural networks (NN) [1–6] represent the most powerful and versatile tools, which, in principle, can approximate any continuous function with arbitrary accuracy [7–9].

Perhaps the best example of large-scale modeling enabled by ML is *ab initio* molecular dynamics (MD) simulations that are based on ML force-field models [10–22]. Contrary to classical MD simulations with empirical force fields, the atomic forces in quantum MD are computed by integrating out electrons on the fly as the atomic trajectories are generated [23]. Over the past decade, various ML models have

been developed to emulate the time-consuming first-principles electronic structure calculations based on, e.g., the density functional theory (DFT). It is worth noting that an ML model here is essentially a complicated classical force-field model, trained from the DFT solutions. ML-based MD simulations thus enjoy the efficiency of classical MD, while maintaining the accuracy of first-principles calculations.

The success of the ML method in quantum MD simulations has further motivated similar ML approaches to achieve large-scale dynamical simulations in correlated electron systems [24–26]. For example, ML methods have been applied to enable large-scale quantum Landau-Lifshitz dynamics simulations of the double-exchange model [24,25], which describes itinerant electrons interacting with magnetic moments of localized d electrons. In such simulations, the exchange forces acting on spins are obtained by solving a tight-binding electron Hamiltonian at every time step, which could be prohibitively expensive for large systems. ML-based exchange-force models are developed to achieve large-scale dynamical simulations of double-exchange systems [24,25]. In another recent work [26], a multilayer NN is employed to enable large-scale quantum kinetic Monte Carlo simulations of the Falicov-Kimball model, which is another canonical example of correlated electron systems. Furthermore, NN has been employed to learn the Gutzwiller solution of Hubbard-type models [22,27], thus enabling large-scale MD simulations of Mott metal-insulator transition in atomic liquids [22].

It is worth noting that the unprecedented efficiency of ML based multiscale modeling is due to the linear scalability of electronic structure calculations enabled by ML methods. This

is in stark contrast to most conventional approaches which scale rather poorly with the system size. For example, exact diagonalization, which is central to effective single-particle methods including Hartree-Fock, Gutzwiller, and DFT, scales cubically $\mathcal{O}(N^3)$ with the system size N. More sophisticated many-body techniques, such as quantum Monte Carlo and configuration interaction, scale even more poorly with increasing system size.

Fundamentally, as first pointed out by Kohn, linear-scaling electronic structure methods are possible mainly because of the locality nature or "nearsightedness" principle [28,29] of many-electron systems. Indeed, in the pioneering work of Behler and Parrinello [10], the locality principle was tacitly assumed in their construction of the NN interatomic potential model. The nearsightedness of electronic systems here does not rely on the existence of well localized Wannier-type wave functions, which only exist in large-gap insulators. Instead, Kohn's locality principle mainly refers to observable quantities such as the correlation function of many-particle systems; the principle is generally a consequence of wave-mechanical destructive interference. It requires the presence of many particles, which need not be interacting.

Although other linear-scaling methods, notably the kernel polynomial method (KPM) [30,31], have been developed for electronic structure calculation, they are mostly restricted to solving the effective single-electron problem. As a result, they cannot be directly applied to strongly interacting or correlated systems, such as the Hubbard or t-J models. Further approximations are required to reduce the many-body Hamiltonian to a single-particle one, which can then be solved by the $\mathcal{O}(N)$ methods such as KPM. On the other hand, assuming nearsightedness for many-electron systems, ML offers a general approach to achieve linear scalability without further approximation. The key is to develop a ML model that can efficiently and accurately emulate the many-body calculations based on a finite local environment.

In this paper, we demonstrate such a scalable ML model for a disordered electron system with Hubbard repulsion. Specifically, we consider the type of on-site potential disorder as described by the Anderson model of localization. Similar on-site disorder can also arise dynamically as in the adiabatic Holstein model. A neural network (NN) model is developed to directly predict local electronic properties such as electron occupation number and double occupancy, based on the disorder configuration in the immediate neighborhood. The NN model is trained from quantum variational Monte Carlo simulations on small lattices. We show that the trained NN model gives accurate predictions on much larger systems of varying Hubbard repulsion and disorder strength. Our work demonstrates the transferability and scalability of the ML approach to Hubbard-type models, paving the way for their applications to multiscale modelings of correlated electron systems.

The remainder of the paper is organized as follows. In Sec. II, we present the variational Monte Carlo (VMC) method and details of its implementation to the disordered Hubbard model. Our focus is on the real-space electronic inhomogeneity and how the local electronic properties depend on the neighborhood disorder configuration. The structure of the NN and the training process are discussed in Sec. III. A

lattice descriptor, based on the group-theoretical method, is developed to incorporate the discrete lattice symmetry into the NN model. Comparisons of the ML predictions versus the VMC results on validation data sets are presented in Sec. IV. Moreover, we discuss the application of the ML model to the Mott transition of the Anderson-Hubbard model. Finally, a summary and discussion for future work are given in Sec. V.

II. VARIATIONAL MONTE CARLO METHOD FOR HUBBARD MODEL

We consider the following two-dimensional Hubbard model with an on-site disorder, also known as the Anderson-Hubbard (AH) model [32–34]:

$$\hat{H} = -t \sum_{\langle ij \rangle, \sigma} \hat{c}_{i,\sigma}^{\dagger} \hat{c}_{j,\sigma} + \sum_i \varepsilon_i \hat{n}_i + \sum_i U \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}. \quad (1)$$

Here t is the nearest-neighbor electron hopping constant, $\hat{c}_{i,\sigma}^{\dagger}$ is the electron creation operator with spin $\sigma = \uparrow, \downarrow$ at site i, $\hat{n}_{i,\sigma} \equiv \hat{c}_{i,\sigma}^{\dagger} \hat{c}_{i,\sigma}$ is the corresponding number operator, and $\hat{n}_i = \hat{n}_{i\uparrow} + \hat{n}_{i\downarrow}$ is the total number operator. The on-site Coulomb repulsion is described by the last term where U is the Hubbard parameter. The second term describes an on-site or potential disorder considered by Anderson; ε_i denotes the on-site potential, which is a random number drawn uniformly from the interval $[-W/2, +W/2]$. The strength of the disorder is thus characterized by the parameter W.

It is worth noting that such on-site disorder can also be of dynamical origin, e.g., due to lattice distortions. A canonical example is the Holstein-Hubbard model [35–37], in which a scalar dynamical variable Q_i is introduced to describe local lattice distortion of A_1 symmetry, such as the breathing mode of oxygen octahedron, associated with site i. The random on-site potential $\varepsilon_i = -gQ_i$ comes from the deformation potential coupling between electrons and lattice: $\mathcal{H}_{\text{el-ph}} = -g \sum_i \hat{n}_i Q_i$, where g is the coupling constant [35]. As will be discussed below, the NN model developed for the AH Hamiltonian can also be used to predict the effective forces acting on Q_i in the adiabatic limit.

The AH model is one of the canonical electron systems. In addition to exhibiting rich phase diagrams [38–44], the AH model offers a simple platform to study the interplay between two important mechanisms of metal-insulator transition, namely the Anderson localization versus correlation-induced Mott transition. The AH model has been extensively studied by a wide variety of numerical methods. Depending on the numerical treatments of the spatial disorder, there are two types of approaches to this problem: the self-consistent theories and the real-space methods. The most representative example of the former approach is the generalization of dynamical mean field theory (DMFT) [45] to include the disorder effects. While several self-consistent theories have been developed for disordered electronic systems, the typical medium theory (TMT) [46] proves successful to capture the Anderson localization phenomena and can be readily combined with DMFT [38,39]. As such self-consistent methods are free of finite-size effect, they provide an overall theoretical picture of the AH model in the thermodynamic limit.

The nonmagnetic phase diagram of the AH model obtained from the TMT-DMFT method includes three distinct phases:

a correlated metallic phase, a Mott insulating phase, and an Anderson insulating phase [38,39]. Importantly, the two insulating phases of the AH model have very different characters. The Mott insulator results from the strong correlation effect which prohibits electrons from hopping to the neighboring sites. On the other hand, strong disorder weakens the constructive interference that allows an electron wave packet to propagate coherently in a periodic potential, leading to the Anderson insulator. TMT-DMFT calculation shows that these two insulating phases are continuously connected [38,39].

On the other hand, the real-space approach, although limited by finite-size effect, can better describe the spatial fluctuations and correlations of the inhomogeneous electronic state due to the disorder, especially for low-dimensional systems. In this real-space approach, the AH Hamiltonian for a particular disorder configuration on a finite lattice is solved by many-body techniques ranging from unrestricted Hartree-Fock [40,47–49] and Gutzwiller [41,50] mean-field type theories to small-cluster exact diagonalization [51–53], inhomogeneous or statistical DMFT [33,54–56], and quantum determinent Monte Carlo [57–60] as well as variational Monte Carlo simulations [61–64]. The calculation results are then averaged over different disorder realizations. In particular, extensive large-scale simulations based on the Gutzwiller/slave-boson methods showed that the strong spatial inhomogeneity gives rise to an electronic Griffiths phase that precedes the metal-insulator transition [41].

In this work, we are interested in the locality of electronic properties and collective behaviors such as the double occupancy. Specifically, our goal is to develop a NN model that can accurately predict on-site quantities based on disorder configuration within a finite neighborhood. To capture these spatial inhomogeneities, we employ the real-space variational Monte Carlo (VMC) method to solve the square-lattice AH model. As we are interested mainly in the competition between localization effect and electron correlation, we restrict ourselves to the paramagnetic phases to avoid complications due to long-range magnetic order. We note that a similar ML approach to disordered Hubbard models has been demonstrated based on the data set from the real-space Gutzwiller/slave-boson method [27]. However, the spatial correlation between local electronic properties is ignored due to the Gutzwiller approximation. The VMC method, on the other hand, can properly take into account these spatial correlations, which are important to test the locality of the ML models.

Next we outline the VMC method for the square-lattice AH model. Following the previous works developed for Hubbard-type models, we consider a variational wave function obtained by applying a Gutzwiller factor $\hat{\mathcal{G}}$ and a Jastrow factor $\hat{\mathcal{J}}$ to a Slater determinant $|\Phi_0\rangle$ [61,62,65]:

$$|\Psi\rangle = \hat{\mathcal{G}}\hat{\mathcal{J}}|\Phi_0\rangle. \tag{2}$$

The uncorrelated Slater determinant state $|\Phi_0\rangle$ is computed from the eigenstates of the following quadratic Hamiltonian:

$$\hat{H}_{\mathrm{MF}} = -t\sum_{ij,\sigma}\hat{c}_{i,\sigma}^\dagger\hat{c}_{j,\sigma} + \sum_{i,\sigma}\tilde{\varepsilon}_{i,\sigma}\,\hat{n}_{i,\sigma}, \tag{3}$$

which can be viewed as a mean-field approximation to the AH model. The on-site energies $\tilde{\varepsilon}_{i,\sigma}$ are part of the variational

parameters. As mentioned above, we focus on the paramagnetic phases and assume spin-independent local energies $\tilde{\varepsilon}_{i,\uparrow} = \tilde{\varepsilon}_{i,\downarrow} = \tilde{\varepsilon}_i$. To account for the crucial on-site electron correlation, the Gutzwiller correlator is introduced [65–67]:

$$\hat{\mathcal{G}} = \prod_i [1 - (1 - g_i)\,\hat{n}_{i\uparrow}\hat{n}_{i\downarrow}] \equiv \prod_i \hat{\mathcal{G}}_i, \tag{4}$$

where g_i are another set of variational parameters that control the on-site double occupancy; as $g_i \to 0$, double-occupied states are completely projected out. The long-range Jastrow factor is defined as [65]

$$\hat{\mathcal{J}} = \exp\left[\frac{-1}{2}\sum_{ij}v_{ij}(\hat{n}_i - 1)(\hat{n}_j - 1)\right]. \tag{5}$$

The Jastrow operator, parametrized by another set of variational parameters v_{ij}, introduces correlation of charge fluctuations $\delta n_i = n_i - 1$ at different sites. For an arbitrary inhomogeneous state, the parameter v_{ij} in principle depends on both sites i and j, giving rise to a total of N^2 parameters to be optimized. In order to make the numerical calculation feasible [61,62], we assume translational invariance for these parameters, i.e., $v_{ij} = v(|\mathbf{r}_i - \mathbf{r}_j|)$, and consider v_{ij} for different pairs up to the eighth nearest neighbors.

The optimal variational parameters $\tilde{\varepsilon}_i$, g_i, and v_{ij} are obtained by minimizing the variational energy $E_{\mathrm{var}} = \langle\Psi|\hat{\mathcal{H}}|\Psi\rangle/\langle\Psi|\Psi\rangle$ using the stochastic reconfiguration method [68,69]. The evaluation of the various expectation values computed from $|\Psi\rangle$ is computed based on Monte Carlo simulations. In the following we apply the VMC methods to study the ground state of the AH model with various disorder strength $W = 6t, 10t, 14t, 18t$ and Hubbard parameter $U = 4t, 8t, 10t, 12t, 16t$. We focus on the case of half filling, where the number of electrons $N_e = N$. Periodic and antiperiodic boundary conditions along x and y directions, respectively, are used. Depending on the convergence, 500–2000 iterations of the stochastic reconfiguration were used to optimize the variational parameters. For each iteration, the various expectation values were obtained from approximately 10^5 Monte Carlo samplings. For each U and W combination, 50 independent realizations were used to generate the training data sets.

The results of the VMC simulation are summarized in Fig. 1, which shows the scatter plots of various local quantities versus on-site potentials ε_i for three different Hubbard U. Here we are interested in the following local quantities: (a) the local electron filling fraction f_i, (b) the electron on-site self-energy Σ_i, (c) the double occupancy D_i, and (d) the local quasiparticle weight q_i. Each point in the scatter plots corresponds to the data of a lattice site from the VMC solution of a given disorder realization. The local electron number is defined as

$$n_i = \langle\hat{n}_i\rangle = \langle\hat{n}_{i,\uparrow}\rangle + \langle\hat{n}_{i,\downarrow}\rangle. \tag{6}$$

For paramagnetic phases, we have $\langle\hat{n}_{i,\uparrow}\rangle = \langle\hat{n}_{i,\downarrow}\rangle$. As shown in Fig. 1(a), this local electron density is reduced with increasing on-site potential, which is expected. At large $U = 16t$, the electron number develops a plateau at $n_i = 1$ for small on-site energies $|\varepsilon_i| \lesssim 6t$. The plateau thus represents lattice
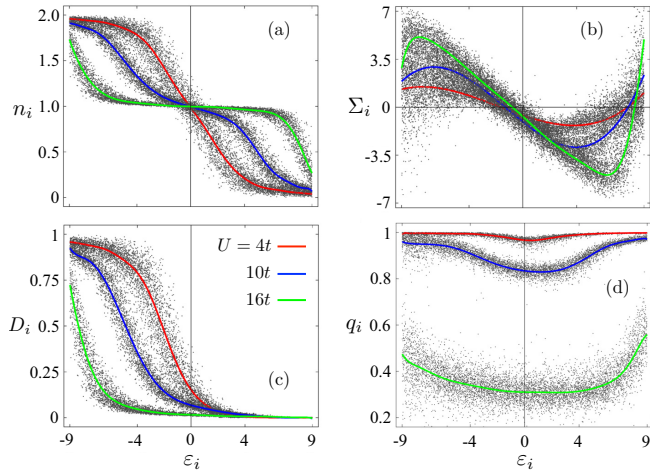
FIG. 1. Summary of the VMC solution for the AH model on a 16×16 square lattice. The four panels show the scatter diagram of (a) on-site electron number n_i, (b) variational local self-energy Σ_i, (c) double occupancy D_i, and (d) effective quasiparticle weight q_i versus the random on-site potential ε_i. The data points were obtained from calculations with random strength $W/t = 6, 10, 14, 18$ and three different $U = 4t, 10t$, and $16t$. The overall dependence of the various local quantities on the on-site potential is highlighted by the three colored curves obtained using polynomial regression with up to 16th-order polynomials. The red, blue, and green curves correspond to $U = 4t, 10t$, and $16t$, respectively.

sites with localized electrons, where the corresponding double occupancy also tends to zero. The electronic properties outside the plateau are dominated by the strong local potential, leading to either almost filled or empty sites. For systems with large $W \gtrsim U$, the interplay between electron correlation and disorder thus gives rise to spatially very inhomogeneous states with coexisting Mott regions and Anderson insulator, consistent with the two-fluid behavior of the Mott-Anderson insulator [39,41,55].

Another quantity of interest is the local self-energy, which is defined as the difference between the renormalized and bare on-site potentials:

$$\Sigma_i = \tilde{\varepsilon}_i - \varepsilon_i. \qquad (7)$$

As shown in Fig. 1(b), this self-energy clearly anticorrelates with the random on-site potential ε_i, meaning that the renormalization due to Σ_i is such that the effective potential $\tilde{\varepsilon}_i$ tends to vanish. Indeed, screening of impurity potential by the electron gas has been demonstrated even in the weak interaction regime. As demonstrated in both real-space DMFT and slave-boson studies on the 2D AH model [41,50,54], the disorder screening persists also in the strong correlation regime, albeit with a rather different nature. In particular, this interplay leads to the enhancement of metallicity in an intermediate regime where the interactions and the disorder are of comparable magnitude.

To detect the correlation-induced electron localization, we compute the local double occupancy from VMC:

$$D_i = \langle \hat{n}_{i,\uparrow} \hat{n}_{i,\downarrow} \rangle. \qquad (8)$$

As expected, the average double occupation is reduced with increasing Hubbard U; see Fig. 1(c). It should be noted that

the large D_i persisting even at large U is due to the deep on-site potential which traps two electrons of opposite spins. On the other hand, the small value of double occupancy at large positive ϵ is a result of empty site, instead of strong correlation. To properly distinguish these two scenarios, we also compute an effective local quasiparticle weight defined as

$$q_i = \frac{4g_i}{(1 + g_i)^2}, \qquad (9)$$

where g_i is obtained from the VMC optimization. For homogeneous electron liquids, the quantity q characterizes the discontinuity of the momentum distribution function at the Fermi surface in the Gutzwiller approximation [66,67]. And, for inhomogeneous systems, q_i plays the role of renormalizing the electron hopping. We emphasize that q_i is not an exact definition, but is meant to be a qualitative indicator of the local quasiparticle weight. As shown in Fig. 1(d), overall the quasiparticle weights decrease with increasing Hubbard U. On the other hand, stronger disorder, e.g., sites with $|\varepsilon_i| \lesssim U$, preserves the itinerant nature of electrons.

The results summarized in Fig. 1 also indicate a continuous trend with respect to increasing disorder for a fixed Hubbard parameter. We note that these scatter plots include VMC data from various disorder strengths $W/t = 6, 10, 14$, and 18. As a result, some of the data points correspond to the Mott-dominated insulating phase, i.e., those $W < U$, while others belong to the Anderson insulator ($W > U$). Nonetheless, the collection of all data points clearly exhibit an overall trend as a function of the on-site potential ε as highlighted by the smooth curves in Fig. 1, which are obtained using polynomial regression up to 18th order. These continuous functions relating the local quantities $n_i, D_i, \ldots$ to the on-site potential ε_i thus provide the zeroth order prediction. Deviations from this smooth curve, as clearly indicated by the scattered points in the figure, thus can be viewed as due to the influence of neighboring random potentials. The effects of the neighborhood disorder are accounted for by the ML model to be described below.

III. MACHINE LEARNING MODEL

We next describe the framework of a scalable ML model for predicting the local electronic properties of disordered Hubbard systems. Our approach is similar to the ML modeling of structure-property relationships in materials science, which play an increasingly important role in accelerated materials discovery. A particularly important application, as mentioned in Sec. I, is the ML modeling of force field for *ab initio* MD simulations. The central idea is to develop a ML model which can accurately predict the force acting on individual atoms and other local properties based on the immediate chemical environment.

A widely used scheme, first pioneered by Behler and Parrinello [10], focuses on local energies. First, the total energy of the system is partitioned into local contributions, $E = \sum_i E_i$, where E_i is called the atomic energy associated with the ith atom. Next, based on locality principle, the atomic energy E_i is assumed to depend on the atomic configuration $\mathcal{C}_i$ in the neighborhood of atom i. Specifically, this local chemical environment is given by $\mathcal{C}_i = \{(Z_j, \mathbf{R}_j) \mid |\mathbf{R}_j - \mathbf{R}_i| < r_c\}$, where
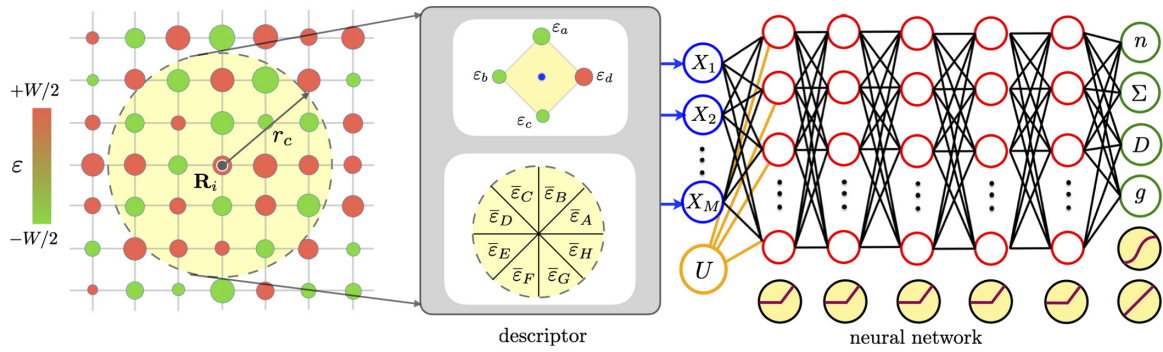
FIG. 2. Schematic diagram of the ML model for the vector function $\mathcal{F}(\mathcal{C}_i; U)$ defined in Eq. (11) for the AH model. The input of the ML model is the disorder configuration $\mathcal{C}_i$ in the neighborhood of a given site i up to a cutoff radius r_c. The output are local electron properties including on-site electron number n_i, self-energy Σ_i, double-occupancy D_i, and Gutzwiller parameter g_i. There are two central components of the ML model: the descriptor and the learning model, which is based on a multilayer neural network. The ReLU activation function is used in the five hidden-feature extraction layers with $256 \times 128 \times 64 \times 32$ nodes. Note the special input node corresponding to the Hubbard parameter U.

Z_j is the atomic number of atom j at position $\mathbf{R}_j$ and r_c is a cutoff radius (soft cutoff is often used). Finally, the complex dependence of atomic energy on the local environment, $E_i = f(\mathcal{C}_i)$, is to be approximated by an ML model, which is trained from electronic structure calculations such as DFT on small systems.

The atomic forces in this scheme are obtained from the derivatives of the total energy: $\mathbf{F}_i = -\partial E / \partial \mathbf{R}_i$. It is worth noting that, instead of direct prediction of atomic forces, which are vectors, the Behler-Parrinello type scheme focuses on the local atomic energy. The fact that the atomic energy, as a scalar, is readily invariant under transformations such as rotations and translations of the system makes it easier to incorporate the symmetry properties into the ML model. Moreover, as forces are derived from an effective energy, this approach also ensures a conservative force field, which is important for quantum MD under Born-Oppenheimer approximation [23]. Most importantly, this ML method is both transferrable and scalable as exactly the same ML model can be used for much larger systems. We note in passing that a similar approach has recently been developed for generalized force fields in condensed matter systems [70].

Here we adapt this ML approach to develop a neural-network (NN) model for the prediction of local electronic properties of the AH model. Again, based on the nearsightedness of many-electron systems [28,29], we assume local electronic properties such as electron density and double occupancy at site i only depend on the neighborhood disorder configuration $\mathcal{C}_i$. Explicitly, it is defined as

$$\mathcal{C}_i = \left\{ \varepsilon_j \,\middle|\, |\mathbf{R}_j - \mathbf{R}_i| < r_c \right\}, \qquad (10)$$

where r_c is a predefined cutoff radius. The complex dependence of local electronic properties on local energies $\mathcal{C}_i$ in the neighborhood is represented by a vector function $\mathcal{F}(\cdot)$:

$$\mathcal{Q}_i = (n_i, \Sigma_i, D_i, g_i, \ldots) = \mathcal{F}(\mathcal{C}_i; U). \qquad (11)$$

For convenience, here we arrange the local quantities associated with site i into a vector or array $\mathcal{Q}_i$. The definitions of these quantities are discussed in Sec. II. Although the approach discussed here can be straightforwardly generalized to include more local properties, here we mainly are concerned with the four quantities shown in Fig. 1. Moreover, we have explicitly included the dependence on the Hubbard parameter U. By setting the nearest-hopping constant $t = 1$, which serves as the unit for energy, the vector function $\mathcal{F}(\cdot)$ is universal for the AH model of a given electron filling fraction $n = N_e/N$. A ML model, shown in Fig. 2, is developed to approximate this universal function for the case of half filling. The input of the model, which is the disorder configuration $\mathcal{C}_i$ in the neighborhood of site i, is first transformed into a set of feature variables $\{x_1, x_2, \ldots, x_M\}$ called the descriptor. These feature variables, along with the Hubbard parameter U, are then fed into the neural network, which produces an array of the local quantities $\mathcal{Q}_i$ at the output. Details of these two central components, namely the descriptor and the neural network, are discussed in the following.

A. Lattice descriptor

It is worth noting that, despite the powerful approximation capability of NNs, symmetries of the electron Hamiltonian are not automatically included in the ML model. One approach to incorporate the required symmetry into the ML model is through the construction of a proper representation of the local environment, which is then used as input to the learning model. A good representation is *invariant* with respect to transformations of the symmetry group of the system. This crucial step of the ML model, namely the construction of the proper representation, is often referred to as feature engineering and the resultant feature variables, also called the generalized coordinates, are termed a descriptor [71–73].

A sescriptor is also crucial to the ML interatomic potentials for quantum MD simulations. A proper description of the chemical environment should respect the fundamental symmetries of interatomic interactions, which are invariant under translational, rotational, and permutational transformations. Over the past decade, a number of descriptors have been proposed together with the learning models based on them [10,11,16,71–81]. A popular atomic descriptor used in many ML models is the atom-centered symmetry functions (ACSFs) built from the two-body (relative distances) and three-body (relative angles) invariants of the atomic
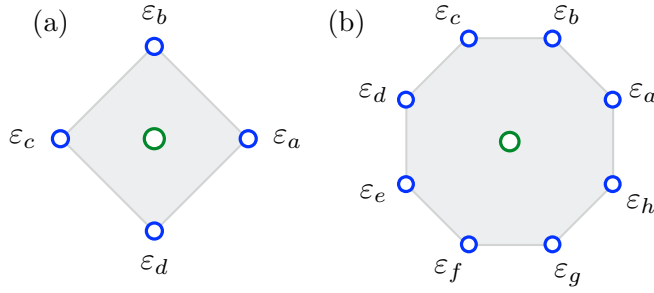
FIG. 3. Schematic diagrams of (a) four-site and (b) eight-site blocks of neighbor sites.

configurations [10,76]. The group-theoretical method, on the other hand, offers a more controlled approach to the construction of atomic representation based on the power-spectrum and bispectrum coefficients [11,72]. It is worth noting that the research of the atomic descriptor is an active ongoing field.

A general theory and several specific implementations of descriptors in condensed matter systems, especially for lattice models, have recently been presented in Ref. [70]. In particular, the group-theoretical bispectrum method was generalized to systematically generate feature variables that are invariant under symmetry operations of the on-site point group [27,70]. Here we apply this method to develop a descriptor of the AH model. To this end, we first note that the on-site potentials in the neighborhood $\mathcal{C}_i$ form a high-dimensional reducible representation of the site-symmetry group, which in the case of a square lattice is equivalent to the point group D_4. The first step of finding invariants under site symmetry is to decompose the neighborhood $\mathcal{C}_i$ into irreducible representations (IRs) of the symmetry group.

While this decomposition can be done following standard procedure in group theory, the calculation here can be greatly simplified by noting that neighboring sites j with the same distance $R_{ij} = |\mathbf{R}_j - \mathbf{R}_i|$ from the center site i form a closed representation of the point group. In the case of D_4, the size of these invariant neighbor blocks is either four or eight; see Fig. 3. The four-site block can be decomposed as $4 = 1A_1 \oplus 1B_1 \oplus 1E$. The expansion coefficients of each IR are

$$f_{A_1} = \varepsilon_a + \varepsilon_b + \varepsilon_c + \varepsilon_d,$$
$$f_{B_1} = \varepsilon_a - \varepsilon_b + \varepsilon_c - \varepsilon_d,$$
$$\boldsymbol{f}_E = (\varepsilon_a - \varepsilon_c, \ \varepsilon_b - \varepsilon_d). \tag{12}$$

The decomposition of the eight-site block is $8 = 1A_1 + 1B_1 + 1A_2 + 1B_2 + 2E$, with the following coefficients for each IR:

$$f_{A_1} = \varepsilon_a + \varepsilon_b + \varepsilon_c + \varepsilon_d + \varepsilon_e + \varepsilon_f + \varepsilon_g + \varepsilon_h,$$
$$f_{A_2} = \varepsilon_a - \varepsilon_b + \varepsilon_c - \varepsilon_d + \varepsilon_e - \varepsilon_f + \varepsilon_g - \varepsilon_h,$$
$$f_{B_1} = \varepsilon_a - \varepsilon_b - \varepsilon_c + \varepsilon_d + \varepsilon_e - \varepsilon_f - \varepsilon_g + \varepsilon_h,$$
$$f_{B_2} = \varepsilon_a + \varepsilon_b - \varepsilon_c - \varepsilon_d + \varepsilon_e + \varepsilon_f - \varepsilon_g - \varepsilon_h,$$
$$\boldsymbol{f}_{(E,1)} = (\varepsilon_a + \varepsilon_b - \varepsilon_e - \varepsilon_f, \ -\varepsilon_c - \varepsilon_d + \varepsilon_g + \varepsilon_h),$$
$$\boldsymbol{f}_{(E,2)} = (\varepsilon_c - \varepsilon_d - \varepsilon_g + \varepsilon_h, \ \varepsilon_a - \varepsilon_b - \varepsilon_e + \varepsilon_f). \tag{13}$$

As the neighborhood $\mathcal{C}_i$ contains several such invariant blocks, we expect the same IRs to appear multiple times in the overall

decomposition of $\mathcal{C}_i$. In the following, we label each IR in the decomposition of $\mathcal{C}_i$ as $\Gamma = (\text{T}, r)$, where $\text{T} = A_1, A_2, \ldots$ denotes the symmetry type of the IR and r indicates a different occurrence of the same IR. For convenience, we arrange the expansion coefficients of an IR Γ into a vector $\boldsymbol{f}_\Gamma = (f_{\Gamma,1}, f_{\Gamma,2}, \ldots, f_{\Gamma,n_\Gamma})$, where n_Γ is the dimension of Γ.

The power spectrum of the representation is given by the amplitudes of the IR coefficients

$$p_\Gamma = |\boldsymbol{f}_\Gamma|^2. \tag{14}$$

Since the power spectrum coefficients are obviously invariant under symmetry transformations, they can be used as feature variables for the ML models. However, a descriptor composed only of power spectrum is incomplete since the relative phases between different IRs are ignored. This also means that the descriptor contains spurious symmetries as the transformation of each IR is independent of each other without the phase information. A complete set of feature variables can be obtained from the bispectrum coefficients $b_{\Gamma_1,\Gamma_2,\Gamma_3}$, which are triple products of the expansion coefficients $\boldsymbol{f}_{\Gamma_{1,2,3}}$ based on the Clebsch-Gordan coefficients of the point group. Intuitively, they can be viewed as the analog of the scalar triple product of three-dimensional vectors. Not only are the bispectrum coefficients invariant under symmetry transformations, they can also be used to faithfully reconstruct the original disorder configuration [72,82].

However, a descriptor based on all the bispectrum coefficients is in fact overcomplete as many of them are redundant. Moreover, since the dimension of most IRs of point groups is rather small, the number of bispectrum coefficients is often a very large number, which makes the implementation infeasible. Instead, here we employ the method of reference IR discussed in Ref. [70] to retain the phase information. The central idea is to first construct an eight-dimensional representation of the neighborhood $\mathcal{C}_i$ based on an average of on-site potentials over symmetry-related finite regions. As shown in Fig. 2, an example is given by $(\bar{\varepsilon}_A, \bar{\varepsilon}_B, \ldots, \bar{\varepsilon}_H)$, where each $\bar{\varepsilon}_K$ is given by the average of all on-site ε_j within wedge K.

The decomposition of this eight-dimensional representation $\bar{\varepsilon}_K$ then gives coefficients $f_{A_1}^*, f_{A_2}^*, \ldots, \boldsymbol{f}_E^*$ for each symmetry type. These coefficients $\boldsymbol{f}_\text{T}^*$ are termed the reference IR coefficients. For each IR, an effective phase can be defined by the following inner product:

$$\eta_\Gamma = (\boldsymbol{f}_\Gamma \cdot \boldsymbol{f}_{\text{T}_\Gamma}^*)/|\boldsymbol{f}_\Gamma||\boldsymbol{f}_{\text{T}_\Gamma}^*|, \tag{15}$$

where T_Γ is the symmetry type of IR Γ. The phase η_Γ, which is an inner product of two IR coefficients, is naturally invariant with respect to symmetry operations. More importantly, by including η_Γ in the descriptor, the relative phases between different $\boldsymbol{f}_\Gamma$ can now be inferred through the intermediate reference IR coefficients.

B. Neural network

The various steps of the descriptor discussed above can be summarized as $\mathcal{C}_i \to \boldsymbol{f}_\Gamma \to (p_\Gamma, \eta_\Gamma)$. Crucially, assuming the various local quantities $\mathcal{Q}_i$ of interest depend on the neighborhood through these feature variables,

$$\mathcal{Q}_i = \mathcal{F}(\{p_\Gamma, \eta_\Gamma\}_i; U). \tag{16}$$

The resultant ML model is ensured to preserve the site symmetry of the AH Hamiltonian. As discussed above, this vector function $\mathcal{F}(\cdot)$ is to be implemented using a NN. The basic unit of a NN is a perceptron or artificial neuron. And a NN is essentially a set of nested linear regression functions with nonlinear activation performed by the neurons; see Fig. 2. For a neuron with m input signals, arranged in a vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$, its output is given by $y = \sigma(\boldsymbol{w} \cdot \boldsymbol{x} + b)$, where $\sigma(\cdot)$ is a nonlinear activation function, $\boldsymbol{w} = (w_1, w_2, \ldots, w_m)$ specifies weights for each input, and b denotes a bias. Here each signal x_k is the output of a neuron from a previous layer. In a NN, each line is associated with a weight, while each node (neuron) is assigned a bias. These weight and bias variables are parameters to be optimized through training processes.

We design a fully connected NN with four hidden layers consisting of $\mathsf{N} = 256 \times 128 \times 64 \times 32$ rectified linear unit (ReLU) neurons, i.e., with an activation function $\sigma(x) = \max(0, x)$. We choose these numbers of neurons in order to optimize the memory access in CUDA-based GPU computation. The input layer, specified by a vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_M)$, are given by the standard-scalar transformation, i.e., by removing the mean and scaling to unit variance, of the power spectrum coefficients p_Γ and the relative phases η_Γ. In addition, the Hubbard parameter U is also used as an input to the NN; see Fig. 2. The neurons in the hidden layers perform a series of nonlinear transformations described above on the input data. The outcome is fed forward to be processed by the output neuron with sigmoid activation function for n, D, and g (whose domain is [0,1]) and linear activation function for Σ. The mean squared error (MSE) is used as the loss function:

$$L = \frac{1}{\mathcal{N}} \sum_{k=1}^{\mathcal{N}} (W_n |n_k - \hat{n}_k|^2 + W_\Sigma |\Sigma_k - \hat{\Sigma}_k|^2$$
$$+ W_D |D_k - \hat{D}_k|^2 + W_g |g_k - \hat{g}_k|^2 + \cdots), \quad (17)$$

where $\mathcal{N}$ is the number of training data, symbols with a hat refer to the predicted values, and the various W denote the weights of each output. The method of batch normalization is used to avoid overfitting, with a minimum batch size of 32. We use randomly mixed $36 \times 256 \times 4 \times 5 = 184320$ data samples as the training set. The Adam optimizer with learning rate of 0.001 is applied for the training process. Once the training process is successful, the trained neural network can rapidly predict the $4 \times 256 \times 4 \times 5 = 20480$ test data samples. Equal weights for the output are used in the current model.

It is worth noting that, instead of developing an independent NN for each of the local quantities, here we adopt a multitask ML framework [83,84]. As shown in Fig. 2, our approach is to build one NN that can simultaneously and consistently predict different local electronic properties. This common NN is trained via a loss function L that includes MSE from all four local quantities introduced in Eq. (11). Such a multitask learning approach allows inductive bias to be acquired via the training signals for related additional tasks drawn from the same domain [83]. The benefit of multitask learning is the additional constraints due to the

interdependences between the multiple outputs; what is learned for each task can help other tasks to be learned better.

IV. RESULTS AND DISCUSSION

We first benchmark the trained ML model by comparing its predictions against the results from VMC simulations for all disorder configurations, including both the training and the validation data sets. As shown in Fig. 4, for all four local quantities, the ML predictions agree reasonably well with the VMC calculations. More quantitatively, we plot the histograms of the prediction error defined as $\delta = A_{\mathrm{ML}} - A_{\mathrm{VMC}}$ in the insets of Fig. 4. In all cases, the error is rather narrowly distributed with a MAE, given by the width of the distribution, of the order of less than one percent of the mean values.

It should be noted that the VMC data set is itself noisy, since the quantities computed from VMC simulations are based on Monte Carlo samplings, for example, $n_i = \frac{1}{M} \sum_{i=1}^{M} \langle \Psi_i | \hat{n}_i | \Psi_i \rangle / \langle \Psi_i | \Psi_i \rangle$ for local electron occupation number, where $| \Phi_i \rangle$ is a trial wave function among the Markov chain samplings. The values from VMC thus depend on the number of samplings and other stochastic factors. Similarly, as the optimization of variational parameters, such as local energy renormalization Σ_i and Gutzwiller parameter g_i, are based on derivatives which are computed stochastically, these quantities are also not without uncertainty. The randomness in the data sets thus partially contributes to the error δ in the ML predictions. Of course, the error due to VMC can be systematically reduced by increasing the number of Monte Carlo samplings.
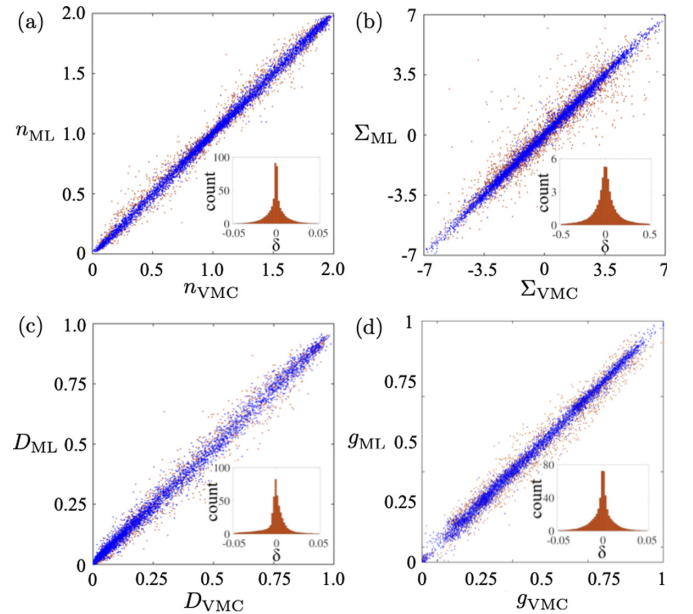

FIG. 4. Comparison of the ML predictions with references obtained from the VMC simulations. (a) Local electron number n, (b) on-site electron self-energy Σ, (c) local double-occupancy D, and (d) the Gutzwiller variational parameter g. The blue and orange data points denote predictions for training and test data sets, respectively. The insets show the normalized count of the error $\delta = A_{\mathrm{ML}} - A_{\mathrm{VMC}}$ defined as the difference between prediction and reference values.
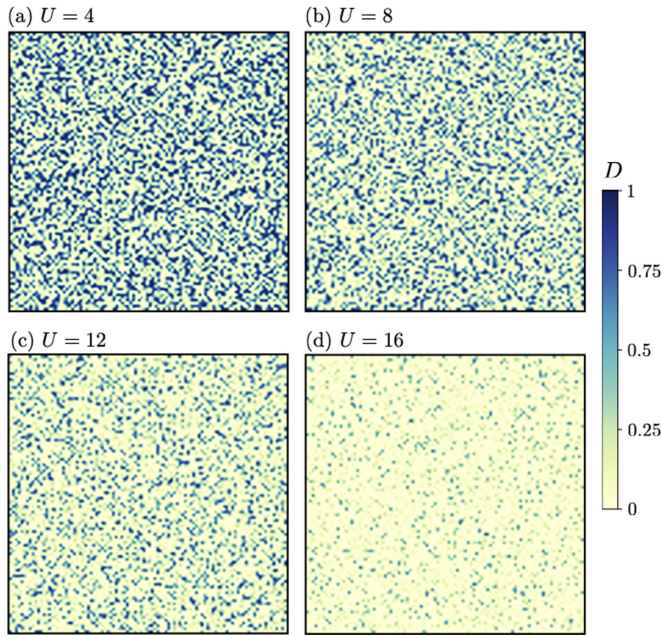
FIG. 5. Distributions of local double-occupancy D_i obtained by applying the trained ML model to a large 100×100 square lattice at various Hubbard parameters.

FIG. 6. (a) Double occupancy D, averaged spatially over sites of the 100×100 lattice and averaged over 50 independent realizations of disorder, as a function of Hubbard parameter U. Panel (b) shows the spatial standard deviation of the double-occupancy $\sigma_D = [\sum_{i=1}^{N}(D_i - \overline{D})^2/N]^{1/2}$, averaged over 50 different disorder realizations. The black dots are the VMC data, while the colored dots are predictions by the trained ML model.

We note in passing that, for Hubbard-type models with electron-lattice coupling through the deformation potential, the ML prediction of the local electron number n_i also provides the forces acting on the atomic displacements or lattice distortions. Our approach is thus an alternative to the more general Behler-Parrinello ML scheme [10]. These include both the Holstein and Jahn-Teller lattice models. For example, the electron-phonon coupling in the Holstein model is described by $\hat{\mathcal{H}}_{\text{el-ph}} = -g \sum_i \hat{n}_i Q_i$, where Q_i denotes the amplitude of local structural distortion. The lattice degrees of freedom play an important role in the emergence of complex electronic textures driven by electron correlation in Hubbard-type models. In such applications, the lattice distortions Q_i can be treated as classical dynamical variables and the electronic forces acting on Q_i can be obtained from the Hellmann-Feynman theorem: $F_i = -\langle \partial \hat{\mathcal{H}}_{\text{el-ph}}/\partial Q_i \rangle = g\langle \hat{n}_i \rangle$. The ML model developed here thus can also be combined with the Langevin dynamics method to enable large-scale dynamical simulations of the Hubbard-Holstein model.

To demonstrate the scalability of our ML approach, we apply the ML model, trained from VMC solutions on a small 16×16 lattice, to compute the real-space electronic properties of the AH model on a 100×100 lattice. Figure 5 shows the profiles of local double-occupancy D_i at various Hubbard U for some random realizations of a large disorder with $W = 18$. Based on the locality of electronic systems, the ML model only depends on on-site potentials of a fixed finite spatial region, e.g., determined by the cutoff radius r_c, independent of the system size N. Consequently, the time complexity of the ML method for computing local electronic properties scales *linearly* with N. The efficiency is thus significantly improved compared with the polynomial scaling $\mathcal{O}(N^s)$ of VMC, where the exponent s ranges from 3 to 6 depending on the specific optimization techniques [69].
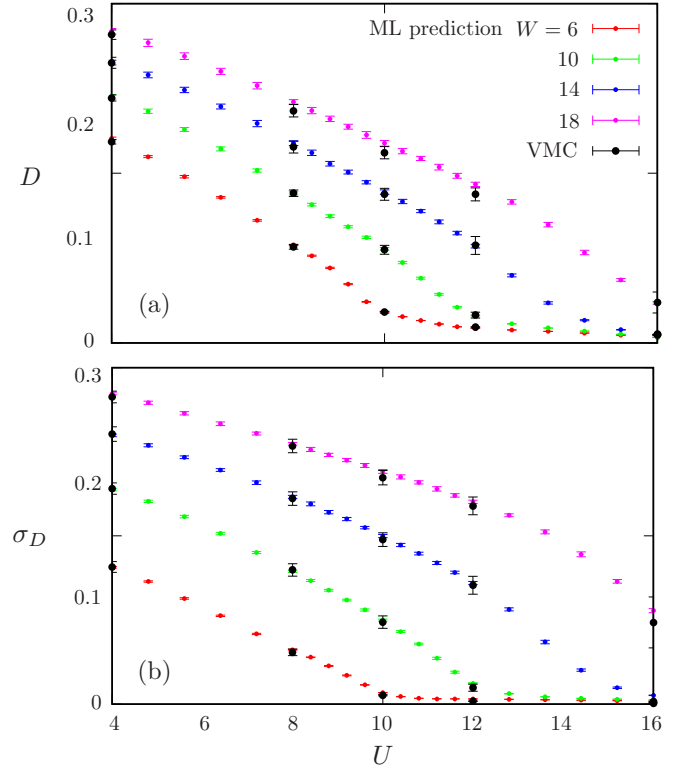
The results shown in Fig. 5 are also consistent with the picture of disorder screening discussed previously [41,50,54]. At small U, the strong disorder results in a highly inhomogeneous distribution of electron double occupation, as demonstrated in Fig. 5(a). With increasing U, in addition to a reduced double occupancy due to Coulomb repulsion, the spatial inhomogeneity of D_i is also reduced due to the renormalization Σ_i of the on-site potential by electron correlation. By applying the ML model to independent realizations of disorder, Fig. 6(a) shows the double-occupancy D, averaged over all lattice sites and different disorder configurations, as a function of U for different disorder strength W. Also shown for comparison are the VMC results on a smaller 16×16 system. The ML predictions not only are consistent with the VMC calculations, but also exhibit a consistent trend towards Mott transition with increasing Hubbard parameter U. In the presence of strong disorder, our results show a continuous crossover from Anderson insulator to the Mott insulating phase [55,62].

As shown in Fig. 5, the rather inhomogeneous states indicate a rather wide distribution of the local double occupancy, especially at small U. In addition to the mean value, our ML model also captures this spatial inhomogeneity of the electron state. To demonstrate this, we plot in Fig. 6(b) the spatial variation of the double occupancy $\sigma_D = [\frac{1}{N}\sum_i (D_i - \overline{D})^2]^{1/2}$,

averaged over several independent disorder realizations, as a function of U. The amount of variation computed from ML predicted local double occupancy agrees very well with the VMC simulations. The ML model also consistently predicts the decrease of the dispersion σ_D as U is increased, which is indicative of the screening of disorder induced by strong electron correlation [41,50,54].

Finally, it is worth pointing out that the nontrivial U dependence, which encapsulates the electron correlation effect, can be incorporated in the ML model by simply adding the Hubbard parameter U at the input node; see Fig. 2. The feasibility of this approach can be partly attributed to the smooth dependence of electronic properties on the Hubbard U in the presence of disorder. Indeed, as discussed in previous works, the first-order Mott transition is smeared by the strong disorder in 2D, leading to a continuous crossover to the Anderson-Mott insulator [55,62]. Moreover, even through a single input node, highly nonlinear dependence on U can be achieved through the fully connected neurons with nonlinear activations.

V. SUMMARY AND OUTLOOK

To summarize, we present a comprehensive ML framework for the predictions of local electronic properties of disordered Hubbard models. By exploiting the universal approximation capability of neural networks, a ML model is developed to encode the complex dependence of local quantities, such as electron number and double occupancy, on the local environment. Based on the group-theoretical method, a descriptor is proposed to represent the neighborhood random potentials with the lattice symmetry properly taken into account. We use the AH model as an example to demonstrate our ML framework. By training the NN with data sets from small-scale VMC simulations, we show that consistent results are obtained by applying the ML model to large systems with approximately 10^4 lattice sites.

Our approach emphasizes the scalability and transferability of the ML model, which are essential in order to achieve the goal of multiscale modeling of correlated electron systems. The fact that most electronic structure methods and many-body techniques for solving strongly correlated models have a polynomial complexity $\mathcal{O}(N^\alpha)$ with $\alpha > 1$ significantly restricts the accessible size and time scales. On the other hand, as pointed out by Kohn [28,29], the locality nature of most electronic systems, namely, local physical observables are determined by the immediate environment, underpins the possibility of linear-scaling electronic structure methods. The ML approach proposed in this work takes advantage of this feature to enable linear-scaling calculations for local electronic properties of correlated electron systems.

The assumption of locality also means that long-ranged interactions such as Coulomb or dipolar interactions are beyond the proposed ML approach. Nonetheless, such slowly decaying long-range interactions are often of classical origin and can be described by explicit analytical formulas. Their effect can thus be included independent of the ML model, which is designed to model the effects of electrons. Moreover, efficient methods such as Ewald summation have been developed to deal with such long-range interactions (Coulomb or dipolar). However, the presence of such long-range interactions does increase the overall computational complexity and, as a result, exact linear scaling might not be possible.

It should also be noted that, despite the generality of the nearsightedness of many-electron systems [28,29], the effective range of influence could be rather large, especially for electronic systems close to criticality. The prediction accuracy of the proposed ML approach then depends crucially on the cutoff radius of the local environment. One thus needs to strike a balance between accuracy and feasibility in practical implementations. Nonetheless, the ML model is developed with the main purpose of modeling spatially inhomogeneous electronic states, of which the electron correlation decays exponentially with a short correlation length. This is because even the electron wave functions are short ranged due to, e.g., Anderson localization effect, in such disordered systems. We thus expect the ML model should perform well for highly inhomogeneous electron systems.

As discussed in Sec. I, similar approaches have been employed to develop ML force-field models for quantum MD simulations. However, in most prior works [10–22], the ML models are derived from electronic structure methods that are based on self-consistent single-particle calculation, such as DFT or Gutzwiller approximation. Our work demonstrates that NN can successfully learn variational Monte Carlo simulation, which is a many-body method beyond effective single-particle or mean-field type approximations.

Finally we note that a similar ML framework has also been applied to enable large-scale dynamical simulations of some correlated electron systems, including the double-exchange and Falicov-Kimball models [24–26]. These electronic models are characterized by dynamical classical degrees of freedom coupled to free electrons described by a tight-binding Hamiltonian. ML models are constructed for the generalized force fields acting on the dynamical classical fields. As there are no direct electron-electron interactions in these models, the electronic structure problems can be solved by exact diagonalization. For correlated electron systems with Hubbard-type interactions, sophisticated many-body methods are required for an accurate solution of the electronic structure problems. As mentioned above, our ML model can be readily combined with Langevin method to enable large-scale dynamical simulations of the Hubbard-Holstein model. Our work paves the way toward multiscale dynamical modeling of strongly correlated systems such as Hubbard or t-J models.

[1] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. **91**, 045002 (2019).

[2] 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 (London) **560**, 41 (2018).

[3] S. Das Sarma, D.-L. Deng, and L.-M. Duan, Machine learning meets quantum physics, Phys. Today **72**, 48 (2019).

[4] E. Bedolla, L. C. Padierna, and R. Castaneda-Priego, Machine learning for condensed matter physics, J. Phys.: Condens. Matter **33**, 053001 (2021).

[5] J. Schmidhuber, Deep learning in neural networks: An overview, Neural Networks **61**, 85 (2015).

[6] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature (London) **521**, 436 (2015).

[7] G. Cybenko, Approximations by superpositions of sigmoidal functions, Math. Control Signals Syst. **2**, 303 (1989).

[8] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, Neural Networks **2**, 359 (1989).

[9] A. Barron, Universal approximation bounds for superpositions of a sigmoidal function, IEEE Trans. Inf. Theory **39**, 930 (1993).

[10] J. Behler and M. Parrinello, Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces, Phys. Rev. Lett. **98**, 146401 (2007).

[11] 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. **104**, 136403 (2010).

[12] Z. Li, J. R. Kermode, and A. De Vita, Molecular Dynamics with On-the-Fly Machine Learning of Quantum-Mechanical Forces, Phys. Rev. Lett. **114**, 096405 (2015).

[13] V. Botu, R. Batra, J. Chapman, and R. Ramprasad, Machine learning force fields: Construction, validation, and outlook, J. Phys. Chem. C **121**, 511 (2017).

[14] 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, J. Chem. Theory Comput. **13**, 4492 (2017).

[15] 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. **8**, 3192 (2017).

[16] 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. **120**, 143001 (2018).

[17] J. Behler, Perspective: Machine learning potentials for atomistic simulations, J. Chem. Phys. **145**, 170901 (2016).

[18] V. L. Deringer, M. A. Caro, and G. Csányi, Machine learning interatomic potentials as emerging tools for materials science, Adv. Mater. **31**, 1902765 (2019).

[19] F. Noé, A. Tkatchenko, K.-R. Müller, and C. Clementi, Machine learning for molecular simulation, Annu. Rev. Phys. Chem. **71**, 361 (2020).

[20] 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 Moller-Plesset perturbation theory with neural networks, J. Chem. Phys. **147**, 161725 (2017).

[21] T. Mueller, A. Hernandez, and C. Wang, Machine learning for interatomic potential models, J. Chem. Phys. **152**, 050902 (2020).

[22] 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 **99**, 161107(R) (2019).

[23] D. Marx and J. Hutter, *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods* (Cambridge University Press, Cambridge, UK, 2009).

[24] P. Zhang, P. Saha, and G.-W. Chern, Machine learning dynamics of phase separation in correlated electron magnets, arXiv:2006.04205.

[25] P. Zhang and G.-W. Chern, Arrested Phase Separation in Double-Exchange Models: Large-Scale Simulation Enabled by Machine Learning, Phys. Rev. Lett. **127**, 146401 (2021).

[26] 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, Proc. Natl. Acad. Sci. USA **119**, e2119957119 (2022).

[27] J. Ma, P. Zhang, Y. Tan, A. W. Ghosh, and G.-W. Chern, Machine learning electron correlation in a disordered medium, Phys. Rev. B **99**, 085118 (2019).

[28] W. Kohn, Density Functional and Density Matrix Method Scaling Linearly with the Number of Atoms, Phys. Rev. Lett. **76**, 3168 (1996).

[29] E. Prodan and W. Kohn, Nearsightedness of electronic matter, Proc. Natl. Acad. Sci. USA **102**, 11635 (2005).

[30] A. Weiße, G. Wellein, A. Alvermann, and H. Fehske, The kernel polynomial method, Rev. Mod. Phys. **78**, 275 (2006).

[31] Z. Wang, G.-W. Chern, C. D. Batista, and K. Barros, Gradient-based stochastic estimation of the density matrix, J. Chem. Phys. **148**, 094107 (2018).

[32] M. Ulmke, V. Janis, and D. Vollhardt, Anderson-Hubbard model in infinite dimensions, Phys. Rev. B **51**, 10411 (1995).

[33] V. Dobrosavljević and G. Kotliar, Mean Field Theory of the Mott-Anderson Transition, Phys. Rev. Lett. **78**, 3943 (1997).

[34] V. Dobrosavljević, D. Tanasković, and A. A. Pastor, Glassy Behavior of Electrons Near Metal-Insulator Transitions, Phys. Rev. Lett. **90**, 016402 (2003).

[35] T. Holstein, Studies of polaron motion: Part I. The molecular-crystal model, Ann. Phys. **8**, 325 (1959).

[36] J. Zhong and H.-B. Schüttler, Polaronic Anharmonicity in the Holstein-Hubbard Model, Phys. Rev. Lett. **69**, 1600 (1992).

[37] M. Weber and M. Hohenadler, Two-dimensional Holstein-Hubbard model: Critical temperature, Ising universality, and bipolaron liquid, Phys. Rev. B **98**, 085405 (2018).

[38] K. Byczuk, W. Hofstetter, and D. Vollhardt, Mott-Hubbard Transition versus Anderson Localization in Correlated Electron Systems with Disorder, Phys. Rev. Lett. **94**, 056404 (2005).

[39] M. C. O. Aguiar, V. Dobrosavljević, E. Abrahams, and G. Kotliar, Critical Behavior at the Mott-Anderson Transition: A Typical-Medium Theory Perspective, Phys. Rev. Lett. **102**, 156402 (2009).

[40] D. Heidarian and N. Trivedi, Inhomogeneous Metallic Phase in a Disordered Mott Insulator in Two Dimensions, Phys. Rev. Lett. **93**, 126401 (2004).

[41] E. C. Andrade, E. Miranda, and V. Dobrosavljević, Electronic Griffiths Phase of the $d = 2$ Mott Transition, Phys. Rev. Lett. **102**, 206403 (2009).

[42] E. Lahoud, O. N. Meetei, K. B. Chaska, A. Kanigel, and N. Trivedi, Emergence of a Novel Pseudogap Metallic State in a Disordered 2D Mott Insulator, Phys. Rev. Lett. **112**, 206402 (2014).

[43] N. D. Patel, A. Mukherjee, N. Kaushal, A. Moreo, and E. Dagotto, Non-Fermi Liquid Behavior and Continuously Tunable Resistivity Exponents in the Anderson-Hubbard Model at Finite Temperature, Phys. Rev. Lett. **119**, 086601 (2017).

[44] J. C. Szabo, K. Lee, V. Madhavan, and N. Trivedi, Local Spectroscopies Reveal Percolative Metal in Disordered Mott Insulators, Phys. Rev. Lett. **124**, 137402 (2020).

[45] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions, Rev. Mod. Phys. **68**, 13 (1996).

[46] V. Dobrosavljević, A. A. Pastor, and B. K. Nikolić, Typical medium theory of Anderson localization: A local order parameter approach to strong-disorder effects, Europhys. Lett. **62**, 76 (2003).

[47] M. A. Tusch and D. E. Logan, Interplay between disorder and electron interactions in a $d = 3$ site-disordered Anderson-Hubbard model: A numerical mean-field study, Phys. Rev. B **48**, 14843 (1993).

[48] H. Shinaoka and M. Imada, Single-particle excitations under coexisting electron correlation and disorder: A numerical study of the Anderson-Hubbard model, J. Phys. Soc. Jpn. **78**, 094708 (2009).

[49] F. Fazileh, R. J. Gooding, W. A. Atkinson, and D. C. Johnston, Role of Strong Electronic Correlations in the Metal-To-Insulator Transition in Disordered $LiAl_yTi_{2-y}O_4$, Phys. Rev. Lett. **96**, 046410 (2006).

[50] E. C. Andrade, E. Miranda, and V. Dobrosavljević, Quantum Ripples in Strongly Correlated Metals, Phys. Rev. Lett. **104**, 236401 (2010).

[51] R. Kotlyar and S. Das Sarma, Disorder and Interaction in 2D: Exact Diagonalization Study of the Anderson-Hubbard-Mott Model, Phys. Rev. Lett. **86**, 2388 (2001).

[52] B. Srinivasan, G. Benenti, and D. L. Shepelyansky, Delocalizing effect of the Hubbard repulsion for electrons on a two-dimensional disordered lattice, Phys. Rev. B **67**, 205112 (2003).

[53] S. Chiesa, P. B. Chakraborty, W. E. Pickett, and R. T. Scalettar, Disorder-Induced Stabilization of the Pseudogap in Strongly Correlated Systems, Phys. Rev. Lett. **101**, 086401 (2008).

[54] D. Tanasković, V. Dobrosavljević, E. Abrahams, and G. Kotliar, Disorder Screening in Strongly Correlated Systems, Phys. Rev. Lett. **91**, 066603 (2003).

[55] M. Y. Suárez-Villagrán, N. Mitsakos, T.-H. Lee, V. Dobrosavljević, J. H. Miller, Jr., and E. Miranda, Two-dimensional disordered Mott metal-insulator transition, Phys. Rev. B **101**, 235112 (2020).

[56] D. Semmler, K. Byczuk, and W. Hofstetter, Anderson-Hubbard model with box disorder: Statistical dynamical mean-field theory investigation, Phys. Rev. B **84**, 115113 (2011).

[57] A. W. Sandvik and D. J. Scalapino, Effects of weak random disorder in the one-dimensional Hubbard model, Phys. Rev. B **47**, 10090 (1993).

[58] A. W. Sandvik, D. J. Scalapino, and P. Henelius, Quantum Monte Carlo study of the one-dimensional Hubbard model with random hopping and random potentials, Phys. Rev. B **50**, 10474 (1994).

[59] M. Ulmke and R. T. Scalettar, Magnetic correlations in the two-dimensional Anderson-Hubbard model, Phys. Rev. B **55**, 4149 (1997).

[60] Y. Otsuka, Y. Morita, and Y. Hatsugai, Collapse of the charge gap in random Mott insulators, Phys. Rev. B **58**, 15314 (1998).

[61] M. E. Pezzoli, F. Becca, M. Fabrizio, and G. Santoro, Local moments and magnetic order in the two-dimensional Anderson-Mott transition, Phys. Rev. B **79**, 033111 (2009).

[62] M. E. Pezzoli and F. Becca, Ground-state properties of the disordered Hubbard model in two dimensions, Phys. Rev. B **81**, 075106 (2010).

[63] A. Farhoodfar, X. Chen, R. J. Gooding, and W. A. Atkinson, Real-space variational Gutzwiller wave functions for the Anderson-Hubbard model, Phys. Rev. B **80**, 045108 (2009).

[64] A. Farhoodfar, R. J. Gooding, and W. A. Atkinson, Variational Monte Carlo study of Anderson localization in the Hubbard model, Phys. Rev. B **84**, 205125 (2011).

[65] M. Capello, F. Becca, M. Fabrizio, S. Sorella, and E. Tosatti, Variational Description of Mott Insulators, Phys. Rev. Lett. **94**, 026406 (2005).

[66] M. C. Gutzwiller, Effect of Correlation on the Ferromagnetism of Transition Metals, Phys. Rev. Lett. **10**, 159 (1963).

[67] M. C. Gutzwiller, Correlation of electrons in a narrow s band, Phys. Rev. **137**, A1726 (1965).

[68] S. Sorella, Wave function optimization in the variational Monte Carlo method, Phys. Rev. B **71**, 241103(R) (2005).

[69] F. Becca and S. Sorella, *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press, Cambridge, UK, 2017).

[70] S. Zhang, P. Zhang, and G.-W. Chern, Machine learning model of generalized force field in condensed matter systems: Importance of descriptors, arXiv:2201.00798.

[71] 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. **114**, 105503 (2015).

[72] A. P. Bartók, R. Kondor, and G. Csányi, On representing chemical environments, Phys. Rev. B **87**, 184115 (2013).

[73] L. Himanen, M. O. J. Jäger, E. V. Morooka, F. F. Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke, and A. S. Foster, DScribe: Library of descriptors for machine learning in materials science, Comput. Phys. Commun. **247**, 106949 (2020).

[74] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning, Phys. Rev. Lett. **108**, 058301 (2012).

[75] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, Bond-orientational order in liquids and glasses, Phys. Rev. B **28**, 784 (1983).

[76] J. Behler, Atom-centered symmetry functions for constructing high-dimensional neural network potentials, J. Chem. Phys. **134**, 074106 (2011).

[77] A. V. Shapeev, Moment tensor potentials: A class of systematically improvable interatomic potentials, Multiscale Model. Simul. **14**, 1153 (2016).

[78] R. Drautz, Atomic cluster expansion for accurate and transferrable interatomic potentials, Phys. Rev. B **99**, 014104 (2019).

[79] 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, J. Phys. Chem. Lett. **6**, 2326 (2015).

[80] F. Faber, A. Lindmaa, O. A.von Lilienfeld, and R. Armiento, Crystal structure representations for machine learning models of formation energies, Int. J. Quantum Chem. **115**, 1094 (2015).

[81] H. Huo and M. Rupp, Unified representation of molecules and crystals for machine learning, arXiv:1704.06439.

[82] R. Kondor, A novel set of rotationally and translationally invariant features for images based on the non-commutative bispectrum, arXiv:cs/0701127.

[83] R. Caruana, Multitask learning, Mach. Learn. **28**, 41 (1997).

[84] R. Collobert and J. Weston, A unified architecture for natural language processing: Deep neural networks with multitask learning, in *ICML '08 Proceedings of the 25th International Conference on Machine Learning* (ACM, New York, NY, 2008), pp. 160–167.