

HubbardNet: Efficient Predictions of the Bose-Hubbard Model Spectrum with Deep Neural Networks

Ziyan Zhu,^{1,*} Marios Mattheakis,² Weiwei Pan,² and Efthimios Kaxiras^{3,2}

¹*Stanford Institute of Materials and Energy Science,
SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*

²*John A. Paulson School of Engineering and Applied Sciences,
Harvard University, Cambridge, Massachusetts 02138, USA*

³*Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*

We present a deep neural network (DNN)-based model (**HubbardNet**) to variationally find the ground state and excited state wavefunctions of the one-dimensional and two-dimensional Bose-Hubbard model. Using this model for a square lattice with M sites, we obtain the energy spectrum as an analytical function of the on-site Coulomb repulsion, U , and the total number of particles, N , from a single training. This approach bypasses the need to solve a new hamiltonian for each different set of values (U, N). We show that the DNN-parametrized solutions are in excellent agreement with results from the exact diagonalization of the hamiltonian, and it outperforms the exact diagonalization solution in terms of computational scaling. These advantages suggest that our model is promising for efficient and accurate computation of exact phase diagrams of many-body lattice hamiltonians.

I. INTRODUCTION

In condensed matter physics, simple many-body lattice models have been employed to capture the essential physics of intriguing phenomena like phase transitions, strongly correlated states, and exotic phases. A particularly popular and powerful simple model is the Hubbard model, which has proven useful in elucidating, for instance, the phase diagram and the spectra of high-temperature superconductors [1–3], of unconventional superconductivity in twisted multilayered van der Waals heterostructures [4–11], and the superfluid-to-insulator transition in bosonic gases [12–14]. However, the solution of the deceptively simple Hubbard model remains a challenge due to the fact that, as a function of the number of particles in the system, an exponentially large space is required to fully describe the many-body wavefunction.

In this work, we adopt a variational approach by considering a class of wavefunctions parameterized by a deep neural network (DNN) to reduce the parameter space to search for the optimal solution. Similar machine learning (ML)-based variational approaches have proven useful in providing insights into the ground state of interacting spin system [15–17], bosonic systems [18–20], fermionic systems [21–24], dynamics in many-body systems [25], and open quantum systems [26, 27]. While the amount of information encoded in these ML-based models in principle scales linearly as a function of the number of particles and sites, the actual training of a DNN is known to be computationally costly and in many cases more costly than the exact diagonalization (ED) of the hamiltonian, because it requires a new training for every different value of the parameters on which the hamiltonian depends.

The very large size of the space needed to represent the wavefunction has so far restricted the search for solutions,

typically, only to the ground state of the many-body system. Our proposed model (in the following referred to as the **HubbardNet**) employs a variational approach to solve for the ground state as well as for the excited states of the one-dimensional (1D) and two-dimensional (2D) Bose-Hubbard model. Our DNN-parametrized solutions are an analytic function of the on-site Coulomb repulsion between electrons, described by the parameter U , and the total number of particles N . This bypasses the need to perform an optimization for every value of the (U, N) parameters of the hamiltonian. Our results provide the proof of principle that variational approach employing DNN's can be used to efficiently map out the phase boundaries in (U, N) space of low-dimensional lattice models.

The paper is organized as follows. In Section II we introduce the Bose-Hubbard model and its construction. In Section III we present the details of **HubbardNet** and use it to solve for the ground state and excited states of the 1D and 2D Bose-Hubbard model on a square lattice with M sites, using both periodic and open boundary conditions. Finally, we summarize our findings in Section V and discuss the possible extensions to fermionic systems.

II. THE BOSE-HUBBARD MODEL

The Bose-Hubbard Hamiltonian is given by

$$\hat{\mathcal{H}} = -t \sum_{\langle ij \rangle} \hat{a}_i \hat{a}_j^\dagger + \frac{U}{2} \sum_{i=1}^M \hat{n}_i (\hat{n}_i - 1) \quad (1)$$

where the sum on i runs over the M lattice sites and $\langle ij \rangle$ denotes the nearest neighbor pairs in the lattice. The first term in the hamiltonian represents the kinetic energy of the quantum particles, with $\hat{a}_i^\dagger, \hat{a}_i$ being the creation and annihilation operators for a particle at the lattice site i ,

* ziyanzhu@stanford.edu

and $-t$ being the amount of energy gained by a particle during a “hop” between two neighboring sites $\langle ij \rangle$. The next term in the hamiltonian represents the interaction between particles, assumed to be charged (electrons or ions). We assume the system consists of only one kind of charged particles, with U describing the Coulomb repulsion between each pair; $\hat{n} = \hat{a}_i^\dagger \hat{a}_i$ is the number operator, whose expectation value gives the occupation of the lattice site i . In the following we take $t = 1$ and express all values of energy in units of t . This simple model can capture the superfluid phase for $U/t \gg 1$ and the Mott insulator phase for $U/t \leq 1$ [11].

We will denote a state, referred to as a Fock state, of the system as $|\mathbf{n}_j\rangle$, which is defined by the occupation of each site $n_i^{(j)}$, with the index i running over all M sites, that is,

$$|\mathbf{n}_j\rangle = |n_1^{(j)}, n_2^{(j)}, \dots, n_M^{(j)}\rangle, \quad (2)$$

with $n_i^{(j)} \in [0, N]$ for $i = 1, \dots, M$, which satisfies

$$\sum_{i=1}^M n_i^{(j)} = N. \quad (3)$$

The number \mathcal{N}_B of these Fock states for the Bose-Hubbard model with N particles in M sites is given by

$$\mathcal{N}_B = \binom{M+N-1}{N}. \quad (4)$$

Using this basis of Fock states, the many-body wavefunction $|\Psi_k\rangle$ that describes the entire system is written as:

$$|\Psi_k\rangle = \sum_{j=1}^{\mathcal{N}_B} \psi_j^{(k)} |\mathbf{n}_j\rangle, \quad (5)$$

and the solution involves finding all the many-body wavefunctions that satisfy the eigenvalue equation

$$\hat{\mathcal{H}}|\Psi_k\rangle = E_k|\Psi_k\rangle, \quad (6)$$

with E_k being the k -th energy eigenvalues. The states $|\Psi_k\rangle$ form an orthonormal set, $\langle\Psi_j|\Psi_j\rangle = \delta_{ij}$, which implies

$$\sum_{j=1}^{\mathcal{N}_B} |\psi_j^{(k)}|^2 = 1 \quad (7)$$

for each value of k . Thus, the solution involves finding the complex numbers $\psi_j^{(k)}$, $j = 1, \dots, \mathcal{N}_B$, that describe each of the many-body states.

The solution is found by constructing the hamiltonian matrix

$$\mathcal{H}_{ij} = \langle \mathbf{n}_i | \hat{\mathcal{H}} | \mathbf{n}_j \rangle, \quad i, j = 1, \dots, \mathcal{N}_B, \quad (8)$$

and finding all its eigenvalues and eigenvectors. To this

end, every Fock state needs to be associated with an integer label, in a unique and known ordering. This can be done iteratively by Ponomarev ordering [28, 29], a procedure described in detail by Reventós *et al.* [30]. We note that the kinetic-energy term produces contributions only to the off-diagonal hamiltonian matrix elements, because it involves hops of particles between nearest-neighbor sites, which changes the occupations of these sites. For example, for a model with $M = 5, N = 3$, supposing the i^{th} state is $\mathbf{n}_i = [2, 1, 0, 0, 0]$ and the j^{th} state is $\mathbf{n}_j = [1, 2, 0, 0, 0]$ by Ponomarev ordering, $\mathcal{H}_{ij} = -t$. This is because a particle at site 1 in state \mathbf{n}_i can hop to site 2, which results in state \mathbf{n}_j . This nearest neighbor hop contributes a term of $-t$ to the energy. By similar arguments, the Coulomb repulsion term produces contributions to the diagonal matrix elements only, because it does not involve any changes in the site occupations.

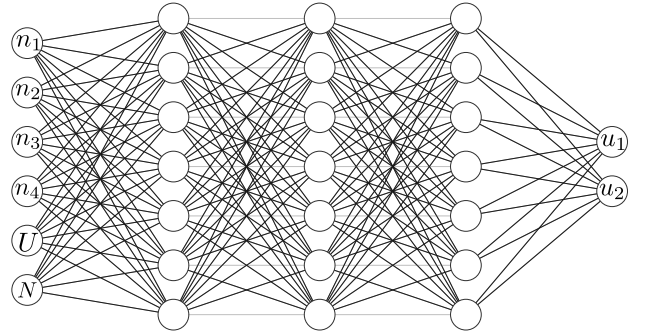


FIG. 1: Schematic diagram of the data-free neural network to solve for the Bose-Hubbard model. The network takes $(M+2)$ inputs, which are the number of particles n_i at all $i = 1, \dots, M$ sites as well as U , the Coulomb repulsion parameter, and N , the total number of particles in the system. The output layer is two-dimensional, with the outputs u_1 and u_2 corresponding to the real and imaginary part of the wavefunction component corresponding to this particular configuration of particles.

III. NEURAL NETWORK-PARAMETRIZED SOLUTION

We parametrize the many-body wavefunction $|\Psi_k\rangle$ by a fully connected DNN

$$|\Psi_k\rangle = \sum_{j=1}^{\mathcal{N}_B} g(\mathbf{n}_j, U, N; \mathbf{W}^{(k)}) |\mathbf{n}_j\rangle \quad (9)$$

where $\mathbf{W}^{(k)}$ denotes the network parameters for this state; a schematic representation of the network is shown in Fig. 1. Thus, the network takes as input the pair of values (U, N) and the set of occupation numbers $n_i^{(j)}$, $i = 1, \dots, M$, that correspond to a specific Fock state $|\mathbf{n}_j\rangle$, and gives as output the coefficient $\psi_j^{(k)}$ that corresponds to this Fock state, which is a component of

the many-body wavefunction $|\Psi_k\rangle$ with energy E_k . There is apparently a redundancy in these definitions, because of the constraint given in Eq. (3), so not all of the input values to the DNN are independent. However, a better description is that for a given value of N , only the Fock states within the space defined by Eqs. (2) and (3) are allowed as inputs to the network. In this sense, N , which is an important variable in the problem, can be considered as an independent variable. To obtain the full wavefunction $|\Psi_k\rangle$, the trained network needs to be applied to each of the Fock states once.

The size of the hamiltonian matrix, $\mathcal{N}_B \times \mathcal{N}_B$, increases as the factorials of the number of particles N and sites M . The exact diagonalization of the hamiltonian matrix quickly becomes intractable as the system size increases. In contrast, the number of input parameters scales linearly with M in the DNN approach [15]. Another important advantage of the DNN-parameterized solution is that with U and N being the input parameters to the network, the network becomes an analytic function of U and N and we essentially solve for a class of hamiltonian during a single training. The network can then be used to make predictions at different values of U that are not in the training set. In contrast, other techniques such as ED require a new calculation for every different set of values (U, N) .

In this work, we use 4 hidden layers and use a hyperbolic tangent, $\tanh()$, activation function for all hidden layers. In the examples that we considered in this work, the wavefunctions are real. However, the wavefunction can in general be complex, and thus we need two real-number outputs for the real and imaginary parts of the coefficients

$$\psi_j^{(k)} = u_{1,j}^{(k)} + iu_{2,j}^{(k)}. \quad (10)$$

For the ground state wavefunction, $|\Psi_0\rangle$ (denoted by the index 0), we choose the activation function of the output layer to be an exponential, namely

$$\psi_j^{(0)} = \exp\left(u_{1,j}^{(0)} + iu_{2,j}^{(0)}\right). \quad (11)$$

This choice is made for better convergence with uniformly sampled initial weights. Note that with the exponential function, the wavefunction $\psi_j(\mathbf{n})$ is positive. The coefficients of the Fock states have the same sign for the ground state, which is guaranteed by the structure of the hamiltonian for $-t < 0$ on all the off-diagonal matrix elements. For excited states, we take the linear activation function for the output layer, given in Eq. (10), because the exponential activation function would no longer work, as the wavefunction needs to be orthogonal to the ground state wavefunction, and the coefficients $\psi_j^{(k)}$ must have mixed signs. Note that for the 1D and 2D square lattices, the wavefunctions are purely real and the output $u_{2,j}^{(k)}$ is not necessary.

The average of any operator $\hat{\mathcal{A}}$ that applies to Fock

states $|\mathbf{n}_j\rangle$ can be obtained through the expression

$$\langle \hat{\mathcal{A}} \rangle_k \equiv \langle \Psi_k | \hat{\mathcal{A}} | \Psi_k \rangle = \sum_{i,j=1}^{\mathcal{N}_B} \psi_i^{(k)*} \langle \mathbf{n}_i | \hat{\mathcal{A}} | \mathbf{n}_j \rangle \psi_j^{(k)}. \quad (12)$$

Using $\hat{\mathcal{H}}$ as the operator in Eq. (12) we can obtain the energy E_k of the eigenstate $|\Psi_k\rangle$ of the hamiltonian. In the network representation of the many-body wavefunction, Eq. (9), the energy takes the form:

$$E_k(U, N) = \sum_{i,j=1}^{\mathcal{N}_B} g^*(\mathbf{n}_i, U, N; \mathbf{W}^{(k)}) \mathcal{H}_{ij} g(\mathbf{n}_j, U, N; \mathbf{W}^{(k)}) \left[\sum_{i=1}^{\mathcal{N}_B} |g(\mathbf{n}_i, U, N; \mathbf{W}^{(k)})|^2 \right]^{-1}, \quad (13)$$

where \mathcal{H}_{ij} are the hamiltonian matrix elements defined in Eq. (8). In the above equation we have indicated explicitly the dependence of the energy eigenvalues on U and N . Accordingly, we choose the loss function as follows:

$$\mathcal{L}_k = \frac{1}{\mathcal{N}_U \mathcal{N}_N} \sum_{l=1}^{\mathcal{N}_U} \sum_{m=1}^{\mathcal{N}_N} E_k(U, N), \quad (14)$$

where $E_k(U, N)$ is the expression defined in Eq. (13), and $\mathcal{N}_U, \mathcal{N}_N$ are the numbers of values of U and N included in the training, respectively, that are included in the training of the network. However, the loss functions for the ground-state energy E_0 and for excited-state energies $E_k (k > 0)$, need to be treated differently, as described in detail next.

A. Ground state

To obtain the ground state energy and wavefunction, we minimize the loss as defined in Eq. (14) directly. As an example, we first discuss the solution for a 1D system with periodic boundary conditions. In this case, every site has 2 nearest neighbors. We train the network with three different U values, namely $U = 2.0, 5.0, 8.0$, for a system with $M = 6, N = 5$, which yields $\mathcal{N}_B = 252$ Fock states. For here and the rest of the paper, unless otherwise specified, we choose the network width to be $D = 400$. We use stochastic gradient descent (SGD) for the optimization of the network weights $\mathbf{W}^{(0)}$ that minimize the loss function. We use momentum for learning $= 0.9$. We adopt a Cosine Annealing scheme for the learning rate, with the maximum learning rate to be 0.01 and decaying to zero following a cosine function. We reset the learning rate every 1000 iterations. We train the network until the variance of the loss function during the last 200 steps is less than 1×10^{-8} (see Fig. 2(a)).

In Fig. 2(b), we compare the ground-state energy values as a function of U , as obtained by HubbardNet with the values obtained by diagonalizing of the hamiltonian

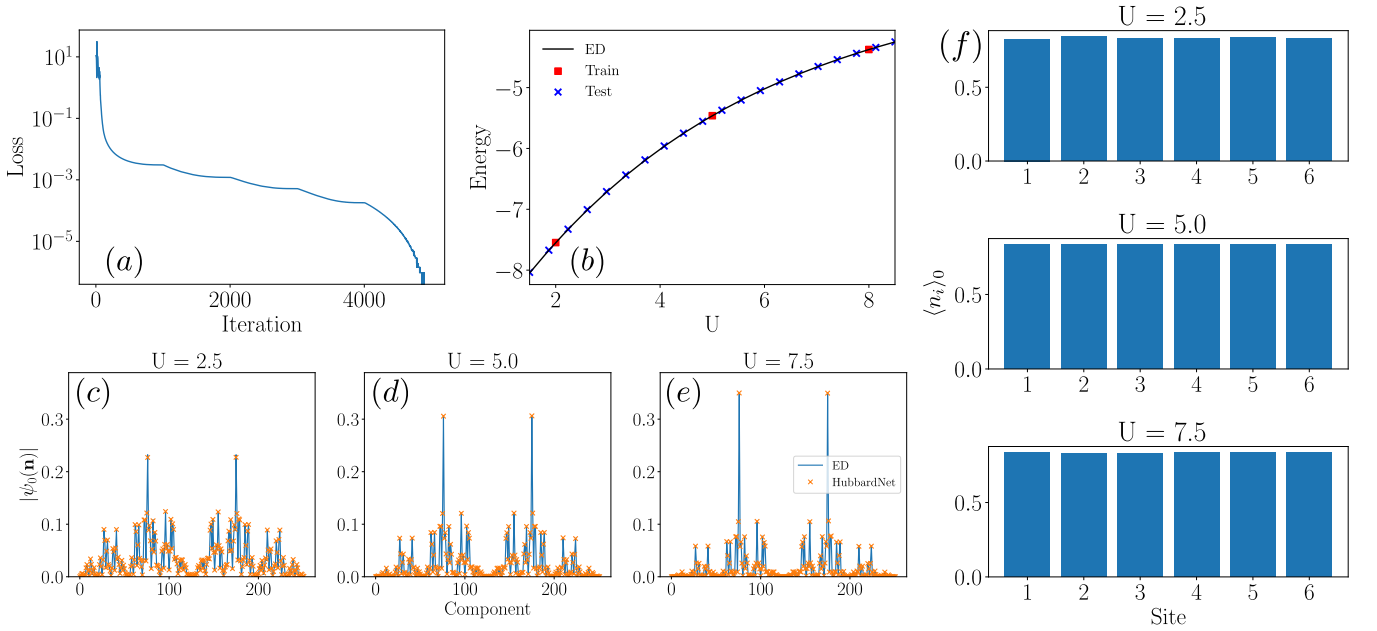


FIG. 2: One-dimensional Bose-Hubbard model solution with a periodic boundary condition and $M = 6, N = 5$ obtained using the **HubbardNet**. (a) Trace of the loss function offset by the negative of the minimum total energy. (b) Comparison between the ground truth ground state energies obtained from ED (blue dashed line) and the neural network results for training set (red crosses) and prediction (blue scattered points). (c)-(e) Absolute value of the wavefunction $|\psi_0(\mathbf{n})|$ from ED (blue dashed line) and neural network (orange scattered crosses) for (d) $U = 2.5$, (e) $U = 5.0$ (f) $U = 7.5$. (f) The occupation number $\langle n_i \rangle_0$ at different sites that correspond to (c)-(e).

matrix (namely ED). The comparison includes the values of U that are part of the training set (red scattered points), as well as values beyond the training set (blue scattered crosses). We note that even though the energy is a smooth function of U , the output of the network includes all the wavefunction components $\psi_j^{(0)}, j = 1, \dots, \mathcal{N}_B$, which are a more complex function of the parameter U . For example, in Fig. 2(c)-(e) we compare the wavefunction components (labeled with an arbitrary ordering to reflect the symmetry of the basis) from the neural network output, shown as a scatter plot, and the corresponding values obtained from ED, for three different values of U , only one of which was included in the training set, namely $U = 5.0$ (see Fig. 2(d)). In all three cases, there is an excellent agreement between the two results, indicating that with as little as $\mathcal{N}_U = 3$ points in the training sets, **HubbardNet** is capable of predicting the wavefunctions and energies of the 1D Bose-Hubbard model accurately for a wide range of U 's. In Fig. 2(f), we show the average occupation number $\langle \hat{n}_i \rangle_0$, obtained from Eq. (12), using \hat{n}_i as the operator, with the trained weights, for three different sets of U 's, both within and beyond the training set. As expected from the periodic boundary conditions, and the large value of U , the average occupation number is uniform for all values of U 's in this example. In other words, the large repulsion between the charged particles distributes them one per lattice site.

We next proceed to examine the capability of **HubbardNet** to handle more complicated cases, namely

lattices with a larger number of sites and particles. We consider a 2D Hubbard model with open boundary conditions (a cluster of lattice sites), with $M = 16$ (a 4×4 square lattice) and $N = 3$. These choices lead to a large number of Fock states, $\mathcal{N}_B = 816$. The hamiltonian construction is similar to the 1D case, but now every site has 4 nearest neighbors instead of 2, except for the sites on the boundary which have only 2 or 3 neighbors, from the open boundary condition. In this way, the hamiltonian becomes denser compared to the 1D case.

In Fig. 3 we show the solution, using again three values of U as the training set, namely $U = 2.0, 6.0, 9.0$. Despite the increased size of the problem, **HubbardNet** performs extremely well: both the energies, shown in Fig. 3(b), and the wavefunction components, shown in Fig. 3(d)-(f), match the results from the ED solution, even for the value of U that was not part of the training set (Fig. 3(d)-(f)). The occupation number distribution is no longer uniform and instead concentrated in the center of the cluster due to the open boundary condition.

In addition to including several values of U as part of the network input, the number of particles in the system, N , can also be part of the network input. Although the space of Fock states has different size for different values of N , **HubbardNet** can still be used as long as M is fixed, since the energy is an average over all particle configurations in the Fock basis. In Fig. 4, we show the results from using **HubbardNet** for a system with multiple values of both N and U as inputs, so that the output is a

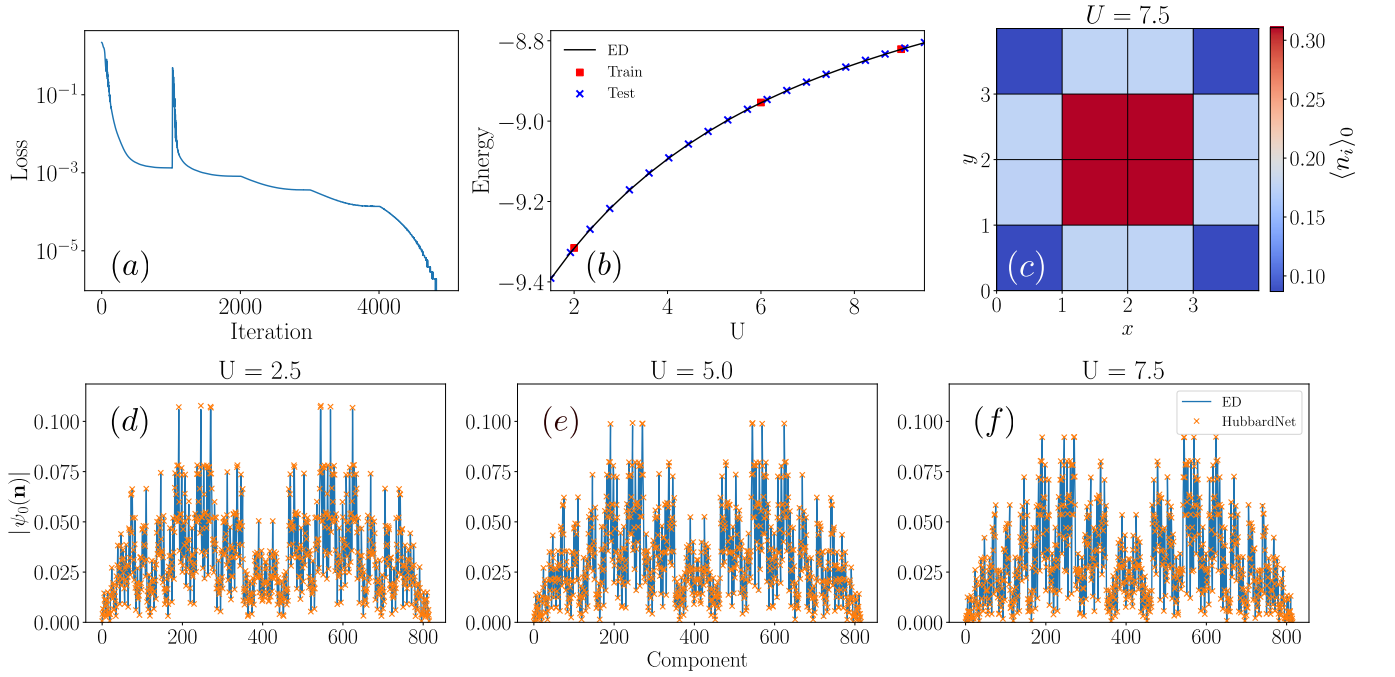


FIG. 3: Two-dimensional Bose-Hubbard model solution with an open boundary condition and $M = 16, N = 3$ with ($M_x = 4, M_y = 4$) obtained using the **HubbardNet**. (a) Trace of the loss function offset by the negative of the minimum total energy. (b) Comparison between the ground truth ground state energies obtained from exact diagonalization (blue dashed line) and the neural network results for the training set (red crosses) and prediction (blue scattered points). (c) The occupation number $\langle n_i \rangle$ at different sites for $U = 5.0$. (d)-(f) Absolute value of the wavefunction $|\psi_0(\mathbf{n})|$ from the exact diagonalization (blue dashed line) and neural network (orange scattered crosses) for (d) $U = 2.5$, (e) $U = 5.0$ (f) $U = 7.5$.

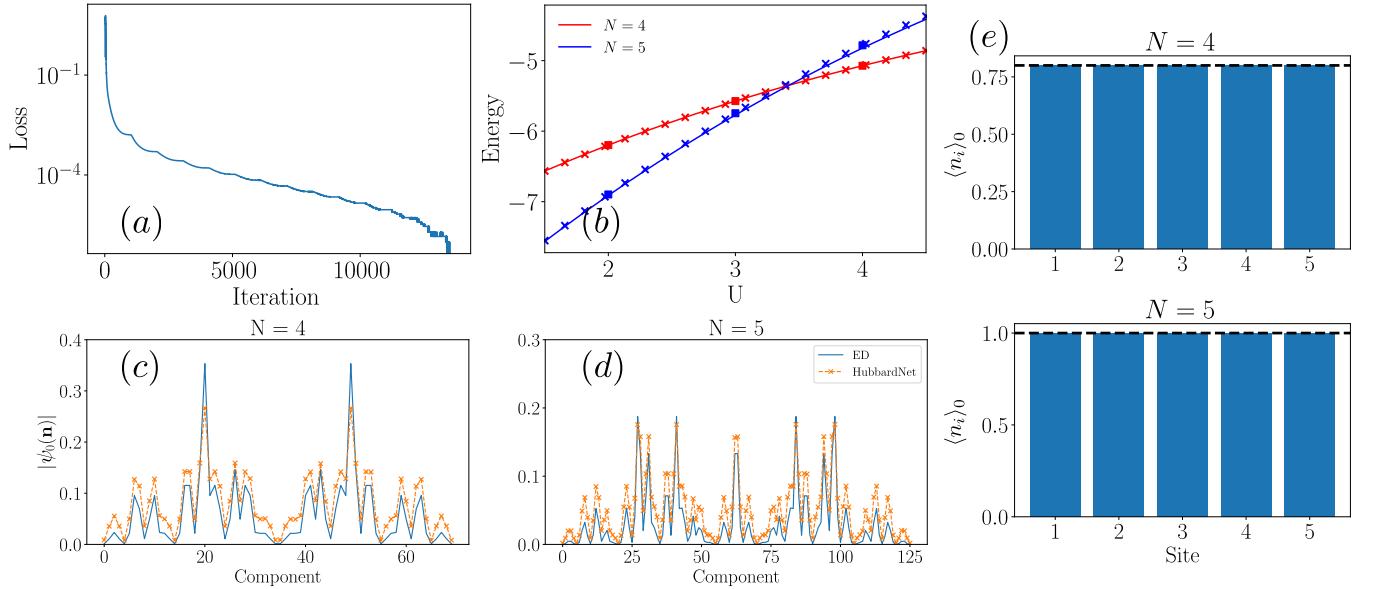


FIG. 4: One-dimensional Bose-Hubbard model ground state solution with a periodic boundary condition for $M = 5$, and the training set including both the $N = 5$ and $N = 4$ states. (a) Trace of the loss function offset by the negative of the minimum total energy. (b) Energy vs. U from ED (solid lines) and **HubbardNet** (scattered points) with $N = 4$ (red) and $N = 5$ (blue) respectively. Scattered points are from training and scattered crosses are from testing set. (c)-(d) Comparison between wavefunction components $|\psi_0(\mathbf{n})|$ obtained from ED (blue solid lines) and **HubbardNet** (orange dashed lines) for $N = 4$ (left) and $N = 5$ (right) with $U = 1.5$. (e) Occupation numbers that correspond to (c) and (d) (bottom). Black dashed lines on top represent a constant occupation number $\langle n_i \rangle = N/M = 0.8$ and bottom $\langle n_i \rangle = N/M = 1$.

function of both variables. Fig. 4(b) shows that the two curves describing the energy as a function of U for the

two different values of N cross each other, making the the optimization much more difficult. Remarkably, energies obtained from **HubbardNet** both at the values of U included in the training set and outside of the training set agree with the ED results (scattered points in Fig. 4), despite the crossover between the $N = 4$ and $N = 5$ energy vs. U curve. The **HubbardNet** wavefunction components, while retaining the approximate shape of the ED result, are slightly different from the latter (Fig. 4(c)-(d)). Despite the slight deviation of the wavefunction, the occupation numbers for both cases remain uniform (Fig. 4(e)) and $\langle n_i \rangle_0 = N/M \forall i \in [1, M]$, as expected for the Bose-Hubbard model ground state with a periodic boundary condition.

B. Excited states

In order to obtain excited states with the approach described above, we employ the fact that the eigenstates of the hamiltonian are orthogonal to each other, that is, $\langle \Psi_i | \Psi_j \rangle = \delta_{ij}$. An excited state is a local minimum of the energy. Therefore, we can obtain excited states by minimizing the expectation value in Eq. (13) for state $|\Psi_k\rangle$ with the constraint that it is orthogonal to $|\Psi_0\rangle, \dots, |\Psi_{k-1}\rangle$. The constraint is achieved iteratively through the Gram-Schmidt orthogonalization process. To facilitate the calculations, we express the wavefunctions $|\Psi_k\rangle$ as vectors \mathbf{v}_k of dimension \mathcal{N}_B , with entries $\psi_j^{(k)}$. In this notation, assume that $\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_{k-1}$ are the vectors that represent the output directly from the neural network, which are linearly independent, with $|\Psi_0\rangle = \mathbf{v}_0$ being the ground state wavefunction, and the rest being the wavefunctions of excited states. The excited state wavefunctions are obtained iteratively as follows:

$$|\Psi_k\rangle = \mathbf{v}_k - \sum_{j=1}^{k-1} \frac{\mathbf{v}_j \cdot \mathbf{v}_k}{\mathbf{v}_j \cdot \mathbf{v}_j} \mathbf{v}_j. \quad (15)$$

This expression for state \mathbf{v}_k incorporates the projection out of this state of components that correspond to all previous states, $j = 0, \dots, k-1$. The projection is performed at every iteration and the set of \mathbf{v}_j vectors is used to compute the loss function. In other words, for an excited state k , we train a new network with the outputs being constrained to be orthogonal to the space of all states with lower energy, from the ground state to the $(k-1)$ excited state. With the excited state optimization for multiple values of U as the input, we add the following penalty term to the loss function:

$$\mathcal{L}_P = -\exp(-|\bar{E}_k - \bar{E}_0|), \quad (16)$$

where the index k labels the excited state, \bar{E}_k is the average energy over the values of U as obtained at each epoch and \bar{E}_0 is the ground state energy. This choice of the loss function encourages the excited-state energy E_k

to be as close to the ground state energy as possible.

Similarly to the ground state, we can train the network as a function of U for a given excited state k . In Fig. 5(a) we show the energy spectra for 4 different values of U , namely $U = 2.0, 3.0, 4.0, 5.0$, with one training for a given k , for $M = 5, N = 3$ with an open boundary condition.

We apply an L2 regularization 10^{-4} to the network weights in order to discourage energies to be drastically different for similar values of U , an issue that we observed in the absence of such regularization. In addition, since our training set only contains a very small number of values of U but we would like to predict energies and wavefunctions as a smooth function of all values of U , we randomly perturb the value of the input U during every epoch, with a perturbation amplitude of 0.01. In Fig. 5(a) we show that for the U values in the training set, there is an excellent agreement between the results obtained with **HubbardNet** and those obtained with ED. Using the network, we can plot the full spectrum of the Bose-Hubbard model for the given M and N and an arbitrary value of U . In Fig. 5(b) we show the spectrum produced using the trained network at inference. The spectrum qualitatively agrees with the ED results, except for $U \lesssim 2$ and at high energies. This is expected because $U < 2$ corresponds to values beyond the range of training-set values, and higher-energy states are more difficult to obtain from the network due to error accumulation in the iterative orthogonalization scheme, Eq. (15).

We note that there is a gap in the energy spectrum at approximately $U = 2.0$, separating a few of the high-energy excited states from the rest. It is also possible that the network has more difficulty in making accurate predictions near the phase boundary, which by definition means that a small perturbation in the value of U leads to a large change in the qualitative behavior, and thus a denser sampling is more desirable near such values of U .

In Fig. 6, we show the corresponding wavefunctions for three different excited states. For the results presented in this figure, we used only one value of U in the training set, namely $U = 5.0$. For excited states with relatively low energy, for example, the 4th excited state, the predicted wavefunctions agree with the ground truth even for values of U not in the test set. However, for excited states with higher energy, such as the 15th or the 30th state, there is a significant difference between the **HubbardNet**-predicted wavefunctions and the ground truth, especially for $U \lesssim 2$. The discrepancy in the wavefunction is also reflected in the corresponding energy spectrum in Fig. 5.

An alternative approach to the iterative optimization state by state, one could treat the state k as an input to the network and minimize the total energy of multiple states simultaneously. With this approach, we first project the neural network onto the nullspace of the ground state and minimize the average energy over k states in the nullspace. In this case, the loss function landscape becomes very complex and has many degenerate local minima. While this approach avoids the error accumulation of the iterative optimization approach, we

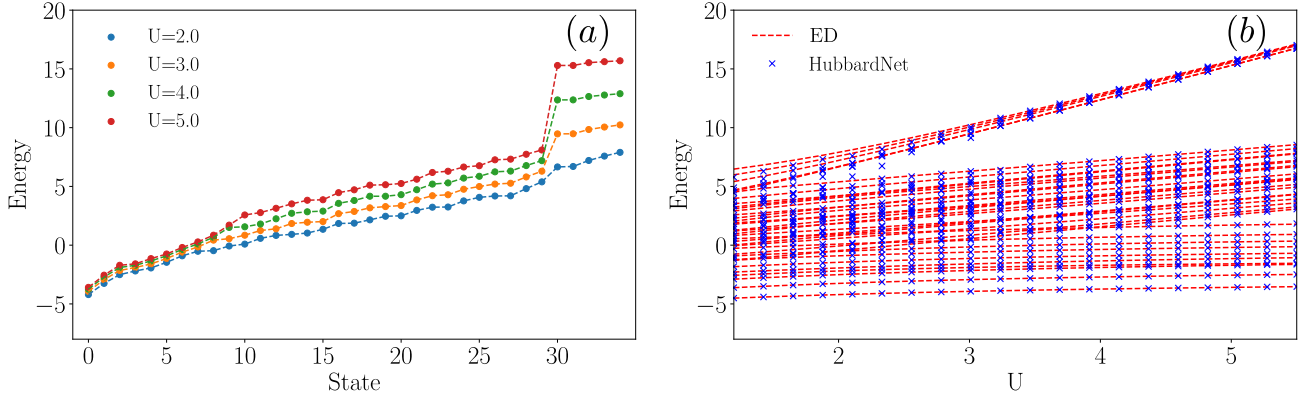


FIG. 5: **HubbardNet** energy spectrum trained using $U = 2, 3, 4, 5$ as the input, with L2-regularization with the coefficient $= 1 \times 10^{-4}$ and perturbing the U by an amplitude 0.01. (a) Energy from the **HubbardNet** (scattered points) vs. exact diagonalization (dashed lines) for the training set (b) Energy spectrum for different U 's in the testing set (scatter crosses). Eigenvalues from ED are in red dashed lines.

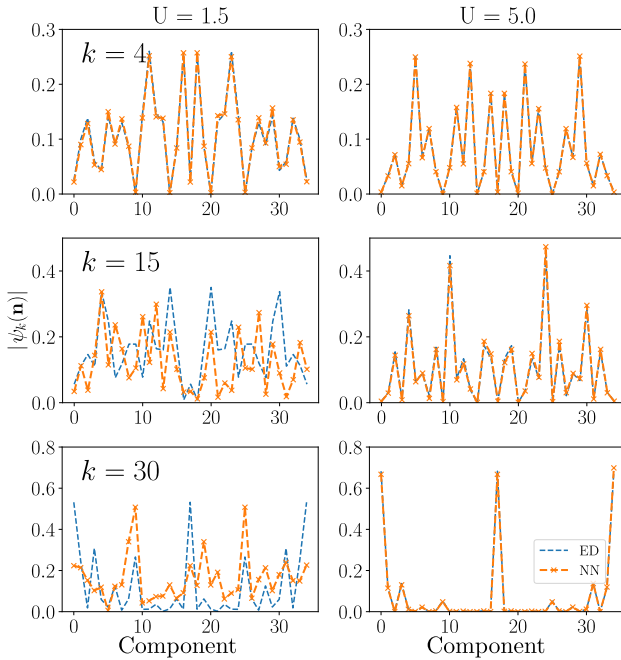


FIG. 6: Comparison between excited state wavefunctions from ED and **HubbardNet** for the 4th (top row), the 15th (middle row), and the 30th (bottom row) excited states for $U = 1.5$ (left) and $U = 5.0$ (right). In all subpanels, blue dashed lines are from ED and orange dashed lines are from **HubbardNet**.

found that the network produces less accurate predictions and the results depend more sensitively on the training parameters.

IV. PERFORMANCE EVALUATION

In all the simulations presented in Sec. III, we used a single Nvidia K80 GPU and most of the calculations can be carried out on CPUs, which suggests that our approach is relatively computationally inexpensive. In Fig. 7 we compare the performance of obtaining the ground state energy for a single value of U using ED and the **HubbardNet** method, performed on a 2.7-GHz Quad-Core Intel Core i7. Here, we take $M = N$ and a network width $D = 200$. We adopt an optimizer constant learning rate $= 0.01$ without L2 regularization, and our convergence criterion is a loss smaller than 1×10^{-6} for all cases. In Fig. 7(a), we show that the ground state energies obtained from ED and **HubbardNet** are in excellent agreement for all cases. In Fig. 7(b) we show the computational time as a function of the system size: **HubbardNet** has a better scaling compared to the ED method, even though the latter method is faster for smaller systems, with crossover being in the range $N_B \sim 5000$. Since the goal of numerical studies of the Hubbard model is to reach as large systems as possible, the advantage of the **HubbardNet** method is obvious.

An alternative approach to further improve the scaling of the DNN-based approach is that instead of calculating the sum over every pair of $\mathbf{n}_i, \mathbf{n}_j$ vectors in Eq. (12), one could evaluate the expectation value using a Monte Carlo sampler, which has been implemented in the package **netket** [31]. However, due to the stochastic nature of the Monte Carlo sample, we expect the energy to deviate from the ground truth. The accuracy of predictions is expected to deteriorate even more for excited states of higher energy due to the error accumulation in the iterative scheme.

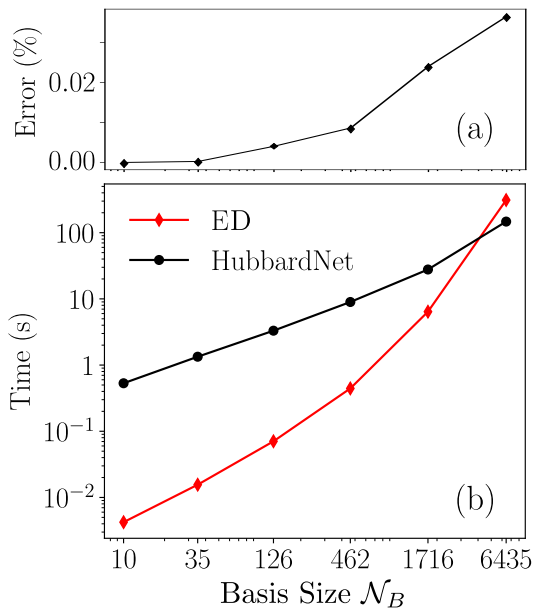


FIG. 7: Performance comparison between ED and HubbardNet using a 1D chain (for $M = N = 3, \dots, 8$) and periodic boundary conditions, as a function of the system basis-size N_B , see Eq. (4). (a) Percentage error of the ground state energies obtained with HubbardNet versus ED (ground truth). (b) Computation time for HubbardNet (black) and ED (red).

V. DISCUSSION AND OUTLOOK

In addition to the scaling advantage discussed in the previous section, our approach provides the analytical solution for a family of hamiltonians with different U and N values without the need to diagonalize a new hamiltonian with every change of each parameter. This suggests that after one training over a set of values of U and N , the network could in principle predict the phase diagram accurately. We provide a proof-of-principle demonstration in Fig. 5(b) that after training on a very small training set, we were able to obtain the energy spectrum of the system as a function of U . In particular, the network is able to predict the gap opening near $U = 2.0$. As we point out in Sec. IIIB, energies and wavefunctions from HubbardNet deviate at the gap opening point, and we suggest that this problem can be solved by increasing the training set size. The choice of training set size is a tradeoff between accuracy and training time: the larger the training set, the more accurate the predictions should be, but the harder it is to converge and the more effort is needed to tune the values of the hyperparameters. We also note that while the scaling of HubbardNet is promising, the performance evaluation we showed is only for a single state. The task of obtaining all excited states is significantly more computationally costly due to the iterative nature of our approach. Nevertheless, if one is

interested in a subset of the solutions and the dependence of the solutions on the inputs, HubbardNet would be an appropriate choice.

While we only provided here a proof-of-concept demonstration of HubbardNet, this approach could be easily adapted to solve more realistic problems including longer-ranged hopping parametrized according to the specific materials' properties as well as systems in the presence of an external potential. In addition, in this work, we solve the Bose-Hubbard model in real space and obtained the wavefunction components in the Fock basis. We could perform a Bloch expansion of Eq. (1) in momentum space and variationally minimize the expectation of the hamiltonian, with the momentum \mathbf{k} taken as an input to the network, which enables very fast prediction of the band structure.

Another interesting generalization is to use HubbardNet for the study of fermionic systems and for obtaining the phase diagrams of low-dimensional quantum systems including twisted bilayer van der Waals heterostructures and high-temperature cuprate superconductors [32–34]. The advantage of HubbardNet is that the fermionic wavefunction symmetry can be directly incorporated into the model hamiltonian and the basis set construction, similar to ED. For example, in twisted bilayer transition metal dichalcogenides (TMDCs), where unconventional superconductivity has been reported [35], the Coulomb interaction U strength is related to the band flatness and t is related to the bandwidth, both of which are functions of the twist angle θ . It has been shown both theoretically [4, 36] and experimentally [37] that twisted bilayer TMD can be accurately represented with a Fermi-Hubbard model on a triangular, honeycomb, or Kagome lattice. From the results of the first-principles calculations, one could derive reduced-band models to obtain the U/t ratio as a function of θ . HubbardNet essentially allows for the mapping of the phase boundary in twisted TMDCs as a function of θ with a single training. One obvious challenge is the large training time and the difficulty to generalize to a large system size, which can be addressed using a Monte Carlo sampler while reducing the basis set using the symmetry of the system. Our code is openly available [38].

ACKNOWLEDGMENTS

Z.Z. and E.K. acknowledge the support from the STC Center for Integrated Quantum Materials, NSF Grant No. DMR-1231319, ARO MURI Grant No. W911NF14-0247, and NSF DMREF Grant No. 1922165. Calculations were performed on the Cannon cluster supported by the FAS Division of Science, Research Computing Group at Harvard University and Google Colaboratory.

-
- [1] P. Anderson, The theory of superconductivity in the high T_c (1998).
- [2] E. W. Huang, C. B. Mendl, S. Liu, S. Johnston, H.-C. Jiang, B. Moritz, and T. P. Devereaux, Numerical evidence of fluctuating stripes in the normal state of high- T_c cuprate superconductors, *Science* **358**, 1161 (2017), [arXiv:1612.05211 \[cond-mat.str-el\]](#).
- [3] K. S. Bedell, D. Coffey, D. Meltzer, D. Pines, and J. Schrieffer, *High temperature superconductivity: Proceedings*, Tech. Rep. (Los Alamos National Lab., NM (USA), 1990).
- [4] F. Wu, T. Lovorn, E. Tutuc, and A. H. MacDonald, Hubbard model physics in transition metal dichalcogenide moiré bands, *Phys. Rev. Lett.* **121**, 026402 (2018).
- [5] Y. Xie, B. Lian, B. Jäck, X. Liu, C.-L. Chiu, K. Watanabe, T. Taniguchi, B. A. Bernevig, and A. Yazdani, Spectroscopic signatures of many-body correlations in magic-angle twisted bilayer graphene, *Nature (London)* **572**, 101 (2019), [arXiv:1906.09274 \[cond-mat.mes-hall\]](#).
- [6] H. Pan, F. Wu, and S. Das Sarma, Band topology, hubbard model, heisenberg model, and dzyaloshinskii-moriya interaction in twisted bilayer wse_2 , *Phys. Rev. Research* **2**, 033087 (2020).
- [7] H. Pan, F. Wu, and S. Das Sarma, Quantum phase diagram of a moiré-hubbard model, *Phys. Rev. B* **102**, 201104 (2020).
- [8] M. Ochi, M. Koshino, and K. Kuroki, Possible correlated insulating states in magic-angle twisted bilayer graphene under strongly competing interactions, *Phys. Rev. B* **98**, 081102 (2018).
- [9] J. F. Dodaro, S. A. Kivelson, Y. Schattner, X. Q. Sun, and C. Wang, Phases of a phenomenological model of twisted bilayer graphene, *Phys. Rev. B* **98**, 075154 (2018), [arXiv:1804.03162 \[cond-mat.supr-con\]](#).
- [10] D. Wong, K. P. Nuckolls, M. Oh, B. Lian, Y. Xie, S. Jeon, K. Watanabe, T. Taniguchi, B. A. Bernevig, and A. Yazdani, Cascade of electronic transitions in magic-angle twisted bilayer graphene, *Nature (London)* **582**, 198 (2020), [arXiv:1912.06145 \[cond-mat.mes-hall\]](#).
- [11] P. Potasz, M. Xie, and A. H. MacDonald, Exact Diagonalization for Magic-Angle Twisted Bilayer Graphene, *Phys. Rev. Lett.* **127**, 147203 (2021), [arXiv:2102.02256 \[cond-mat.str-el\]](#).
- [12] D. J. Scalapino, S. R. White, and S. C. Zhang, Superfluid density and the drude weight of the hubbard model, *Phys. Rev. Lett.* **68**, 2830 (1992).
- [13] J. K. Freericks and H. Monien, Phase diagram of the bose-hubbard model, *Europhysics Letters (EPL)* **26**, 545 (1994).
- [14] T. D. Kühner and H. Monien, Phases of the one-dimensional bose-hubbard model, *Phys. Rev. B* **58**, R14741 (1998).
- [15] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017), [arXiv:1606.02318 \[cond-mat.dis-nn\]](#).
- [16] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, Restricted boltzmann machine learning for solving strongly correlated quantum systems, *Phys. Rev. B* **96**, 205152 (2017).
- [17] S. Bravyi, D. Gosset, R. König, and K. Temme, Approximation algorithms for quantum many-body problems, *Journal of Mathematical Physics* **60**, 032203 (2019), [arXiv:1808.01734 \[quant-ph\]](#).
- [18] H. Saito, Solving the Bose-Hubbard Model with Machine Learning, *Journal of the Physical Society of Japan* **86**, 093001 (2017), [arXiv:1707.09723 \[cond-mat.dis-nn\]](#).
- [19] H. Saito and M. Kato, Machine Learning Technique to Find Quantum Many-Body Ground States of Bosons on a Lattice, *Journal of the Physical Society of Japan* **87**, 014001 (2018), [arXiv:1709.05468 \[cond-mat.dis-nn\]](#).
- [20] K. McBrien, G. Carleo, and E. Khatami, Ground state phase diagram of the one-dimensional Bose-Hubbard model from restricted Boltzmann machines, in *Journal of Physics Conference Series*, Journal of Physics Conference Series, Vol. 1290 (2019) p. 012005, [arXiv:1903.03076 \[cond-mat.str-el\]](#).
- [21] K. Choo, A. Mezzacapo, and G. Carleo, Fermionic neural-network states for ab-initio electronic structure, *Nature Communications* **11**, 2368 (2020), [arXiv:1909.12852 \[physics.comp-ph\]](#).
- [22] J. Robledo Moreno, G. Carleo, A. Georges, and J. Stokes, Fermionic Wave Functions from Neural-Network Constrained Hidden States, *arXiv e-prints*, [arXiv:2111.10420](#) (2021), [arXiv:2111.10420 \[cond-mat.str-el\]](#).
- [23] N. Yoshioka, W. Mizukami, and F. Nori, Solving quasi-particle band spectra of real solids using neural-network quantum states, *Communications Physics* **4**, 106 (2021), [arXiv:2010.01358 \[cond-mat.str-el\]](#).
- [24] E. R. Bennewitz, F. Hopfmueller, B. Kulchytskyy, J. Carrasquilla, and P. Ronagh, Neural Error Mitigation of Near-Term Quantum Simulations, *arXiv e-prints*, [arXiv:2105.08086](#) (2021), [arXiv:2105.08086 \[quant-ph\]](#).
- [25] M. J. Hartmann and G. Carleo, Neural-Network Approach to Dissipative Quantum Many-Body Dynamics, *Phys. Rev. Lett.* **122**, 250502 (2019), [arXiv:1902.05131 \[quant-ph\]](#).
- [26] A. Nagy and V. Savona, Variational Quantum Monte Carlo Method with a Neural-Network Ansatz for Open Quantum Systems, *Phys. Rev. Lett.* **122**, 250501 (2019), [arXiv:1902.09483 \[quant-ph\]](#).
- [27] F. Vicentini, A. Biella, N. Regnault, and C. Ciuti, Variational Neural-Network Ansatz for Steady States in Open Quantum Systems, *Phys. Rev. Lett.* **122**, 250503 (2019), [arXiv:1902.10104 \[quant-ph\]](#).
- [28] A. V. Ponomarev, S. Denisov, and P. Hänggi, ac-Driven Atomic Quantum Motor, *Phys. Rev. Lett.* **102**, 230601 (2009), [arXiv:0902.0489 \[physics.atom-ph\]](#).
- [29] A. V. Ponomarev, S. Denisov, and P. Hänggi, Lévy distribution in many-particle quantum systems, *Phys. Rev. A* **81**, 043615 (2010), [arXiv:0907.4328 \[cond-mat.quant-gas\]](#).
- [30] D. Raventós, T. Graß, M. Lewenstein, and B. Juliá-Díaz, Cold bosons in optical lattices: a tutorial for exact diagonalization, *Journal of Physics B Atomic Molecular Physics* **50**, 113001 (2017), [arXiv:1410.7280 \[cond-mat.quant-gas\]](#).
- [31] G. Carleo, K. Choo, D. Hofmann, J. E. T. Smith, T. Westerhout, F. Alet, E. J. Davis, S. Efthymiou, I. Glasser, S.-H. Lin, M. Mauri, G. Mazzola, C. B. Mendl, E. van Nieuwenburg, O. O'Reilly, H. Théveniaut, G. Torlai, F. Vicentini, and A. Wietek, Netket: A machine learning toolkit for many-body quantum systems, *Soft-*

- [wareX](#), 100311 (2019).
- [32] Y. Cao, V. Fatemi, S. Fang, K. Watanabe, T. Taniguchi, E. Kaxiras, and P. Jarillo-Herrero, Unconventional superconductivity in magic-angle graphene superlattices, *Nature (London)* **556**, 43 (2018), [arXiv:1803.02342 \[cond-mat.mes-hall\]](#).
 - [33] E. Fradkin, S. A. Kivelson, and J. M. Tranquada, Colloquium: Theory of intertwined orders in high temperature superconductors, *Reviews of Modern Physics* **87**, 457 (2015), [arXiv:1407.4480 \[cond-mat.supr-con\]](#).
 - [34] F. Heidrich-Meisner, A. Honecker, D. C. Cabra, and W. Brenig, Comment on “anomalous thermal conductivity of frustrated heisenberg spin chains and ladders”, *Phys. Rev. Lett.* **92**, 069703 (2004).
 - [35] L. Wang, E.-M. Shih, A. Ghiotto, L. Xian, D. A. Rhodes, C. Tan, M. Claassen, D. M. Kennes, Y. Bai, B. Kim, K. Watanabe, T. Taniguchi, X. Zhu, J. Hone, A. Rubio, A. N. Pasupathy, and C. R. Dean, Correlated electronic phases in twisted bilayer transition metal dichalcogenides, *Nature Materials* **19**, 861 (2020).
 - [36] M. Angeli, G. R. Schleder, and E. Kaxiras, Twistronics of Janus transition metal dichalcogenide bilayers, *arXiv e-prints*, [arXiv:2207.05788](#) (2022), [arXiv:2207.05788 \[cond-mat.mtrl-sci\]](#).
 - [37] Y. Xu, K. Kang, K. Watanabe, T. Taniguchi, K. F. Mak, and J. Shan, A tunable bilayer Hubbard model in twisted WSe₂, *Nature Nanotechnology* **17**, 934 (2022), [arXiv:2202.02055 \[cond-mat.str-el\]](#).
 - [38] Z. Zhu, M. Marios, P. Weiwei, and K. Efthimios, *Model for HubbardNet: Efficient Predictions of the Bose-Hubbard Model Spectrum with Deep Neural Networks: <https://github.com/ziyanzhu/HubbardNet>* (2022).