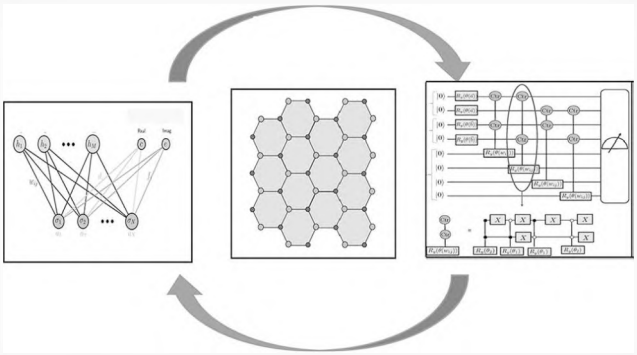
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Quantum Machine-Learning for Eigenstate Filtration in Two-Dimensional Materials

Manas Sajjan, Shree Hari Sureshbabu, and Sabre Kais*

Cite This: https://doi.org/10.1021/jacs.1c06246
Read Online
ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Quantum machine-learning algorithms have emerged to be a promising alternative to their classical counterparts as they leverage the power of quantum computers. Such algorithms have been developed to solve problems like electronic structure calculations of molecular systems and spin models in magnetic systems. However, the discussion in all these recipes focuses specifically on targeting the ground state. Herein we demonstrate a quantum algorithm that can filter any energy eigenstate of the system based on either symmetry properties or a predefined choice of the user. The workhorse of our technique is a shallow neural network encoding the desired state of the system with the amplitude computed by sampling the Gibbs–Boltzmann distribution using a quantum circuit and the phase information obtained classically from the nonlinear activation of a separate set of neurons. We show that the resource requirements of our algorithm are strictly quadratic. To demonstrate its efficacy, we use state filtration in monolayer transition metal dichalcogenides which are hitherto unexplored in any flavor of quantum simulations. We implement our algorithm not only on quantum simulators but also on actual IBM-Q quantum devices and show good agreement with the results procured from conventional electronic structure calculations. We thus expect our protocol to provide a new alternative in exploring the band structures of exquisite materials to usual electronic structure methods or machine-learning techniques that are implementable solely on a classical computer.



1. INTRODUCTION

Machine learning concerned with identifying and utilizing patterns within a data set has gained tremendous importance within the past decade. Even though the germinal idea can be traced back to the 1950s,[1] it is safe to say that the domain has become a pioneering field of research within the past few years due to escalation in computational prowess and data availability and that it has metamorphosed several disciplines including autonomous driving,[2] image recognition,[3] speech recognition,[4] natural language processing,[5] computer games,[6] and even refugee integration.[7] Consequently, the integration of the technique in solving problems of physicochemical interest[8] have also been explored with remarkable success whether in predicting ground-state density functionals,[9,10] self-energy in dynamical mean-field theory (DMFT) for the Anderson model,[11] atomistic potentials and force fields for molecular dynamics,[12,13] or the unsupervised learning of phases of the 2D-Ising Hamiltonian.[14] Similar advancements have also been made in the fields of deep learning[15] and artificial neural networks (ANN) which have been used successfully to learn phase transition parameters[16,17] or in quantum phase recognition.[18] Among the various architectures in this category, restricted Boltzmann machine (RBM)-based generative models being a universally powerful approximator for any probability density[19,20] have particularly gained attention. RBMs have been successfully used to reconstruct quantum states in tomography from measurement statistics.[21] Carleo and Troyer showed how a neural network encoding a shallow RBM ansatz requires fewer parameters than certain kinds of matrix product states and predicts the ground-state energy and unitary dynamical evolution of simple spin models with high accuracy.[22]

However, all the algorithms discussed above have trained machine-learning or deep-learning models on a classical computer to effectively recreate either a quantum state or its essential features. The past decade has also witnessed unprecedented development in quantum computing as a new paradigm which is fundamentally different than its classical counterpart in processing and storing data and performing logical operations[23] harnessing the power of quantum

Received: June 17, 2021

superposition and nonclassical correlations like entanglement. A natural question that has spawned is whether such quantum machines can interpret and produce statistical patterns in data which are either difficult for classical machine-learning algorithms or the performance of machine-learning algorithms on quantum computer can outperform the classical variants in efficiency.[24] This has naturally motivated the development of a host of quantum machine-learning algorithms like Quantum Principal Component Analysis (PCA),[25] Quantum Support Vector Machines (QSVM),[26] Quantum Reinforcement Learning,[27] quantum supervised and unsupervised learning,[28] kernel design for Gaussian processes,[29] Gaussian process regression,[30] quantum classifier,[31] or a plethora of linear algebra routines like HHL,[32] QSVD,[33] and qBLAS[34] which form the backbone of the quantum versions of many other machine-learning algorithms. Each of these methods has reported theoretical speedup over the best-known classical algorithm under certain specific circumstances.[35] Similar investigations have also been undertaken for artificial neural networks to discover any unforeseen quantum advantage. For instance, Amin and coworkers have demonstrated a quantum Boltzmann machine[36] by adding an off-diagonal transverse field to the training model thereby making it more expressive to treat larger classes of problems.[37] Weibe et al. have shown how sampling from a Gibbs distribution as is required for training an RBM can be distinctly accelerated using a quantum processor.[38]

Motivated by such recent developments, Xia and Kais[39] proposed an actual quantum circuit using polynomial resources to correctly learn the amplitude of the RBM ansatz encoded within a neural network representing the state of a quantum system. The work also extended the neural network to three layers to learn the sign of the various components of the encoded wave function. The algorithm was benchmarked by showing the evaluation of ground states on simple molecular systems like H_2, LiH, and so on, thereby formally extending the efforts mentioned above to actual electronic structure calculations which are considered to be powerful applications of near-term quantum devices. Indeed, interesting algorithmic advances have been made recently that can capture both the ground and excited states of such electronic structure problems with good accuracy.[40−46] Kanno et al.[47] modified the above method to encompass the complex phase of each component of the wave function by adding an additional neuron to the third layer. However, both the works simulated the performance of the algorithm for ground states only on noiseless classical devices. In fact, due to conditional dependence on the sequence of measurements of the ancilla register, straightforward implementation of the algorithm on a present-day actual NISQ device is difficult.

The main contributions of this manuscript are as follows: (1) Unlike previous efforts, we focus our attention beyond just the ground state and devise a quantum machine-learning algorithm with a three-layered RBM being trained to learn any arbitrary state of the system retaining the quadratic resource requirements. To train the network with the RBM ansatz, we employ a hardware-implementable version of the above quantum circuit which as we shall discuss explicitly makes our algorithm require quadratic resources in all fronts like circuit width, circuit depth, and parameter count. (2) A generic lower bound for the successful sampling of the quantum circuit in the algorithm is derived in terms of the parameters of the network. The performance of the lower bound is thoroughly characterized, and specific limiting cases leading to known bounds are formally deduced and discussed. On the basis of this we also discuss in detail the measurement statistics of our algorithm for systems studied in this report and in general. (3) We also present a simple yet formal proof of feasibility of the cost function used to train the network. Even though such functions have been used in classical algorithms and are beginning to being used in other quantum algorithms beyond the precincts of quantum machine learning, a formal proof is lacking in literature which we supply here for completeness. (4) Furthermore, unlike most reports on quantum machine learning and quantum computing in general which have studied molecular systems only, we apply our algorithm on important 2D materials like monolayer transition metal dichalcogenides (TMDCs) which are hitherto unexplored on a quantum computer using any algorithm let alone quantum machine learning. These materials have been shown to possess tunable band gap for many novel applications.[48−51] We make a comprehensive study of such materials by showing how our algorithm can not only learn the true band gap but also by resolving finer yet important features like trigonal warping and spin−orbit coupling (SOC) which dictates the low-energy physics near the *K*-valley. The importance of understanding excited states beyond just the valence band for such periodic materials underlies its function in photovoltaics.[52,53] If applied to other systems, then excited states can be an insightful resource like in elucidating the reaction pathways across conical intersections arising in processes like vision,[54,55] photosynthesis,[56,57] magneto-reception,[58,59] and even the biochemistry of luciferin[60] to name a few. (5) We further demonstrate in a unified way how a user can sieve any desired state in such materials using not only energy as in point (4) but also other inherent symmetries of the Hamiltonian. (6) All numerical experiments are implemented on not only quantum simulator (Qiskit) but also actual NISQ devices using the quantum processors at IBM.[61] The performance of the algorithm is benchmarked thoroughly in each case using quantifiers like energy errors of the target state, state composition, constraint violation, infidelity with the target state learnt by the neural network, and so on. The usage of certain kinds of error-mitigation techniques and the role of initial parametrization enhancing the capacity of the model through additional spins in the network is thoroughly discussed. We have also included results from a molecular example wherein multireference correlation is important due to geometric distortion. To the best of our knowledge, all of these are first of their kind in any flavor of quantum machine learning. We show that the performance of our algorithm is in excellent agreement with the exact value in each case.

The organization of this paper is as follows. In section 2, we discuss the theoretical underpinning of our algorithm with an original proof of the feasibility of our cost function employed for training the network. In section 3, we elaborate on the geometry of the network and the details of the algorithm required for learning the desired feature with the associated resource requirements and implementation techniques. We prove an explicit lower bound on the probability of successful events on our algorithm (see Supporting Information section 2). In section 4, we discuss the application of the algorithm in simulating excited states or any arbitrary states in two important TMDCs: MoS_2- and WS_2-based on user-defined constraints. We conclude in section 5 with a brief discussion of possible future extensions.

https://doi.org/10.1021/jacs.1c06246
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

2. THEORY

Our objective is to develop an efficient algorithm to train a neural network to perform the following minimization in a d-dimensional space

$$\min_{\forall \psi \in S} \langle \psi | \hat{H} | \psi \rangle$$
$$S = \{ |x\rangle | \hat{O}|x\rangle = \omega |x\rangle \ \forall \ |x\rangle \in \mathbb{C}^d \} \quad (1)$$

where $\hat{H} \in \mathbb{C}^{d \times d}$ is the hermitian Hamiltonian defining the problem. Similarly, $\hat{O} \in \mathbb{C}^{d \times d}$ is the user-defined hermitian operator. ω is the eigenvalue (real-valued) of the operator $\hat{O}$ and $|x\rangle$ is the corresponding eigenvector. The set S is the collection of all such eigenvectors with a specific eigenvalue ω. The operators $\hat{O}$ which we shall discuss will generally have more than one element in set S due to degeneracy in the eigenspace labeled by ω. By construction, the form of the algorithm shall always normalize the state $|\psi\rangle$; hence, normalization as a further constraint is unnecessary. We will return to this point later. The primary goal of the network is to then encode a normalized state-vector $|\psi\rangle$ which is a formal solution to eq 1. The corresponding state so obtained is from the eigenspace of $\hat{O}$ with eigenvalue ω. If several such choices exist, then the network learns the one with minimum energy.

To solve the quadratic minimization problem with quadratic constraint in eq 1, we will define a penalty procedure as

$$F(|\psi\rangle, \hat{H}, \hat{O}, \lambda) = \langle \psi | \hat{H} | \psi \rangle + \lambda \langle \psi | (\hat{O} - \omega)^2 | \psi \rangle \quad (2)$$

where $\lambda \geq 0$ is the penalty parameter. We provide a formal and original proof of equivalence of eq 2 with respect to eq 1 based on the following Theorem.

Theorem 2.1. Let $\{\lambda_i\}_{i=1}^{\infty}$ be a sequence in the penalty parameter such that $\lambda_1 \leq \lambda_2 \leq \lambda_3 \ldots \lambda_\infty \to \infty$. Also let $P = \{|\psi_i\rangle\}_{i=1}^{\infty}$ such that $\forall \ |\psi_i\rangle \in P$ the following is true.

$$|\psi_i\rangle = \arg\min_{\psi} F(\lambda_i, \hat{H}, \hat{O}, |\psi\rangle) \quad (3)$$

In other words, P is the set of minimizers for eq 2 for each penalty parameter $\lambda \in \{\lambda_i\}_{i=1}^{\infty}$. If $|\psi^*\rangle \in P$ is a limit-point of the convergent sequence $\{\psi_i\}_{i=1}^{\infty}$ in P, i.e., $|\psi^*\rangle = \lim_{i \to \infty} |\psi_i\rangle$, then $|\psi^*\rangle \in S$

An original proof of Theorem 2.1 is given in section 1 of the Supporting Information based on the fact that both the first and second term in eq 2 are quadratic forms. An intuitive explanation can be provided that would suffice to appreciate the discussion in this report. One can note that in the cost function defined in eq 2 the term $\langle \psi | \hat{H} | \psi \rangle$ imposes the minimization of energy as required in eq 1. The second term, i.e., $\langle \psi | (\hat{O} - \omega)^2 | \psi \rangle$, is the variance of the operator $\hat{O}$ with the mean being the eigenvalue ω and is non-negative by construction. For large values of the penalty parameter λ, the minimization of the overall cost function is afforded if the variance term is pinned to zero, i.e., the state $|\psi^*\rangle$ so chosen is an eigenstate of the operator $\hat{O}$ with eigenvalue ω. The space of such states is defined by the set S in eq 1. If several such choices exist, then the role of the first term kicks in to guarantee optimality in energy.

While penalized optimization schemes with cost function of the kind in eq 2 have been employed in classical algorithms like Density Matrix Renormalization Group (DMRG),[62] in Quantum Monte Carlo methods in the past,[63] and even recently[64] and are also are beginning to gain attention in recent literature on quantum algorithms beyond quantum machine learning, i.e., in algorithms using Unitary-Coupled Cluster Ansatz (UCC) of variational quantum eigensolver (VQE),[65] yet a formal proof is lacking. Besides a more popular choice that has been studied in some detail is constraining the average value of the operator $\langle \psi | \hat{O} | \psi \rangle$[66,67] with the required eigenvalue instead of penalizing the variance as in eq 2. However, this recent study[65] shows eq 2 is a better penalty procedure in terms of feasibility and final error than restraining the average without providing a formal proof of equivalence between eqs 2 and 1. Ref 65 also implemented the same to target symmetry operators on molecular systems using UCC-VQE using Qulacs[68] which is an ideal simulator of a real quantum computer. However, in this report, we shall use eq 2 to develop and train a shallow neural network using a quantum machine-learning algorithm with quadratic resource requirements in terms of the size of the qubit register, the number of gates, and parameter counts. The ansatz which the neural network would encode for the quantum state $|\psi\rangle$ would correspond to a probability density represented by RBM. We benchmark our algorithm on important 2D periodic materials like TMDCs and show implementations not only on quantum simulators but on actual NISQ devices (IBM-Q). TMDCs have never been studied before using any quantum algorithm. In the next few sections, we shall show how to filter any specific state of these 2D materials using either symmetry operators of the Hamiltonian or user-defined constructions of operator $\hat{O}$ in a unified manner using the same algorithm. Such an attempt to the best of our knowledge is the first of its kind in QML as all previous reports have focused exclusively on targeting the ground state of the system alone.[47,69]

2.1. Filter for Specific Excited States. To target the first excited state of the system, one can use a user-defined operator ($\hat{O} = |g\rangle\langle g|, \omega = 0$) where $|g\rangle$ is the ground state of the system obtained by training the network in a previous computation with $\lambda = 0$ in eq 2. In essence, we require the neural network to return a state vector in the null space of operator $|g\rangle\langle g|$. Since the null space is $(d - 1)$-dimensional, the minimum energy criterion as enforced by the first term in eq 2 guarantees the first excited state. This method using the penalty program in eq 2 is formally equivalent to the deflation technique if one recognizes the idempotency of $\hat{O} = |g\rangle\langle g|$. Deflation has been the cornerstone of many classical algorithms in the past for obtaining excited states[70,71] and even a quantum algorithm as well with UCC-VQE.[72] However, the formal reduction of our penalty procedure to deflation in eq 2 based on Theorem 2.1 offers a slightly different perspective. Moreover, as we shall see shortly, the penalty program in eq 2 is more general and can be used to sieve any state based on arbitrary operator $\hat{O}$. For higher excited states (say the tth), one can add similar terms to eq 2 with the set $\{\hat{O}_i\}_{i=i}^{t-1}$ which forms a set of commuting operators with progressively refined null space. For the choice of the penalty parameter λ in eq 2, one can choose any number greater than the spectral range of the Hamiltonian $\hat{H}$ as that would always work. The spectral range can be computed from the knowledge of the ground state and $\|\hat{H}\|_2$.

2.2. Filter for Arbitrary States Using Symmetry Operators. Equation 2 can be used to solve a more general problem with any symmetry operator of the system $\hat{O}$ (by definition such operators satisfy $[\hat{O}, \hat{H}] = 0$ and hence share the same eigenspace). The corresponding user-desired eigenvalue ω labels the symmetry sector (set S in eq 1). Unlike in the previous case in section 2.1, the usual symmetry

https://doi.org/10.1021/jacs.1c06246
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

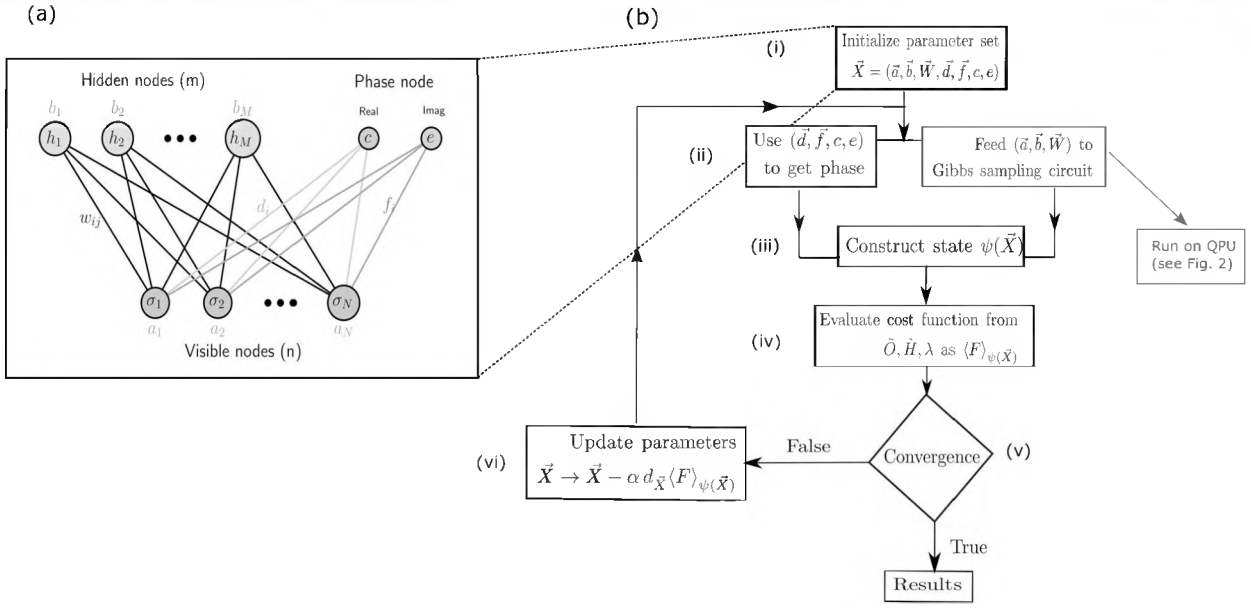


Figure 1. (a) RBM architecture used in this work. The visible node contains n neurons (green); the hidden node has m neurons (blue). The phase node contains 2 neurons: one to model the real part (orange) of the phase of the wave function and the other to model the imaginary part (gray). The weights and biases of the respective units are displayed. The RBM ansatz for the required state is defined from the Boltzmann distribution over the state-space of the visible-hidden units. (b) QML algorithm used to perform the variance penalized optimization. The part of step (ii) marked within the red box is performed on a quantum processor (QPU). All other steps are performed on a classical computer. Each step is marked with a roman numeral. We follow each of these roman numerals for discussing the algorithm in section 3.2.

operators need not satisfy idempotency and hence relaxation to deflation is impossible. To demonstrate our point, here we shall use $\hat{O} = L^2$ where L^2 is the squared-orbital angular momentum operator, a symmetry for 2D materials. ω would be set to the desired eigenvalue of L^2. We shall see that the network will always learn the lowest energy eigenstate correctly despite multiple-fold degeneracy. To sieve other states from the entire degenerate subspace, one can use a combination filter of $\hat{O}_1 = L^2$ and $\hat{O}_2 = |v\rangle\langle v|$ where $|v\rangle$ is the lowest energy state in the symmetry subspace obtained from the RBM. The penalty parameter λ can be chosen using the prescription in ref 65.

3. ALGORITHM

3.1. The Model. In the early 1980s, Hopfield networks[73] defined a probability distribution over a set of random variables which is encoded within the nodes of a unidirected graph using the physical notion of energy of interaction between the nodes. Boltzmann machines (BM) are extensions of such a network that categorizes the node space into visible/physical layer and hidden/latent layers maintaining all to all connectivity.[74] A restricted Boltzmann machine (RBM)[8,20,75−78] is a practically useful subcategory of BM which permits interaction only between the visible layer and hidden layer. The energy function used in the RBM model is thus that of a partially connected classical Ising network and the ansatz for the probability distribution is the corresponding thermal distribution. The ansatz is optimized to mimic the underlying probability distribution of the given data using free parameters called *weights* and *biases*.[19,76−81] The goal of this paper is to use the RBM distribution to encode the amplitude field of an arbitrary quantum state $|\psi\rangle$ which is a solution to eq 1. Such neural-network quantum states (NQS) have been successfully employed in a variety of problems recently[8,20,22,82] by training the *weights* and *biases* using a classical computer. Herein, we shall train the network by constructing the RBM distribution using a quantum circuit and discuss the quantum advantages.

The RBM network we use in this report consists specifically of three layers each having multiple neurons. The schematic of the network architecture is presented in Figure 1a. The first layer is the visible node consisting of n neurons; the second layer is the hidden node consisting of m neurons. The last layer is a phase node consisting of two neurons. While the n neurons are responsible for encoding the actual state, the purpose of the hidden neurons m is to add more controllable parameters to make the joint probability distribution (to be defined in eq 4) more expressive and induce higher order correlation among n neurons.[20] Variables encoded by the visible node neurons (henceforth denoted by $\{\sigma_i\}_{i=1}^n$) and those by the hidden node neurons (henceforth denoted by $\{h_j\}_{j=1}^m$) are both binary random variables as σ_i and $h_j \in \{1, -1\}$. As depicted in Figure 1a, the bias vector of the visible neurons is denoted as $\vec{a} \in \mathbb{R}^n$, the bias vector of hidden neurons is denoted as $\vec{b} \in \mathbb{R}^m$, and the interconnecting weights of the visible and hidden neurons are denoted as $\vec{W} \in \mathbb{R}^{n\times m}$. The joint RBM distribution[8,20,76−78] $P(\vec{a}, \vec{b}, \vec{W}, \vec{\sigma}, \vec{h})$ defined over the variables $(\vec{\sigma}, \vec{h})$ is

$$P(\vec{a}, \vec{b}, \vec{W}, \vec{\sigma}, \vec{h}) = \frac{e^{\sum_i a_i\sigma_i + \sum_j b_j h_j + \sum_{ij} w_{ij}\sigma_i h_j}}{\sum_{\{\sigma h\}} e^{\sum_i a_i\sigma_i + \sum_j b_j h_j + \sum_{ij} w_{ij}\sigma_i h_j}} \tag{4}$$

For an electronic Hamiltonian with r spin−orbitals and N electrons, a naive Jordan Wigner mapping (JW)[83] would make $n = r$ or (or $n \approx O(\log_2(r))$ for Bravi−Kitaev mapping).[83] However, it is now well-understood that qubit requirements can be tapered by using additional symmetries like Z_2.[84] The chemically inspired process of reducing qubit cost like using an active space[85] (wherein number of physical qubits required is

https://doi.org/10.1021/jacs.1c06246
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

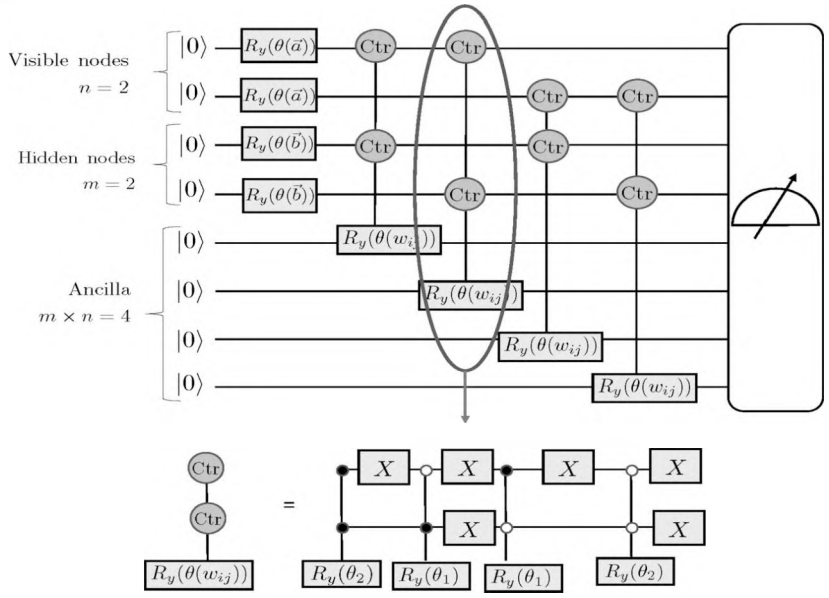


Figure 2. Gibbs sampling quantum circuit used to create the Boltzmann distribution in Figure 1b (highlighted within the red box in Figure 1(b) step (ii)) for the case of $n = m = 2$. The circuit contains single-qubit R_y gates parametrized by biases $(\vec{a}, \vec{b})$ of hidden and visible neurons and $C-C-R_y$ gates parametrized by weights $\vec{W}$ between the hidden and visible neurons. Each $C-C-R_y$ gate is conditioned to rotate by different angles θ_1 and θ_2 for different choices of configurations of the control qubits. This can be implemented by use of X gates as illustrated at the bottom. The open circles show a node in state $|0\rangle$, and the closed circles show a node in state $|1\rangle$. At the end of the circuit, all qubits are measured and configurations wherein the ancilla qubits are all in state $|1\rangle$ are postselected (see text for details). For $(n + m)$ visible and hidden neurons, there will be $(n + m)$ visible and hidden qubits and $(n + m)$ single R_y gates as there are that many biases. However, since the $C-C-R_y$ gates are always controlled by 1 visible and 1 hidden qubit, there will be $m \times n$ such possibilities each of which targets one ancilla thereby making the size of the ancilla register $m \times n$. Thus, there will be $O(n \times m)$ gates and number of qubits in the circuit. We discuss this further in section 3.3.

still r, but logical qubits required are much less as some qubits have frozen occupation/eigenvalue with Z operator) or using point-group symmetry or angular-momentum symmetry of the required state[84,86] is also being recently employed. Thus, a direct relationship of n with r would depend on the specification of the mapping and tapering used. Whatever may be the method, if the final $\hat{H}$ matrix is $\mathbb{C}^{d\times d}$ (as used in section 2), then it is safe to say that in our model $n = \log_2(d)$. The number of hidden units m in our model is user-defined (for almost all data in this manuscript we have used $n = m$), but the hidden node density $\alpha = m/n$ can be tuned to enhance the final accuracy desired. We shall return to this point later. Neurons in the phase node are always 2 in number.

The purpose of the neurons in the phase node is to account for complex values and capture the phase of the wave function[47] unlike in conventional two-layer RBM networks[21] which faithfully recovers only the amplitude. As shown in Figure 1a, for the phase node, the biases are denoted by $\{c, e\} \in \mathbb{R}^2$ where c is the bias for the neuron capturing the real part of the phase and e is the bias for the neuron encoding the corresponding imaginary part. The phase node shares interconnections with the visible node only and is defined by $\vec{d} \in \mathbb{R}^n$ for the real part of the phase and $\vec{f} \in \mathbb{R}^n$ for the associated imaginary part. The corresponding phase function for the quantum state $|\psi\rangle$ defined using these nodes is

$$s(\vec{d}, \vec{f}, c, e, \vec{\sigma}) = \tanh\left[\left(c + \sum_i d_i\sigma_i\right) + i\left(e + \sum_i f_i\sigma_i\right)\right] \tag{5}$$

Together, the set $\vec{X} = (\vec{a}, \vec{b}, \vec{W}, \vec{d}, \vec{f}, c, e)$ thus defines the complete set of trainable parameters of the model which the network shall learn iteratively to mimic the coefficients of the quantum state $|\psi\rangle$ in the chosen basis. We shall discuss the algorithm to do this in the next section.

3.2. Outline of the Method. The entire algorithm is schematically depicted in Figure 1b. It goes as follows.

(i)The first step is to initialize the parameters in the parameter vector $\vec{X} = (\vec{a}, \vec{b}, \vec{W}, \vec{d}, \vec{f}, c, e)$ on a classical computer. All parameters are randomly initialized in the parameter range $[-0.02, 0.02]$ to avoid the vanishing gradient of the activation function for the phase node.[39] Sometimes if random initialization returns a poorly converged result, then we use the initial parameter set of a converged point in a similar problem as the starting guess, a process known as warm optimization.

(ii) In the second step, the set $(\vec{a}, \vec{b}, \vec{W})$ is fed into a quantum circuit for Gibbs sampling shown in Figure 2. This step is performed on a quantum computer. The circuit requires $n + m$ qubits to encode the visible node and the hidden node respectively and additionally $m \times n$ ancillary qubits. The entire register is initialized to $|0\rangle$. The purpose of the circuit is to sample a bit string $(\vec{\sigma}, h) \in \{1, -1\}^{m+n}$ from the RBM distribution $P(\vec{a}, \vec{b}, \vec{W}, \vec{\sigma}, \vec{h})$ defined in before in eq 4.[69] In reality, the circuit actually draws a sample $(\vec{\sigma}, \vec{h})$ from

$$Q(\vec{a}, \vec{b}, \vec{W}, \vec{\sigma}, \vec{h}) = \frac{e^{1/k(\sum_i a_i\sigma_i + \sum_j b_jh_j + \sum_{ij} w_{ij}\sigma_ih_j)}}{\sum_{\{\sigma h\}} e^{1/k(\sum_i a_i\sigma_i + \sum_j b_jh_j + \sum_{ij} w_{ij}\sigma_ih_j)}} \tag{6}$$

https://doi.org/10.1021/jacs.1c06246
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

and then reconstructs $P(\vec{a}, \vec{b}, \vec{W}, \vec{\sigma}, \vec{h}) \propto Q(\vec{a}, \vec{b}, \vec{W}, \vec{\sigma}, \vec{h})^k$. The real-valued parameter k will be discussed shortly.

The state of the visible node qubits and hidden node qubits are denoted henceforth as $|\sigma_i\rangle$ and $|h_j\rangle$ respectively. When σ_i (or h_j) $= -1$, $|\sigma_i\rangle$ (or $|h_j\rangle$) $= |0\rangle$ and $|1\rangle$ otherwise. In the circuit shown in Figure 2, the single-qubit R_y gates acting only on the visible and hidden units have rotation angles parametrized by $(\vec{a}, \vec{b})$and are responsible for creating the noninteracting part of the distribution in $Q(\vec{a}, \vec{b}, \vec{W}, \vec{\sigma}, \vec{h})$, while the interaction terms $\{\sum_{i,j} w_{ij}\sigma_i h_j\}$ are turned on through using $C-C-R_y$ gates acting on ancilla register as the target. The rotation angles of these doubly controlled R_y gates are parametrized by $\vec{W}$ and are different for different configurations of the control qubits (always 1 hidden and 1 visible). Various such choices can be realized by using X gates as shown in Figure 2. After all such operations, we measure all the $(m + n + m \times n)$ qubits and after select the results wherein the ancilla qubits have collapsed to state $|1111\ldots 1_{mn}\rangle$ only. We show that the probability of such a successful event has a generic lower bound determinable in terms of the parameters of the network $(\vec{a}, \vec{b}, \vec{W})$ (for details of the derivation of the generic bound, refer to section 2 of the Supporting Information). This master lower bound generalizes the previously noted one[39] as a special case. The role of the real-valued parameter k kicks in here. It serves as a regulator and is chosen in simulation to make the aforesaid lower bound a constant value (see section 2 in the Supporting Information). After the postselection, the corresponding states of the visible and hidden units are equivalent to all possible bit strings sampled from the distribution $Q(\vec{a}, \vec{b}, \vec{W}, \vec{\sigma}, \vec{h})$ from which the desired distribution $P(\vec{a}, \vec{b}, \vec{W}, \vec{\sigma}, \vec{h})$ is constructed. The primary quantum advantage in our algorithm comes at this step where the full RBM distribution is constructed. Indeed, we shall elaborate in section 3.3 that there are no polynomial-time classical algorithms for the construction of full RBM distribution. In our case, we can access the full distribution using quadratic resources by leveraging a quantum computer. The physical reason for this advantage is rooted in quantum parallelism which before a projective measurement allows the general state of the $(m + n + m \times n)$ qubits to be a superposition of all possible bit-strings with the coefficients sampled from the full RBM distribution. Many such measurements are necessary to construct the RBM distribution encoding the target state as postmeasurement we can retrieve only one such bit-string. As explained above, the k parameter in our model is useful here as it can be adaptively chosen by the user to control the measurement statistics (see section 2 in the Supporting Information). Besides, for all systems primarily treated in this manuscript, we shall show that the chances of the ancilla register collapsing in the favorable state are naturally high even for modest values of the k parameter (see section 5 in the Supporting Information). With $P(\vec{a}, \vec{b}, \vec{W}, \vec{\sigma}, \vec{h})$ constructed, one can now compute the marginal distribution over the state space of the visible units only as $\tilde{p}(\vec{a}, \vec{b}, \vec{W}, \vec{\sigma})$ where $\tilde{p}(\vec{a}, \vec{b}, \vec{W}, \vec{\sigma}) = \sum_h P(\vec{a}, \vec{b}, \vec{W}, \vec{\sigma}, \vec{h})$. Now $\sqrt{\tilde{p}(\vec{a}, \vec{b}, \vec{W}, \vec{\sigma})}$ defines the amplitude of wave function over basis states of the visible units, i.e., $|\sigma_1\sigma_2\ldots \sigma_n\rangle$. The phase of each component of the wave function is now constructed classically using $(\vec{d}, \vec{f}, e)$ and tanh activation of neurons in the phase node as defined before in eq 5.

(iii) With the two information from step (ii), the target wave function can now be constructed classically as

$$\psi(\vec{X}) = \sum_{\sigma} \sqrt{\tilde{p}(\vec{a}, \vec{b}, \vec{W}, \vec{\sigma})}\ s(\vec{d}, \vec{f}, c, e, \vec{\sigma})|\sigma_1\sigma_2\ldots\sigma_n\rangle \tag{7}$$

(iv) With the wave function, the cost function in eq 2 can now be constructed classically with the $(\hat{H}, \hat{O}, \lambda)$ from the user where $\hat{H}$ and $\hat{O}$ are the Hamiltonian and filter operator for the system being investigated, respectively, and λ is the penalty parameter.

(v) The next step is to check for convergence criterion or maximum number of iterations (to be discussed later). If either of the criterion is satisfied, then the results are printed.

(vi) If either of the criterion from the previous step is not satisfied, then the parameter set $\vec{X}$ is updated using steepest-descent algorithm with a learning rate (set to 0.005 in all our calculations). The updated parameter vector $\vec{X}$ is fed into step (ii) for the next iteration of the algorithm. We have also used the ADAM optimizer,[87] but there is no significant change in convergence for the systems treated in this report. It must be emphasized that unlike in classical supervised deep-learning models the learning of our network does not require prior training against a preassigned labeled data set. The network learns the target eigenstate directly through minimization of the cost function (see eq 2) using the optimizer of choice (gradient descent in this case).

3.3. Resource Requirements. The power of an RBM ansatz even though underutilized in material science is beginning to gain attention in many areas of Fermionic and bosonic physics.[20,82,88] Using n visible neurons and m hidden neurons, a recent study[89] has shown explicitly how a shallow RBM ansatz ($\alpha = m/n = 1$) like ours already captures several orders of perturbation theory and is a good approximant to the exact state. Classically, constructing such a full RBM distribution will require tracking amplitudes from a 2^{m+n} dimensional state space and hence has exponential resource requirements in preparation. Ref 90 formalizes and consolidates this statement by proving that a polynomial-time algorithm for classically simulating or constructing a full RBM distribution is not only absent now but also is unlikely to exist even in the future as long as the polynomial hierarchy remains uncollapsed. However, such analysis does not preclude the existence of efficient quantum algorithms such as the one considered in this work. The quantum circuit in our algorithm (see Figure 2) uses $m + n + m \times n$ qubits only for constructing the state indicating an $O(m \times n)$ scaling in the qubit resource which if expressed in terms of hidden node density $\alpha = m/n$ is $O(\alpha n^2)$ The gate-set comprising single-qubit R_y gates also scales as $m + n$, one for each of the bias terms $(\vec{a}, \vec{b})$ of the visible and hidden node qubits. Each $C-C-R_y$ gate in the circuit mediate a single interaction term within the $\vec{W}$ matrix between a spin of the visible layer σ_i and a spin of the hidden layer h_j. Since there are $m \times n$ such terms, the number of $C-C-R_y$ gates are also $m \times n$, with the targets being each qubit in the ancilla register. Toggling between the various configurations of the control qubits (1 visible + 1 hidden) would require 6 R_x gates additionally in each $C-C-R_y$ (see Figure 2); hence, the total number of such R_x gates is $6mn$. This indicates

https://doi.org/10.1021/jacs.1c06246
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

the total gate requirements of our sampling circuit is also $O(m \times n)$ which is equivalent to $O(\alpha n^2)$. The number of variational parameters in our algorithm for amplitude encoding using RBM is $m + n$ for the biases of the two nodes and $m \times n$ for the $\vec{W}$ matrix. For the phase encoding, the variables are two n-dimensional vectors $(\vec{d}, \vec{f})$ and two scalars (c, e). Thus, the total number of variational parameters is $m \times n + m + 3n + 2 = \alpha n^2 + \alpha n + 3n + 2$ which is also quadratic. The upshot is then that our algorithm for an RBM ansatz uses $O(\alpha n^2)$ qubits (circuit width), gate-set (circuit depth), and variational parameters to encode any arbitrary quantum state of n qubits in a $d = 2^n$-dimensional Hilbert space. Removing redundancy in the global phase and normalization, a general such state would require $2(2^n - 1)$ parameters. One must know that in the RBM construction circuit no specific structure or sparsity has been assumed in the $\vec{W}$ matrix which if present may lower the requirements further. Quantum advantages have also been observed in supervised learning using the RBM distribution.[38] The study indicated that for the data set of size N, a quantum circuit with amplitude amplification reduces the complexity of the algorithm from the conventional $O(N)$ to $O(\sqrt{N})$, a quadratic boost. It must also be emphasized that all the results in this manuscript are primarily treated for the case of $\alpha = 1$ as that suffices for the description of the system we study. We show how the results change for changing hidden node density α in section 6 of the Supporting Information. Even though $\alpha = 1$ is good for systems in this report, for the case where the state is highly entangled, the user may be required to enhance the hidden node density as that increases the number of variational parameters and make the ansatz more expressive.[20] That may also be the case for molecular systems under geometric distortion wherein multireference correlation is important (we explore this point briefly in section 10 of the Supporting Information). In this work, all our results are compared against exact diagonalization as it affords the best accuracy in a given basis. The exact diagonalization results are obtained using "Numpy" package[91] in python 3.0 with LAPACK routine.

3.4. Implementation Methods. We implement the algorithm in three flavors of computation. The first flavor henceforth designated as **"RBM-cl"** involves implementing the entire gate set of the Gibbs sampling circuit on a classical computer. This computation returns to us the exact state after the termination of the circuit. The second flavor is henceforth designated as **"RBM-qasm"**. This has been implemented by simulating the Gibbs sampling circuit using Qiskit which stands for IBM's Quantum Information Software Kit (Qiskit).[61] We specifically used the *qasm_simulator* at Aer provider (hence the name **RBM-qasm**) which is a quantum computer simulator and hence can mimic calculations performed on a noisy-intermediate scale quantum computing device even using a classical computer with options to incorporate customizable noise models. Unlike in **"RBM-cl"** where the exact state is returned, in **"RBM-qasm"**, the Gibbs sampling circuit in Figure 2 is interrogated multiple times to build measurement statistics. From the observed bit-strings, the measurement probabilities $P(\vec{a}, \vec{b}, \vec{W}, \vec{\sigma}, \vec{h})$ are computed; hence, the results are subjected to statistical fluctuations due to finite sampling errors. No noise model was used during the simulation in **"RBM-qasm"**. Finally, to see the effect of noise, we also investigated the performance of our algorithm on real IBM-Q quantum computers using the Qiskit interface. We used IBM-Q Sydney[92] and IBM-Q Toronto[93] interchangeably, both of which are 27 qubit machines and hence suitable for our case studies. Calculations of this flavor are henceforth referred to as **"RBM-IBMQ"**. To reduce the effect of noise on the sampling probabilities we employ measurement error mitigation (MEM)[94] directly implementable on Qiskit. We show in this report that MEM alone guarantees smooth and clear self-convergence in training (see section 4 in the Supporting Information). The final accuracy of the results is affected by both MEM and warm-starting. We have seen without warm-starting convergence can not only be slow but sometimes the network can even be trapped in a local minima. It is in general difficult to assess apriori when the need for warm-starting can arise without a knowledge of the optimization surface as the objective function being optimized for the amplitude and the phase are nonconvex in the arguments (see eqs 4 and 5). It has been noted that the algorithm converges better without the need for warm-starting near optima (symmetry points for the system being treated in this report as discussed later). For the **"RBM-qasm"** and **"RBM-cl"** simulations, the maximum number of iterations within which well-converged results to be discussed below were obtained is $\leq$30 000 either with a warm-start or randomly initialized parameter set depending on the case. The **"RBM-IBMQ"** simulations were performed by breaking into two sessions/runs with the maximum iteration $\leq$700 for each session to reduce the job queue. Normally most calculations converged well before 700 iterations were reached within the first run as warm-starting and MEM has been used as described above. For the few that did not, the final parameter set of the first run is punched for initializing the second session to ensure one continuous run. It must be emphasized that the entire code-base for training the network is home-built in Python 3.0 using standard packages like Numpy.[91] As mentioned before, we have extensively used Qiskit though as an interface to communicate with the IBMQ hardware and with *qasm_simulator*.

4. RESULTS AND DISCUSSION

As a test of our method, we target state filtration of energy eigenstates of two well-established TMDCs: monolayer molybdenum disulfide (MoS_2) and monolayer tungsten disulfide (WS_2). Monolayer TMDCs have so far eluded attention in quantum simulations even though it is imperative to study their electronic structures to understand novel properties[51,96] like high carrier mobility, high photoluminescence due to the direct band gap, lack of inversion symmetry leading to large spin−orbit coupling, intravalley transport, and so on. Indeed, such features have made them attractive candidates for applications in field-effect transistors,[97] supercapacitors,[98] spintronics,[99] optoelectronics,[100,101] and "valleytronics".[102] We first show how the entire conduction band (CB) in such materials can be simulated using an appropriate choice of operator $\hat{O}$ as the ground-state projector as discussed before and then later show how to "sieve" eigenstates based on angular momentum symmetry. In all cases, we implement our algorithm on three flavors of RBM calculations as discussed: RBM-cl, RBM-qasm, and RBM-IBMQ.

4.1. Filter for Target Excited States: Simulation of Low-Energy Bands in MoS_2 and WS_2 and the Effect of Spin−Orbit Coupling. The geometrical structure of monolayer TMDCs like MoS_2 or WS_2 indicates the presence of a trigonal prismatic real space unit cell[51] with D_{3h} point group symmetry as shown in Figure 3. The transition metal is

https://doi.org/10.1021/jacs.1c06246
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

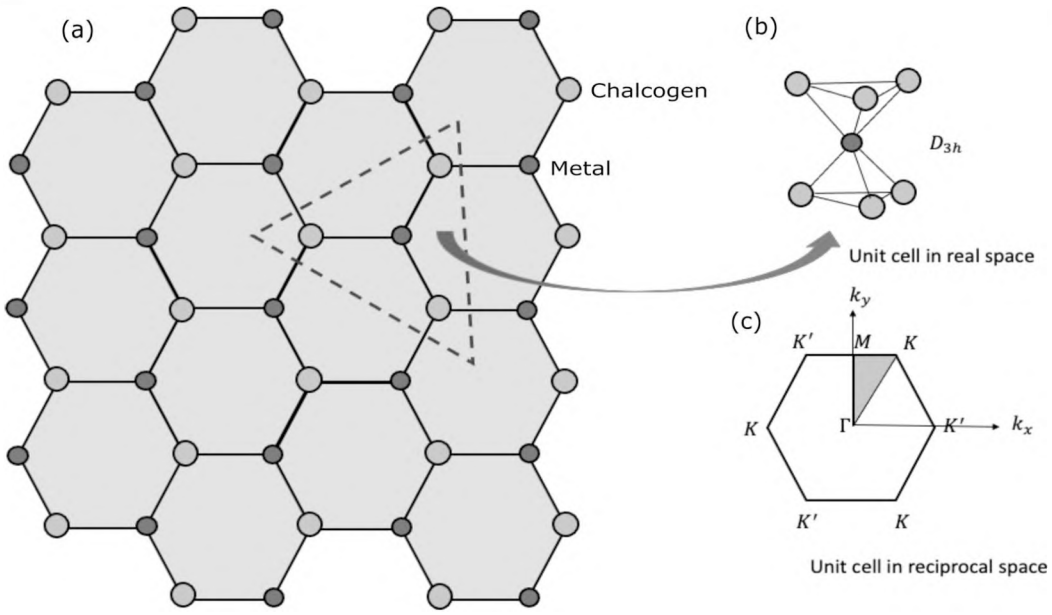


Figure 3. (a) Top view of the TMDC monolayer as studied in this report. The orange atoms are a chalcogen whereas the blue atoms are the metal center. (b) Real-space trigonal prismatic unit cell highlighting D_{3h} symmetry. This shows that in the TMDC monolayer unlike in graphene, the constituent atoms have a noncoplanar arrangement. (c) Unit cell in reciprocal space showing the important symmetry points (Γ, K, M, K'). We shall investigate the energy and other properties within the sector marked in green following the usual $\Gamma-K-M-\Gamma$ path as in ref 103. The coordinates of the symmetry points as (k_x, k_y) are $\Gamma = (0, 0)$, $K = \left(\frac{4\pi}{3a_0}, 0\right)$, and $M = \left(\frac{\pi}{a_0}, \frac{\pi}{\sqrt{3}a_0}\right)$ where a_0 is the metal−chalcogen bond length. For systems studied in this report the metal center is Mo, W, and the chalcogen is S.

at the center and the sulfur atoms are at the six corners of the triangular prism (see Figure 3b). Consequently, the best orbital decomposition to evaluate the band structure of this periodic material should involve not only the s, p, and d orbitals of the central metal atom but also of the surrounding sulfur atoms. Indeed several reports exist which treats the electronic structure of such materials using a tight-binding description obtainable from a 5, 7, or an 11 band model using a varying degree of inclusion of the orbital set of the metal and the chalcogen.[104−107] However, recently a three-band parametrization has been demonstrated to yield remarkable accuracy in energy over the entire Brillouin zone.[103] A tight-binding Hamiltonian in this description is obtained by fitting the energy curves against DFT calculations (with GGA and LDA functionals) only employing the d_{z^2}, d_{xy} and $d_{x^2-y^2}$ orbitals of the metal center.[103] This choice is based on the fact that for trigonal prismatic coordination, the d orbital set of the metal splits into three groups: A_1' containing d_{z^2} orbital only, E' containing d_{xy} and $d_{x^2-y^2}$, and E'' containing d_{xz} and d_{yz} orbitals. However, reflection symmetry of D_{3h} restricts intercoupling between the orbitals of E'' set with the remaining two groups. Indeed E'' contributes exclusively to higher energy bands and has no role to play in the low-energy physics of the valence and conduction band which is considered in this work. The absence of chalcogen p orbitals is definitely an approximation albeit a good one as seen from ref 103. We shall return to this point shortly.

We henceforth use a tight-binding model comprising of third-nearest neighbor (TNN) metal−metal hopping[103] of the aforesaid three-band Hamiltonian for all our calculations. The parameters of the model are obtained from the more accurate GGA calculation set.[103] Section 3 of the Supporting Information gives the details of the Hamiltonian and parameters for completeness and brevity. Our working Hamiltonian, for both the systems are thus a 3×3 Hermitian matrix. For qubitization we convert it into a 4×4 Hermitian matrix by padding an additional 1×1 block with a diagonal entry chosen to be greater than or equal to the spectral range of $H_{3\times3}$ as that would keep the low-lying eigenvalue structure of the resultant matrix undisturbed for the training to successfully proceed. Thus, for both the systems, our neural network comprises of a visible node with 2 neurons to encode the state, two hidden neurons, and additional 2 neurons for the phase node. For the Gibbs sampling circuit in Figure 2, we thus need 2 qubits to represent the entire visible layer and 2 qubits for the hidden layer. In addition, we need 4 ancilla qubits to serve as targets for $(C-C-R_y)$ rotation thereby requiring 8 qubits in total. For the circuit in Figure 2, we use 4 single-qubit rotation gates (R_y), 4 controlled−controlled rotation gates $(C-C-R_y)$, and also 24 bit-flip (X) gates. The optimization in each case starts from a randomly initialized parameter set. In case if the accuracy is poor, then we restart the algorithm by feeding the initial parameter from the results of a nearby converged k-point as a warm start. We see the results are in excellent agreement with the exact diagonalization when a such a warm start is employed along with MEM as described before. For IBMQ implementation, we have used "IBM-Sydney" and "IBM-Toronto" both of which are 27 qubit machines. To reduce the operational time on the actual quantum device for the job queue and isolate the effect of gate-infidelity, IBMQ simulations for each k-point were often warm-started with an initial parameter set obtained from the initial parameters of the *qasm* simulation of a nearby but nonidentical k-point.

The results from the algorithm using the cost function in eq 2 is displayed in Figure 4 for MoS_2 and Figure 5 for WS_2. In Figure 4a, we have overlaid the energies obtained from our algorithm as a function of the wave-vector index sampled from the Brillouin zone following the usual $\Gamma-K-M-\Gamma$ path (see

https://doi.org/10.1021/jacs.1c06246
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

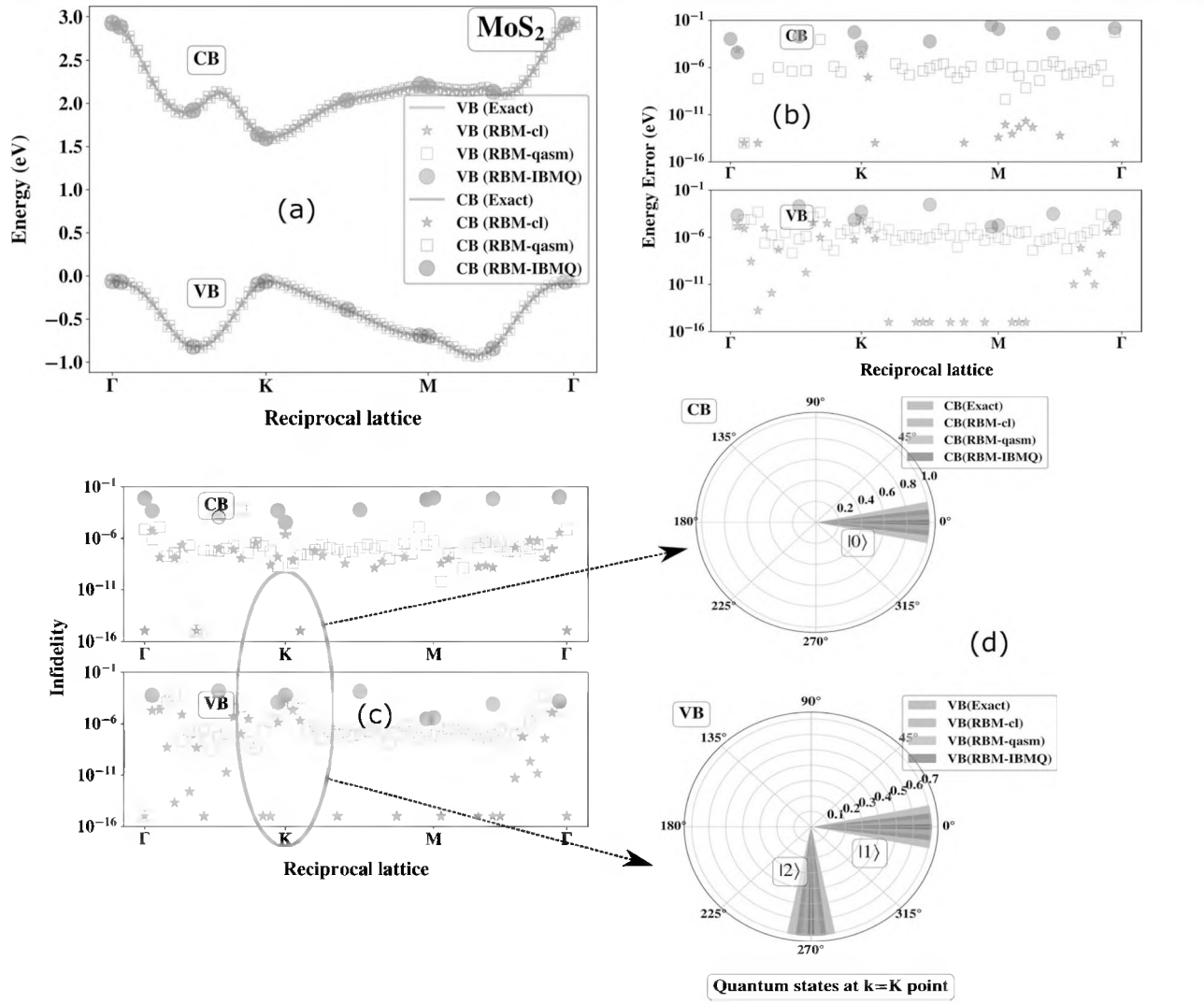


Figure 4. (a) Valence (VB) and conduction band (CB) of MoS_2 calculated using all flavors of RBM and overlaid against exact diagonalization. The valence band is simulated using $\lambda = 0$ in eq 2 and the conduction band using ($O = |v_0\rangle\langle v_0|$, $\omega = 0$, $\lambda = 5$) in eq 2 where $|v_0\rangle$ is the valence band state at each k-point. For IBMQ implementations we used "IBM-Sydney" and "IBM-Toronto". All parameters are randomly initialized (see Figure 2) or warm-started with the initial guess of a converged nearby k-point. (b) Corresponding energy errors from (a) in eV. (c) Corresponding state infidelities ($1 - Fid$) where $Fid = |\langle\Psi_{RBM}|\Psi_{Exact}\rangle|^2$. (d) Orbital decomposition of the states at K-point where $|0\rangle = d_{z^2}$, $|1\rangle = d_{xy}$, and $|2\rangle = d_{x^2-y^2}$. The states from RBM calculations match well with those from exact diagonalization in phase and amplitude. The width for each bar is set differently for visual clarity.

Figure 3c). The result for the valence band (VB) is denoted in blue and is obtained by setting $\lambda = 0$ in eq 2 which corresponds to the usual variational optimization to obtain the ground state at each k-point. The results for the CB are shown in orange in Figure 4a. They are thereafter computed as a separate set of calculations using $O = |v_0\rangle\langle v_0|$ and $\omega = 0$ in the cost function in eq 2 where the corresponding ground state in the VB is denoted as $|v_0\rangle$. The penalty parameter is $\lambda = 5$. The cost-function now samples a state orthogonal to ground state (null space of the projector $|v_0\rangle\langle v_0|$) for each of the k-points. The minimum energy criterion imposed by the first term in the cost function in eq 2 guarantees obtaining the next higher excited state which happens to be the state space in the conduction band.

We see for all flavors of our algorithm (RBM-cl, RBM-qasm, and RBM-IBMQ) that the simulated energy values for both the valence and the conduction band are in good agreement with the ones obtained from exact diagonalization. The corresponding errors in energy are displayed in Figure 4b and are usually $\leq 10^{-4}$ eV for RBM-cl and RBM-qasm which are noiseless pristine implementations, but the error is around 10^{-2}–10^{-4} eV for the VB and the conduction band for RBM-IBMQ, indicating the worsening of performance due to faulty gate implementations in the Gibbs sampling circuit. Figure 4c plots the state infidelities, i.e., $1 - Fid$ where $Fid = |\langle\Psi_{RBM}|\Psi_{Exact}\rangle|^2$. We see that the infidelities are also quite small for each band with the performance worsened only in the IBMQ variant of the RBM implementation.

Like Figure 4a, Figure 5a displays the band structure of WS_2 wherein the energies for both the valence and conduction band are overlaid against the energy values obtained from exact diagonalization. All three flavors of RBM implementation yield reasonably accurate results as in the case for Figure 4a. Figure 5b,c displays the energy error and the state infidelities of the state obtained from the RBM calculations against exact

https://doi.org/10.1021/jacs.1c06246
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

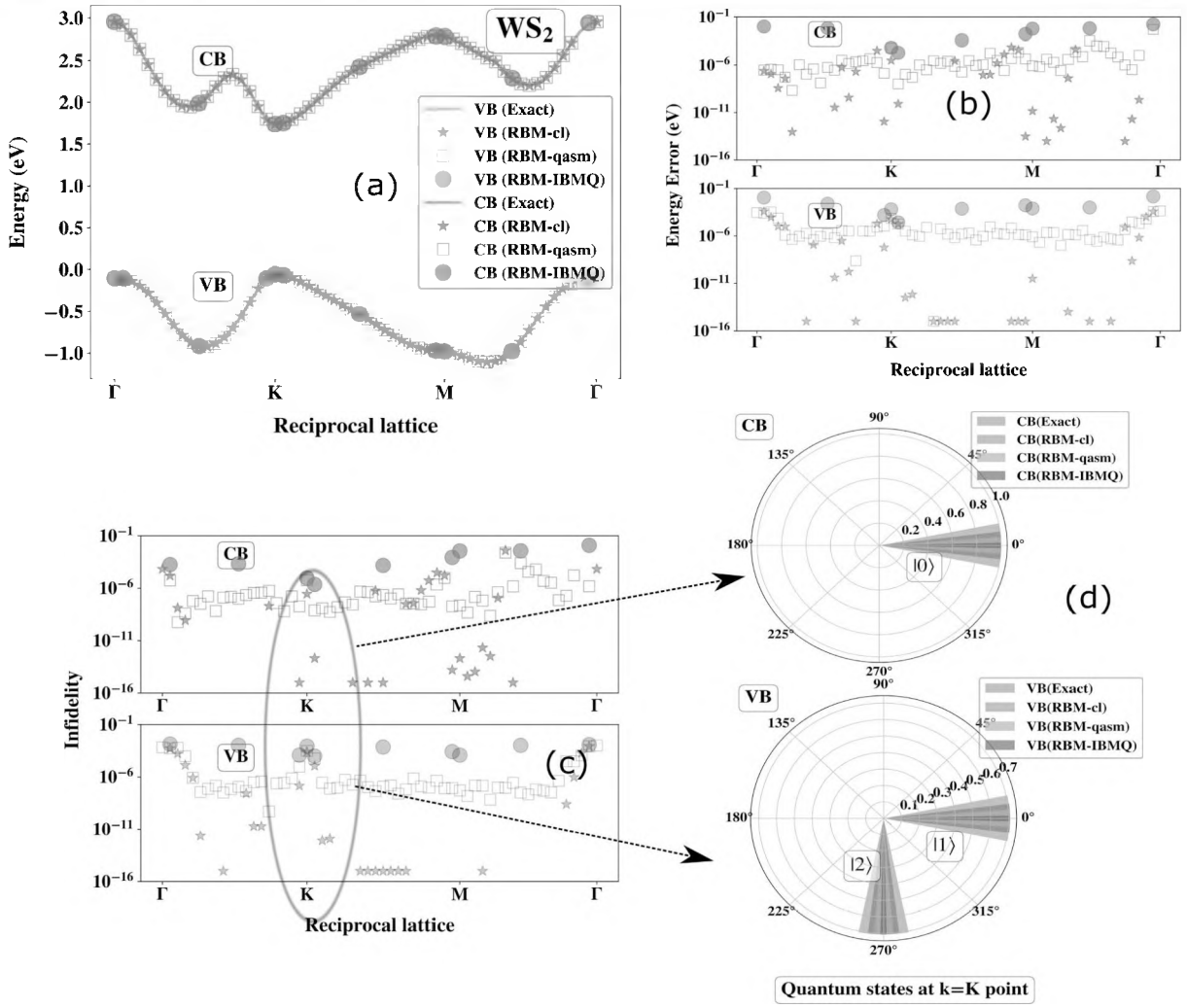


Figure 5. (a) VB and CB of WS_2 calculated using all flavors of RBM and overlaid against exact diagonalization. The valence band is simulated using $\lambda = 0$ in eq 2 and the conduction band using ($O = |v_0\rangle\langle v_0|$, $\omega = 0$, $\lambda = 5$) in eq 2 where $|v_0\rangle$ is the valence band state at each k-point. For IBMQ implementations we used "IBM-Sydney" and "IBM-Toronto". All parameters are randomly initialized (see Figure 2) or warm-started with the initial guess of a converged nearby k-point. (b) Corresponding energy errors from (a) in eV. (c) Corresponding state infidelities ($1 - Fid$) where $Fid = |\langle\Psi_{RBM}|\Psi_{Exact}\rangle|^2$. (d) Orbital decomposition of the states at K-point where $|0\rangle = d_{z^2}$, $|1\rangle = d_{xy}$, and $|2\rangle = d_{x^2-y^2}$. The states from RBM calculations match well with those from exact diagonalization in phase and amplitude. The width for each bar is set differently for visual clarity.

diagonalization. The error ranges in each case are similar to what has been discussed for MoS_2.

Figures 4d and 5d display the orbital decomposition of the states in the conduction and valence band at the most important symmetry point, i.e., the K-point. In our calculations, qubits $|0\rangle = d_{z^2}$, $|1\rangle = d_{xy}$, and $|2\rangle = d_{x^2-y^2}$ where {0, 1, 2} are the integer equivalents of the two-qubit bit strings encoding the neurons of the visible node. We see from our calculations however that the exact state generated from the model lines up correctly against the RBM states in both amplitude and phase. While the state of the conduction band at K-point is exclusively populated by d_{z^2}, that in the valence band is a superposition of d_{xy} and $d_{x^2-y^2}$ with a phase shift of $3\pi/2$. This is consistent with the orbital decomposition given in Figure 2 of ref 103 and is partly the reason given by the authors to use this three orbitals for generating the tight-binding Hamiltonian as the model yields correct state description near the band gap. However, as is clear from Figure 2 of ref 103, the orbital composition of the states at the Γ- and M-points has contribution from the p-orbitals of S and s-orbitals of both the metal and the S atoms. This makes the three-band model an approximation for the exact character of the states even though it can replicate the energy very well throughout the Brillouin zone.

We further concentrate in this report on describing the low-energy physics near the K or K' valley for which, as mentioned before, the state description of the three-band model suffices. We construct the Hamiltonian[108−111] near the K-valley in the basis of the states of the conduction band, i.e., $|d_{z^2}\rangle$ (see Figures 4d and 5d) and that of the valence band, i.e., $\frac{1}{\sqrt{2}}(|d_{x^2-y^2}\rangle + i|d_{xy}\rangle)$ (see Figures 4d and 5d). The states at the K' valley are related to those at the K valley due to time-reversal symmetry[110,112] and hence are ignored from further discussion. The Hamiltonian is

https://doi.org/10.1021/jacs.1c06246
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

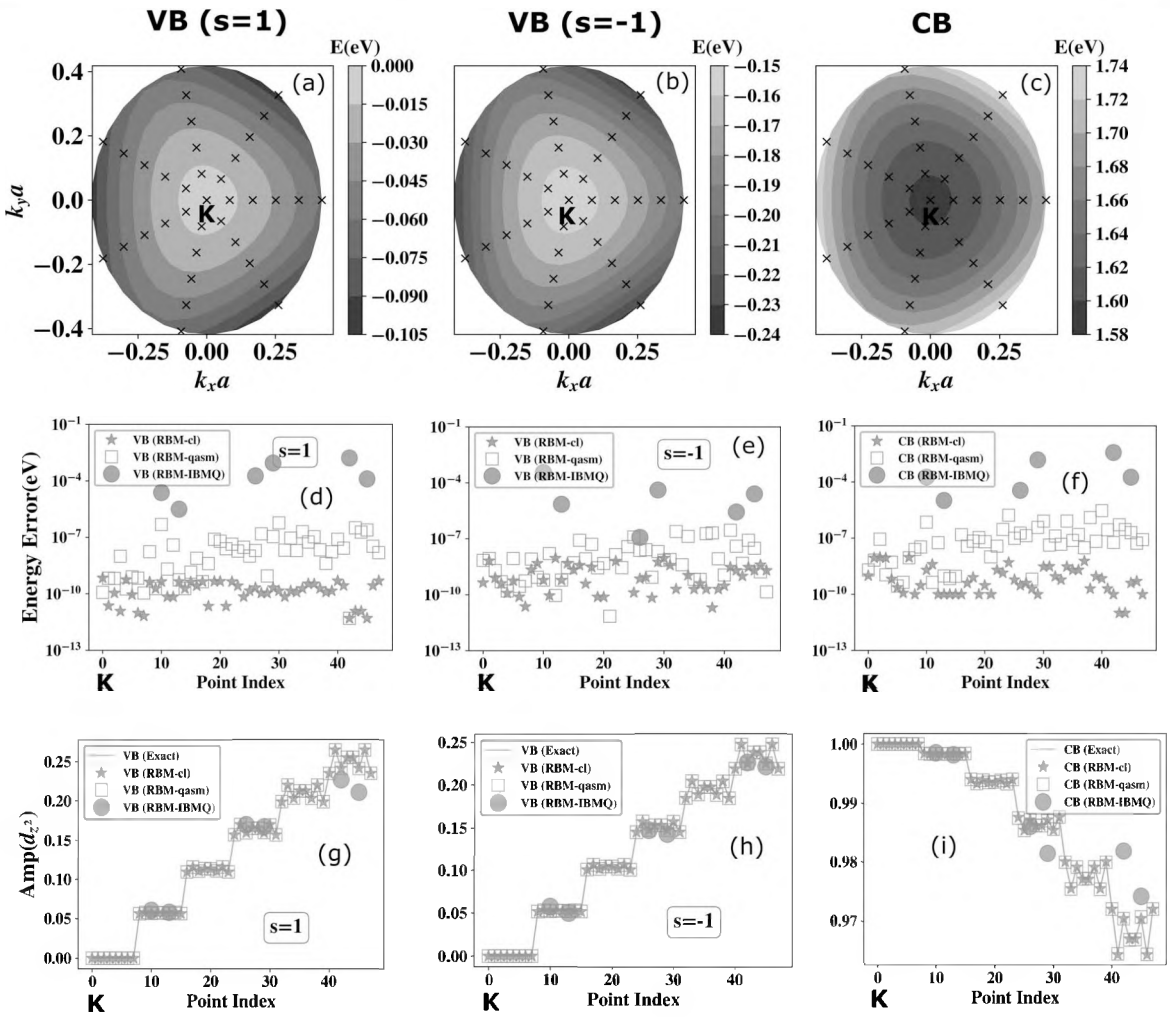


Figure 6. (a) Exact energy contours in the VB for $s = 1$ within the three-band approximation for the Hamiltonian in eq 8 as a function of (k_x, k_y) near the K-point in MoS_2. (b) Same as in (a) but for $s = -1$. (c) Same as in (a) for the CB. The crosses in (a), (b), and (c) denote the (k_x, k_y) pair wherein calculations for all three flavors of RBM have been executed. (d) Energy errors in eV from three flavors of RBM calculations for points denoted as cross in (a) for the valence band ($s = 1$) case computed using $\lambda = 0$ in eq 2 in MoS_2. The x-axis is a flattened point index with (k_x, k_y) pairs marked as crosses in (a) mapped to integers such that the origin is at the K-point. From the K-point, the flattened point index scale moves spirally outward grouping all (k_x, k_y) pairs satisfying $|k| = \sqrt{k_x^2 + k_y^2}$ as consecutive integers and then proceeding to the next $|k|$ (e) Same as in d) but with points denoted in b) as crosses for other valence band with $s = -1$. (f) Same as in (d) but for points denoted in (c) as crosses for the conduction band computed with $(\lambda = 5, \omega = 0, \hat{O} = |\nu_0\rangle\langle\nu_0|)$ in eq 2. (g) Amplitude for the occupancy of d_{z^2} orbital on the metal for states computed at (k_x, k_y) pairs near the K-point from all three flavors of RBM as well as the exact states in valence band ($s = 1$) for MoS_2. The amplitude of states with the same $|k| = \sqrt{k_x^2 + k_y^2}$ appear bunched together as "steps" due to the flattened point-index scale used. Near the K-point, the amplitude is the same for all such pairs within a given step due to isotropy of the energy surface. However, away from the K-point deviations appear due to trigonal warping owing to the D_{3h} symmetry of the unit cells in TMDCs. The states from all flavors of RBM can resolve the influence of warping accurately with the performance worsened for the noisy variant. (h) Same as in (g) for valence band ($s = -1$). (i) Same as in (g) for conduction band.

$$H = \left(\frac{E_c}{2} + \frac{\lambda_c s}{2}\right)(\hat{I} - \hat{\sigma_z}) + \left(\frac{E_v}{2} + \frac{\lambda_v s}{2}\right)(\hat{I} + \hat{\sigma_z}) + \gamma k_x \hat{\sigma_x} + \gamma k_y \hat{\sigma_y} + \frac{\alpha}{2}(k_x^2 + k_y^2)(\hat{I} + \hat{\sigma_z}) + \frac{\beta}{2}(k_x^2 + k_y^2)(\hat{I} - \hat{\sigma_z}) + \kappa_{TW}(k_x + ik_y)^2(\hat{\sigma_x} + i\hat{\sigma_y}) + \kappa_{TW}(k_x - ik_y)^2(\hat{\sigma_x} - i\hat{\sigma_y}) \quad (8)$$

A effective description such as eq 8 is often referred in literature as the two-band $k{\cdot}p$ model constructed using Lowdin Partitioning.[103,109] The first two terms in eq 8 is the massive term required to create the band gap (Δ) in the material at the K-point. These terms are absent in graphene. In most reports this term is written as $\Delta/2\sigma_z$ with a symmetrically located origin but we choose to use the E_c and E_v values obtained from our calculations in Figures 4 and 5. The additional summands in each of the first two terms (λ_v, λ_c) refer to band-splitting at the K-point due to spin−orbit coupling (SOC). In the three-band basis, SOC is entirely due to the L_z operator (more on this in the next section) contribution of which in the chosen basis of can be effectively modeled as the first two terms.[108,111,113] Unlike the Bloch state in the conduction band, the valence band is exclusively dominated by metal orbitals $|d_{x^2-y^2}\rangle$ and $|d_{xy}\rangle$ with nonzero angular momentum

https://doi.org/10.1021/jacs.1c06246
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

leading to strong splitting.[103] The spin−orbit splitting in the conduction band is weak[103,108,112] and below the resolvable limit of NISQ devices and hence has been ignored herein, i.e., $\lambda_c = 0$. The parameter $s \in \{1, -1\}$ is the spin index and labels the SOC split valence bands. The third through sixth terms are the linear and quadratic extrapolations away from the K-point and yield a spherically isotropic band surface. The seventh and eighth terms (parametrized by κ_{TW}) break the isotropy and lead to the well-known effect of trigonal warping (TW). The warped band surfaces in these materials are a consequence of the presence of a perpendicular C_3 axis due to the D_{3h} symmetry of the associated real-space unit cells (see Figure 3b). Further terms in ref 109 which removes anisotropy between valence and conduction band are ignored due to their small unresolvable contributions.

Since the Hamiltonian in eq 8 is 2×2, we require a single visible neuron to encode the eigenstates, a single hidden neuron consistent with $\alpha = 1$, and 1 additional ancillary qubit. The number of single-qubit R_y gates is 2, and the number of $C-C-R_y$ gates is 1 and 6 R_x gates. Calculations are performed using $\lambda = 0$ in eq 2 for the two SOC split valence bands with $s = \pm 1$ and $(\hat{O} = |\nu_0\rangle\langle\nu_0|, \lambda = 5, \omega = 0)$ for the conduction band. For NISQ devices we use "IBM-Sydney" and "IBM-Toronto" interchangeably as before. All calculations are performed for (k_x, k_y) pairs centered at the K-point and with a cutoff $|k|$ of $0.1K$ point to probe the low-energy regime. Since the (k_x, k_y) pairs are near a symmetry point (K-point), warm-starting was rarely observed to be required in RBM-cl and RBM-qasm but has been occasionally used in RBM-IBMQ for hastening convergence and reducing job queue. Each point on RBM-IBMQ are performed within a single run with MEM as before for smooth self-convergence and consistency with other results. Parameters for warping are obtained from ref 109.

In Figure 6a−c, we plot the exact 2D band surfaces obtained from eq 8 for the two SOC split valence bands ($s = \pm 1$) and the conduction band. The crosses in each plot refer to the (k_x, k_y) pairs wherein all flavors of RBM calculations have been performed. The results of such RBM calculations for each such pair are displayed as energy errors (eV) in Figure 6d−f. The x-axis in each such plot is a flattened point index mapping (k_x, k_y) pairs to integers by starting from pairs closest to the K-valley at the origin and proceeding spirally outward. In other words, for a given $|k|$ the flattened point index groups all (k_x, k_y) pairs satisfying $|k| = \sqrt{k_x^2 + k_y^2}$ as consecutive integers and then proceeds to the next $|k|$. We see that the energy error in each case is low for the RBM-cl and RBM-qasm variant ($\lesssim 10^{-4}$ eV) for all three bands and $\lesssim 10^{-2}$ eV for the IBMQ variant. Thus, given the energy scale and extent of the splitting in the valence bands ($s = \pm 1$) in Figure 6a,b and the scale of the energy errors in Figure 6d,e, it suffices to say that the performance of our algorithm is good enough to resolve band splitting due to features like spin−orbit coupling. To study the effect of warping parameters in eq 8 in the state, we plot in Figure 6g−i the amplitude of the corresponding states in the basis of $|d_{z^2}\rangle$ for the two SOC split valence bands ($s = \pm 1$) and the conduction band. The x-axis in each case is the flattened pair index as in Figure 6d−f. At the K-point (origin), the conduction band is exclusively populated by $|d_{z^2}\rangle$ as discussed before, but the reverse is true for the valence bands. In each of the plots in Figure 6g−i, all (k_x, k_y) pairs which satisfy $|k| = \sqrt{k_x^2 + k_y^2}$ are bunched together as "steps" due to the flattened point index scale chosen. We see that near the K-point wherein the effect of warping is not prominent, all such points within a given "step" (same $|k|$) share the same amplitude. However, away from the K-point deviation starts to become predominant. The amplitudes computed from the states of all three variants of RBM calculations line up well against the exact curve with the IBMQ variant showing some deviations albeit small considering the y-scale in these plots. Our algorithm thus can successfully resolve finer features like trigonal warping too in these Bloch states. A similar panel for WS_2 is presented in section 7 of the Supporting Information. Accurate computation of such Bloch states with these finer features preserved is necessary as momentum matrix elements between these states become important in simulating important properties of materials like optical conductivity,[114,115] electrical and thermal conductivity,[116] and so on.

4.2. Filter for Arbitrary States Using Symmetry Operators. In this section, we shall use the same set of TMDCs discussed above to explore how one can sieve arbitrary states based on symmetry constraints. To demonstrate the point we use orbital angular momentum symmetry. The L_z operator in the three-band approximation commutes with the Hamiltonian[103] in absence of spin−orbit coupling as has been considered in this work. The operators L_x, L_y are essentially null matrices in the three-band basis of $\{d_{z^2}, d_{xy}, d_{x^2-y^2}\}$ as mentioned in.[103] Hence L^2 enjoys exclusive contribution from L_z and is a symmetry operator in the system. For computation, we use the Hamiltonian of the system at the K-point because the three-band approximation as discussed before is extremely accurate therein.

The complete set of eigenvalues and eigenstates of L_z and hence of L^2 operator is given in section 8 of the Supporting Information. From the knowledge of the spectrum of L^2 operator, we see that it has two distinct eigenvalues which are $\{0, 4\}$ in atomic units. One of the eigenvectors of the doubly degenerate eigenspace with eigenvalue 4 is the state in the valence band, and the other is a higher energy excited state above the conduction band (not shown in Figures 4 or 5). Both these states are exclusively made from the contribution of $\{d_{xy}, d_{x^2-y^2}\}$ as seen from the state decomposition in section 8 of the Supporting Information. The sector with eigenvalue 0 has single-fold degeneracy and is made from the excited state in the conduction band. As discussed before in Figures 4d and 5d (and in section 8 of the Supporting Information) this state is exclusively made from the contribution of the d_{z^2} which explains the absence of z-component angular momentum. We would thus expect that if we choose $\hat{O} = L^2$ and $\omega = \{0, 4\}$ in eq 2 for training the network then we should yield the excited state in the conduction band for $\omega = 0$ and should yield the ground state in the valence band for $\omega = 4$ as that is of lower energy (in compliance with the first term in eq 2) than the other degenerate eigenstate.

The qubit and gate resource requirements of this simulation are exactly the same as discussed in section 4.1 with 2 visible node neurons and 2 hidden node neurons for each of the two systems, MoS_2 and WS_2. The Gibbs sampling circuit in Figure 2 would need a total of 8 qubits as before (2 for visible node + 2 for hidden node + 4 ancillary qubits). The gate requirements for the circuit to reproduce the amplitude are thus 4 single qubit rotation gates (R_y), 4 controlled−controlled rotation gates ($C-C-R_y$), and 24 bit-flip (X) gates. We start the optimization with randomly initialized parameters.

https://doi.org/10.1021/jacs.1c06246
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

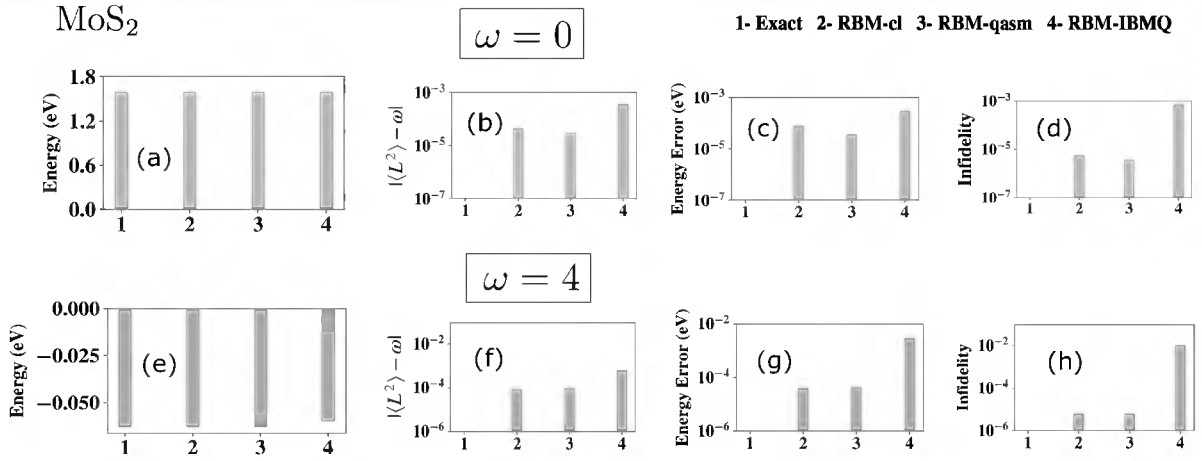


Figure 7. (a) Energy comparison between exact (1), RBM-cl (2), RBM-qasm (3), and RBM-IBMQ (4) for computation with $\hat{O} = L^2$ and eigenvalue $\omega = 0.0$ a.u. in eq 2. The exact energy is 1.5950 eV and is the conduction band energy at K-point in MoS_2 shown in Figure 4. (b) Constraint violation error $|\langle L^2\rangle - \omega|$ of the state obtained from different flavors of RBM and the desired value ω. (c) Energy error in eV from (a) of the states obtained from RBM. (d) State infidelities ($1 - Fid$ where $Fid = |\langle\Psi_{RBM}|\Psi_{Exact}\rangle|^2$) obtained from RBM and the exact one (e–h) corresponds to an equivalent set of plots as in (a–d) just described but with the other eigenspace of L^2 with eigenvalue $\omega = 4$ a.u. The exact energy here is the valence band energy at K-point for MoS_2 shown in Figure 4 and is −0.0629 eV.

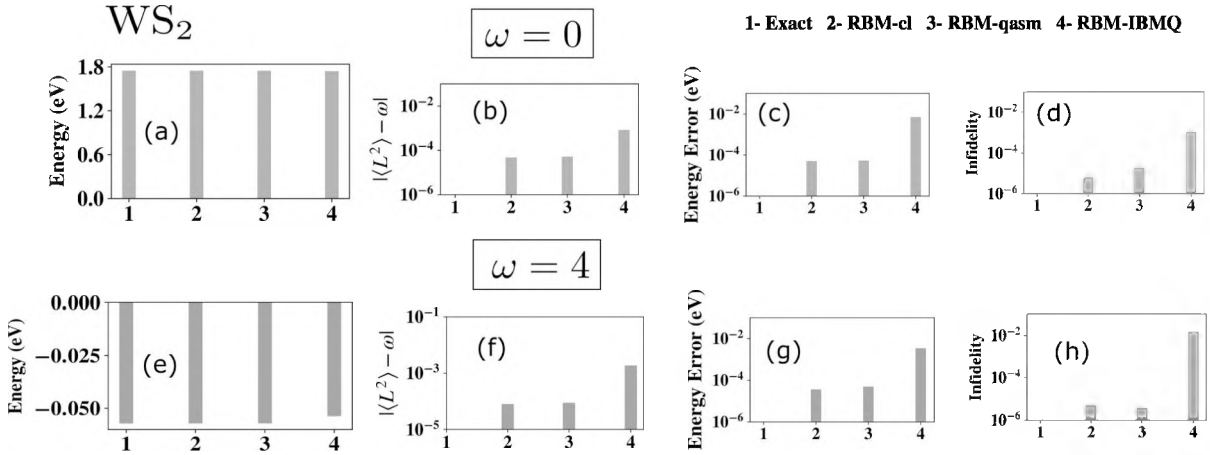


Figure 8. (a) Energy comparison between exact (1), RBM-cl (2), RBM-qasm (3), and RBM-IBMQ (4) for computation with $\hat{O} = L^2$ and eigenvalue $\omega = 0.0$ a.u. in eq 2. The exact energy is 1.749 eV and is the conduction band energy at K-point in WS_2 shown in Figure 5. (b) Constraint violation error $|\langle L^2\rangle - \omega|$ of the state obtained from different flavors of RBM and the desired value ω. (c) Energy error in eV from (a) of the states obtained from RBM. (d) State infidelities ($1 - Fid$ where $Fid = |\langle\Psi_{RBM}|\Psi_{Exact}\rangle|^2$) obtained from RBM and the exact one (e–h) corresponds to an equivalent set of plots as in (a–d) just described but with the other eigenspace of L^2 with eigenvalue $\omega = 4$ a.u. The exact energy here is the valence band energy at K-point for WS_2 shown in Figure 5 and is −0.0572 eV.

In Figure 7, we display the results of our simulation. Like before, the results from all three flavors of RBM (marked as 2 = RBM-cl, 3 = RBM-qasm and 4 = RBM-IBMQ) are compared against the exact expected state (marked as 1 = Exact). In Figure 7a, the results of energy (in eV) from the three RBM simulations and the exact one are displayed for the eigenvalue sector $\omega = 0$ a.u. This happens to be the CB energy in Figure 4a. We find an extremely good agreement for all flavors of RBM with the exact value. The corresponding energy error is displayed in Figure 7c and is in the range of 10^{-5}–10^{-4} eV for RBM-cl and RBM-qasm, but it is within 10^{-4}–10^{-3} eV for the RBM-IBMQ variant. Figure 7b displays the constraint violation error, i.e., how much the state encoded in the neural network after training has an $\langle L^2\rangle$ equal to the target value of ω (in this case $\omega = 0$ a.u.). We see that the violations are quite small for the noiseless implementations. Even for implementation on actual NISQ devices of IBM-Q, it is close to 10^{-3} a.u.

Figure 7c displays the energy error, and Figure 7d displays the state infidelity error ($1 - Fid$ where $Fid = |\langle\Psi_{RBM}|\Psi_{Exact}\rangle|^2$). We see that for all flavors of RBM implementation the infidelities are quite small with the performance worsened for implementation on the actual IBM-Q device. Figure 7e–h corresponds to similar plots as discussed above but this time in the other eigenvalue sector with $\omega = 4$ a.u. We again see that the energy values (in eV) in Figure 7e matches with the exact for all flavors of RBM-implementation. This state happens to be the ground state in the VB shown in Figure 4a. The corresponding energy errors shown in Figure 7g are like in the previous case ($\omega = 0$ a.u.) low for RBM-cl and RBM-qasm but in the range of 10^{-3}–10^{-2} eV for RBM-IBMQ. Similar analysis as in the case of $\omega = 0$ a.u. can also be made for the constraint violation error in Figure 7f and the state infidelity in Figure 7h. Both of these have low errors with the respective ranges as displayed.

https://doi.org/10.1021/jacs.1c06246
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

Figure 8 shows a similar plot for the other system studied, WS_2. Just as before we display the results for $\omega = 0$ a.u. in Figure 8a–d and for $\omega = 4$ a.u. in Figure 8e–h. Figure 8a shows the energy match between the RBM implementations and the exact value for $\omega = 0$ a.u., and Figure 8e shows the same for $\omega = 4$ a.u. The former is equal to the state in the conduction band at *K*-point (see Figure 5a) and the latter is the corresponding state in the valence band (see Figure 5a). We see good agreement for all RBM variants and the exact expected value. The corresponding energy errors are low (see Figure 8c,g) with the range for IBMQ implementation being 10^{-3}–10^{-2} eV and even lesser for the pristine implementations. The respective constraint violation errors are displayed in Figure 8b,f and are small too as seen from the scale. A similar statement can also be made for the state infidelity displayed in Figure 8d,h. We have seen that in both the systems, MoS_2 and WS_2, the state infidelity and energy errors are higher in the $\omega = 4$ a.u. eigensector than in $\omega = 0$ a.u. eigensector in the IBMQ implementation especially. In fact, the relative energy errors for the said sector are close to 5% for RBM-IBMQ. However, the corresponding errors (both relative and absolute) are low for the noiseless implementation (RBM-cl and RBM-qasm) indicating that the higher percent error is attributable to the imperfect implementation of gates in the Gibbs sampling circuit in an IBM-Q machine and hence can be mitigated with future quantum computing devices with better gate fidelities and error-correction schemes.

5. CONCLUSION

In this study, we have demonstrated an algorithm which can filter arbitrary energy eigenstates in 2D materials like TMDCs using a quantum circuit with quadratic resources. We provided an original proof of feasibility for our cost function employed for the constrained optimization. We also proved a generic lower bound for the successful sampling of our quantum circuit from which previously known bounds can be extracted. Our circuit trains a three-layered neural network that encodes the desired state using an RBM ansatz for the probability density. As an illustration, we were able to filter energy eigenstates in the conduction band of important TMDCs like MoS_2 and WS_2 and faithfully reproduce the band gap. We were also able to filter arbitrary states based on a user-defined orbital angular momentum symmetry constraint. We trained the network on various flavors of computation using not only a classical computer, *qasm* backend quantum simulator in Qiskit, but also a real IBMQ machine (IBM Sydney and IBM Toronto) with the objective to see the performance of the algorithm on actual NISQ devices. In all flavors of computation, our algorithm demonstrated very high accuracy when compared to the exact values obtained from direct diagonalization.

Venturing beyond the ground state to obtain arbitrary states based on user-defined restrictions is the first of its kind in all flavors of QML. Furthermore, the systems of our choice happen to be TMDCs, an important class of 2D-periodic systems which have never been studied using any quantum algorithm. Periodic systems in general have received scanty attention as far as quantum algorithms are concerned. Only two such reports exist,[47,69] both of which have simulated just the valence band in graphene and hexagonal boron nitride (h-BN).

It must also be emphasized that a host of classical algorithms have been developed in traditional quantum chemistry that are extremely accurate and polynomially efficient. Over the past few decades, density functional theory (DFT) has emerged into a leading candidate for accurate computation of wide-variety of electronic structure problems in molecules and materials.[117,118] Variants of it are being developed for cases wherein multireference correlation would be important too.[119] Reduced density-matrix based methods are also polynomially scaling[120] and have shown excellent accuracy in strongly correlated systems.[121,122] Tensor-network-based methods like Density-Matrix Renormalization Group (DMRG)[123−125] have been developed which even though capable of exploiting rank sparsity in strongly correlated one-dimensional systems yet loses the polynomial advantage in multiple dimensions. Like our algorithm which attempts to construct the many-body state, a plethora of similar wave function based *ab initio* methods exist in traditional quantum chemistry too starting from the uncorrelated Hartree–Fock method to post-Hartree methods which can recover dynamic correlation like perturbative approaches (like MP2),[126−128] Truncated Configuration-Interaction or CI (like CISD),[127,129] Couple-Cluster (CC) methods[130] (like CCSD, CCSD(T), CCSDT, or EOM-CCSD for excited states), recently developed SHCI methods[131,132] to ones which are good for capturing static correlation like Multi-Configurational Self-Consistent Field (MCSCF).[133] A direct comparison of a quantum algorithm like ours with these classical algorithms can be attempted to be made in terms of accuracy and resource cost. In terms of resource requirements, the comparison is made difficult by the fact that certain parameters like circuit width, circuit depth, and so on which affect the performance of quantum algorithms like ours, have no classical analogues. If we consider an N_{elec} electron system with $r = r_o + r_{uo}$ spin orbitals/Fermionic modes such that $r_o = N_{elec}$ is the occupied orbital set in Hartree–Fock reference and r_{uo} are virtual orbitals excluded from Hartree–Fock reference, then under the assumption that the orbital space rank loosely equates to qubits or number of visible neurons n (see section 3.1 and ref 83), we have shown in section 3.3 that the spin–orbital cost of our algorithm would be $\approx O(r^2) = O(r_o^2 + r_{uo}^2 + 2r_{uo}r_o)$. The numerical parameter count of our algorithm is also quadratic, i.e., $O(\alpha n^2) \approx O(r^2) = O(r_o^2 + r_{uo}^2 + 2r_{uo}r_o)$. This is unlike methods like CCSD (Coupled-Cluster Singles Doubles) which has a computational cost of $\approx O(r_o^2 r_{uo}^4)$ (for CCSDT it is $\approx O(r_o^3 r_{uo}^5)$ and for CCSD(T) it is $\approx O(r_o^3 r_{uo}^4)$).[130,134] CCSD also evaluates $\approx O(r_o^2 r_{uo}^2)$ cluster amplitudes as parameters defining the excitations. Chemically important phenomenon, like dissociation events which are no longer single-referenced, are known to be difficult to treat with CCSD,[135] even though pair cluster doubles can ameliorate the situation to some extent.[136] That being said, it must also be noted that traditional variant of CCSD unlike ours is nonvariational. As far as accuracy is concerned, all results in this report are benchmarked against exponentially scaling exact diagonalization as that affords the exact value in a given basis. Not only the physics of Bloch states in the material TMDC but also a molecular example like LiH has been treated using our algorithm (see section 10 in the Supporting Information). For both the ground and excited states of LiH we see good accuracy and improvement of error by enhancing the hidden node density which makes the ansatz more expressive. Studies on larger molecular systems for which the results of exact diagonalization may not be available may be undertaken in future. That will provide a platform for comparison in accuracy with a subset of the aforesaid classical

https://doi.org/10.1021/jacs.1c06246
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

algorithms. Desirable chemical features like size consistency and size extensivity may be probed too.

One must also note that several quantum algorithms already exist which aim at obtaining ground and excited states of Fermionic systems.[137] Nonvariational quantum algorithms like quantum phase estimation (QPE)[40,138,139] have exponential speed up[140] yet require high circuit depth and long coherent operations which are beyond the limits of near-term hardwares.[23] Hybrid Variational Quantum Algorithms (VQA) have also been developed which can ameliorate some of the above problems.[46] The most notable one in the list is Unitary Coupled-Cluster Variational Quantum Eigensolver (UCC-VQE).[43] In its most traditional variant, the unitary ansatz which UCC-VQE uses for state preparation consists of single and double excitations[141−143] (hence often called Unitary Coupled Cluster Singles Doubles or UCCSD) from the reference state. However, the circuit depth in preparing such an ansatz is still large, and the circuit is parametrized by many variables which necessitates a high-dimensional classical optimization routine[144] to update the parameters. To be concrete, for r_o and r_{uo} with the same meaning as described in previous paragraph, the UCCSD-VQE uses[137] $O(r_o)$ qubits, $O(r_o^2 r_{uo}^2)$ cluster amplitudes as parameters, and $O(fr_o^4)$ gates where $f = O(r_o)$ or $O(\log(r_o))$ depending on the qubit-mapping. Besides, the UCC-VQE method can suffer from errors incurred due to operator ordering or Trotterization.[145] Also, the ansatz requires a high degree of qubit connectivity for nonlocal operations which may not always be available in all hardware.[146] A hardware-efficient ansatz[44,147] has been developed to help solve the above issues which use an alternating framework of single-qubit gates and fixed entangling operations which can be chosen with the specific capabilities of the device in hand. However, unlike UCC-VQE, such an ansatz is not physically inspired and often suffers from trainability issues during parameter optimization.[46,148] Besides the number of parameters grow as a function of entangling blocks and can even surpass the size of the Hilbert space.[137,147] A third variant that has low circuit depth and parameter cost is the ADAPT-VQE approach.[149] Unlike in the previous two cases, this variant constructs the circuit from a preselected pool of operators and changes the circuit architecture adaptively by adding operators from the pool which affects the energy gradient the most. The chosen pool decides the parameter count and gate counts in the circuit. In this method, the number of measurement shots can be high for computing the gradients,[144] and it is generally not clear how to preselect the operator pool and what guarantees that the pool is complete, i.e., the ansatz it produces is expressive enough. Many different variants for each method have been constructed for which the reader is referred to many excellent reviews.[46,137,146] For excited states,[46] deflation-assisted VQE as described before[72] exist, but for its implementation it used the UCCSD ansatz which inherits some of the above problems of high parameter count and gates. A recent promising method known as Weighted Subspace-Search VQE uses an input array of several orthogonal states to construct a weighted Lagrangian as the cost function.[150] In this case, the input states are mapped to the excited states of the system using a parametrized ansatz circuit. Depending on the nature of the ansatz circuit, the algorithm can have different gate count or parameter count and hence it is hard to mention a general estimate.

Our algorithm is also a hybrid variational algorithm like the ones in the aforesaid list, but it always requires quadratic resources (see section 3.3). However, there are some key differences as well which need to be acknowledged. Unlike the above list of algorithms which prepare a unitary ansatz on a quantum computer to mimic the state, our algorithm proposes to construct a probability distribution that mimics the amplitude field of the target state on a quantum computer. As a result, our algorithm is a distribution sampling protocol on a quantum computer using a nonunitary ansatz (RBM) which is manifested in the usage of ancilla and its subsequent measurement collapse. The measurement statistics of such a collapse are discussed in detail in sections 2 and 5 in the Supporting Information. Since the distribution encoding the amplitude field is based on RBM, unlike the ADAPT-VQE method, our protocol is largely *problem-agnostic*. This is due to the fact that RBM can act as a universal approximant to any probability density[19] and hence can be used for a variety of problems provided it is made sufficiently expressive with an adequate hidden node density. Also unlike other algorithms wherein the nature of the excitations or operator pool used decides the cost-function gradient, in our case the distribution function being RBM always permits training the network with analytical gradients. Besides, we have already demonstrated in section 3.3 using ref 90 that an analogous classical construction of RBM distribution has an exponential overhead, whereas by using a quantum algorithm like ours, one can construct it using quadratic resources thereby illustrating the distinct quantum-classical advantage in our algorithm directly.

Further extension of this algorithm can be made to compute operators using Hellmann−Feynmann method,[151] to characterize the influence of noise on the algorithm and to see it being extended to study other interesting phenomena on 2D materials like Rashba splitting in polar TMDCs[152] or even effect of strain.[153] One must also note that in this work we construct the full $d = 2^n$-dimensional eigenstate from the amplitude encoding using the RBM ansatz (eq 4) and the phase encoding using eq 5. This is because the primary quantum advantage of our algorithm lies in the fact we use quadratic resources to learn the full RBM distribution which classically would require exponential resources as necessitated in.[90] Besides, access to the full state allows us to compute matrix elements of arbitrary operator between eigenstates important for spectral information, i.e., learning in excitonic features[114] or thermal and electronic conductivity,[116] which as said before are important future extensions of this work. Also, once trained for a given system, the neural network in our algorithm can be used to learn the eigenstates of a closely related system accurately with faster convergence and fewer iterations, indicating partial transferability of these models (see section 9 of the Supporting Information for details). Benefits and the scope of such "transferable training" for other chemically motivated systems will be investigated in future. It must be noted that the symmetry partitioning of the metal orbitals in TMDCs guaranteed in ref 103 have reduced the effective size of the orbital space and qubit requirements in this study. However, understanding spectral information in excitonic physics would require more involved models with a larger orbital space. A way forward may be focusing on low-energy excitons with a certain symmetry (like overall spin-angular momentum) characteristics only. For molecular systems such symmetry-inspired cost reductions are already beginning to be noticed[84,86] as discussed earlier. However, such an initiative for materials is largely an uncharted territory. Further reduction in qubit resource requirements of our

https://doi.org/10.1021/jacs.1c06246
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

algorithm may also help, even though the nonunitary nature of the ansatz as discussed before makes it harder. From the hardware point of view, robust large-scale error mitigation strategies are beginning to be made available now,[95,154] and devices with over 1000 qubits with low qubit decoherence errors and gate infidelities are also being promised in recent future.[155] Such resources would certainly be beneficial to extensions of studies like these.

From the algorithmic point of view, besides being quadratic scaling in qubit and gate requirements and parameter count, our algorithm does not have any dependence on oracular objects like qRAM[35] which is responsible for creating a superposition of all possible basis states and is known to commonly sought in most quantum machine-learning modules. As futuristic quantum devices are being developed with proper error mitigation schemes, we expect to have more such cross-pollination between machine-learning algorithms and quantum computing with the promise to study electronic structure and dynamics in new complex materials.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.1c06246.

Proof of convergence of Theorem 2.1. in the manuscript, derivation of the generic lower bound for successful sampling, information about the input Hamiltonian and L_z/L^2 operator corresponding to MoS_2 and WS_2, Hamiltonian matrix used for MoS_2 and WS_2, importance of measurement error mitigation, measurement statistics for the systems in the report, how the results are affected by changing hidden node density, the SOC splitting data for WS_2, L^z/L^2 operator corresponding to MoS_2 and WS_2, transferability of learning from trained network to other systems, and molecular example of LiH (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Sabre Kais − *Department of Chemistry, Department of Physics and Astronomy, and Purdue Quantum Science and Engineering Institute, Purdue University, West Lafayette, Indiana 47907, United States;* orcid.org/0000-0003-0574-5346; Email: kais@purdue.edu

Authors

Manas Sajjan − *Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, United States*

Shree Hari Sureshbabu − *Elmore Family School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana 47907, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.1c06246

Notes

The authors declare no competing financial interest.

The codes associated with the classical simulation, simulation on the *qasm* backend, and the implementation on IBM's quantum computing devices is available with the corresponding author upon reasonable request.

■ ACKNOWLEDGMENTS

We thank Dr. Ruth Pachter, AFRL, for many useful discussions. AFRL support is acknowledged. We also thank Sangchul Oh for his help during the preparation of the manuscript. We acknowledge funding by the U.S. Department of Energy (Office of Basic Energy Sciences) under Award No. DE-SC0019215 and the National Science Foundation under award number 1955907. This material is also based upon work supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers. We also acknowledge the use of IBM-Q and thank them for the support. The views expressed are those of the authors and do not reflect the official policy or position of IBM or the IBM Q team.

■ REFERENCES

(1) Cohen, S. Chapter 1 - The evolution of machine learning: past, present, and future. In *Artificial Intelligence and Deep Learning in Pathology*; Cohen, S., Ed.; Elsevier, 2021; pp 1−12.

(2) Chen, C.; Seff, A.; Kornhauser, A.; Xiao, J. DeepDriving: Learning Affordance for Direct Perception in Autonomous Driving. *2015 IEEE Int. Conf. Computer Vision* **2015**, 2722−2730.

(3) He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. *Proc. IEEE Conf. Computer Vision Pattern Recognit.* **2016**, 770−778.

(4) Sak, H.; Senior, A.; Rao, K.; Irsoy, O.; Graves, A.; Beaufays, F.; Schalkwyk, J. Learning acoustic frame labeling for speech recognition with recurrent neural networks. *Proc. ICASSP* **2015**, 4280−4284.

(5) Hoy, M. B. Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Medical reference services quarterly* **2018**, *37*, 81−88.

(6) Silver, D.; et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **2016**, *529*, 484−489.

(7) Bansak, K.; Ferwerda, J.; Hainmueller, J.; Dillon, A.; Hangartner, D.; Lawrence, D.; Weinstein, J. Improving refugee integration through data-driven algorithmic assignment. *Science* **2018**, *359*, 325−329.

(8) Carleo, G.; Cirac, I.; Cranmer, K.; Daudet, L.; Schuld, M.; Tishby, N.; Vogt-Maranto, L.; Zdeborová, L. Machine learning and the physical sciences. *Rev. Mod. Phys.* **2019**, *91*, 45002.

(9) Snyder, J. C.; Rupp, M.; Hansen, K.; Müller, K.-R.; Burke, K. Finding Density Functionals with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 253002.

(10) Brockherde, F.; Vogt, L.; Li, L.; Tuckerman, M. E.; Burke, K.; Müller, K.-R. Bypassing the Kohn-Sham equations with machine learning. *Nat. Commun.* **2017**, *8*, 872.

(11) Arsenault, L. F.; Lopez-Bezanilla, A.; Von Lilienfeld, O. A.; Millis, A. J. Machine learning for many-body physics: The case of the Anderson impurity model. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *90*, 155136.

(12) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.

(13) Li, Z.; Kermode, J. R.; De Vita, A. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.* **2015**, *114*, 096405.

(14) Wang, L. Discovering phase transitions with unsupervised learning. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *94*, 214502.

(15) Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436−444.

(16) Carrasquilla, J.; Melko, R. G. Machine learning phases of matter. *Nat. Phys.* **2017**, *13*, 431−434.

(17) Ch'Ng, K.; Carrasquilla, J.; Melko, R. G.; Khatami, E. Machine learning phases of strongly correlated fermions. *Phys. Rev. X* **2017**, *7*, 031038.

(18) Cong, I.; Choi, S.; Lukin, M. D. Quantum convolutional neural networks. *Nat. Phys.* **2019**, *15*, 1273−1278.

(19) Le Roux, N.; Bengio, Y. Representational Power of Restricted Boltzmann Machines and Deep Belief Networks. *Neural Computation* **2008**, *20*, 1631−1649.
(20) Melko, R. G.; Carleo, G.; Carrasquilla, J.; Cirac, J. I. Restricted Boltzmann machines in quantum physics. *Nat. Phys.* **2019**, *15*, 887−892.
(21) Torlai, G.; Mazzola, G.; Carrasquilla, J.; Troyer, M.; Melko, R.; Carleo, G. Neural-network quantum state tomography. *Nat. Phys.* **2018**, *14*, 447−450.
(22) Carleo, G.; Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **2017**, *355*, 602−606.
(23) Preskill, J. Quantum Computing in the NISQ era and beyond. *Quantum* **2018**, *2*, 79.
(24) Biamonte, J.; Wittek, P.; Pancotti, N.; Rebentrost, P.; Wiebe, N.; Lloyd, S. Quantum machine learning. *Nature* **2017**, *549*, 195−202.
(25) Lloyd, S.; Mohseni, M.; Rebentrost, P. Quantum principal component analysis. *Nat. Phys.* **2014**, *10*, 631−633.
(26) Rebentrost, P.; Mohseni, M.; Lloyd, S. Quantum support vector machine for big data classification. *Phys. Rev. Lett.* **2014**, *113*, 130503.
(27) Dunjko, V.; Taylor, J. M.; Briegel, H. J. Quantum-Enhanced Machine Learning. *Phys. Rev. Lett.* **2016**, *117*, 130501.
(28) Lloyd, S.; Mohseni, M.; Rebentrost, P. Quantum algorithms for supervised and unsupervised machine learning. *arXiv (Quantum Physics)*, Nov. 4, 2013, 1307.0411, ver. 2. https://arxiv.org/pdf/1307.0411 (accessed 2021−11−10).
(29) Otten, M.; Goumiri, I. R.; Priest, B. W.; Chapline, G. F.; Schneider, M. D. Quantum Machine Learning using Gaussian Processes with Performant Quantum Kernels. *arXiv (Quantum Physics)*, April 23, 2020, 2004.11280. https://arxiv.org/pdf/2004.11280. (accessed 2021−11−10).
(30) Zhao, Z.; Fitzsimons, J. K.; Fitzsimons, J. F. Quantum-assisted Gaussian process regression. *Phys. Rev. A: At., Mol., Opt. Phys.* **2019**, *99*, 052331.
(31) Neven, H.; Denchev, V. S.; Rose, G.; Macready, W. G. Training a large scale classifier with the quantum adiabatic algorithm. *arXiv (Quantum Physics)*, Dec. 4, 2009, 0912.0779. https://arxiv.org/pdf/0912.0779 (accessed 2021−11−10).
(32) Harrow, A. W.; Hassidim, A.; Lloyd, S. Quantum Algorithm for Linear Systems of Equations. *Phys. Rev. Lett.* **2009**, *103*, 150502.
(33) Rebentrost, P.; Steffens, A.; Marvian, I.; Lloyd, S. Quantum singular-value decomposition of nonsparse low-rank matrices. *Phys. Rev. A: At., Mol., Opt. Phys.* **2018**, *97*, 12327.
(34) Cui, X.; Shi, Y. *QBLAS: A Quantum Basic Linear Algebra and Simulation Library*. https://github.com/xpclove/qblas (accessed 2021−11−10).
(35) Ciliberto, C.; Herbster, M.; Ialongo, A. D.; Pontil, M.; Rocchetto, A.; Severini, S.; Wossnig, L. Quantum machine learning: a classical perspective. *Proc. R. Soc. London, Ser. A* **2018**, *474*, 20170551.
(36) Amin, M. H.; Andriyash, E.; Rolfe, J.; Kulchytskyy, B.; Melko, R. Quantum Boltzmann Machine. *Phys. Rev. X* **2018**, *8*, 021050.
(37) Kieferová, M.; Wiebe, N. Tomography and generative training with quantum Boltzmann machines. *Phys. Rev. A: At., Mol., Opt. Phys.* **2017**, *96*, 062327.
(38) Wiebe, N.; Kapoor, A.; Svore, K. M. Quantum deep learning. *arXiv (Quantum Physics)*, May 22, 2015, 1412.3489, ver. 2. https://arxiv.org/pdf/1412.3489 (accessed 2021−11−10).
(39) Xia, R.; Kais, S. Quantum machine learning for electronic structure calculations. *Nat. Commun.* **2018**, *9*, 4195.
(40) Aspuru-Guzik, A.; Dutoi, A. D.; Love, P. J.; Head-Gordon, M. Simulated Quantum Computation of Molecular Energies. *Science* **2005**, *309*, 1704−1707.
(41) Smart, S. E.; Mazziotti, D. A. Quantum Solver of Contracted Eigenvalue Equations for Scalable Molecular Simulations on Quantum Computing Devices. *Phys. Rev. Lett.* **2021**, *126*, 070504.
(42) Kais, S. Introduction to Quantum Information and Computation for Chemistry. In *Quantum for Chemistry*; John Wiley & Sons, Ltd, 2014; Chapter 1, pp 1−38.
(43) Peruzzo, A.; McClean, J.; Shadbolt, P.; Yung, M.-H.; Zhou, X.-Q.; Love, P. J.; Aspuru-Guzik, A.; O'Brien, J. L. A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* **2014**, *5*, 4213.
(44) Kandala, A.; Mezzacapo, A.; Temme, K.; Takita, M.; Brink, M.; Chow, J. M.; Gambetta, J. M. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **2017**, *549*, 242.
(45) Daskin, A.; Kais, S. Direct application of the phase estimation algorithm to find the eigenvalues of the Hamiltonians. *Chem. Phys.* **2018**, *514*, 87−94.
(46) Cerezo, M.; Arrasmith, A.; Babbush, R.; Benjamin, S. C.; Endo, S.; Fujii, K.; McClean, J. R.; Mitarai, K.; Yuan, X.; Cincio, L.; Coles, P. J. Variational quantum algorithms. *Nature Reviews Physics* **2021**, *3*, 625−644.
(47) Kanno, S.; Tada, T. Many-body calculations for periodic materials via restricted Boltzmann machine-based VQE. *Quantum Science and Technology* **2021**, *6*, 025015.
(48) Choi, W.; Choudhary, N.; Han, G. H.; Park, J.; Akinwande, D.; Lee, Y. H. Recent development of two-dimensional transition metal dichalcogenides and their applications. *Mater. Today* **2017**, *20*, 116−130.
(49) Liu, M.; Liu, W.; Liu, X.; Wang, Y.; Wei, Z. Application of transition metal dichalcogenides in mid-infrared fiber laser. *Nano Select* **2021**, *2*, 37−46.
(50) Lv, R.; Robinson, J. A.; Schaak, R. E.; Sun, D.; Sun, Y.; Mallouk, T. E.; Terrones, M. Transition Metal Dichalcogenides and Beyond: Synthesis, Properties, and Applications of Single- and Few-Layer Nanosheets. *Acc. Chem. Res.* **2015**, *48*, 56−64.
(51) Manzeli, S.; Ovchinnikov, D.; Pasquier, D.; Yazyev, O. V.; Kis, A. 2D transition metal dichalcogenides. *Nature Reviews Materials* **2017**, *2*, 17033.
(52) Li, C.; Cao, Q.; Wang, F.; Xiao, Y.; Li, Y.; Delaunay, J.-J.; Zhu, H. Engineering graphene and TMDs based van der Waals heterostructures for photovoltaic and photoelectrochemical solar energy conversion. *Chem. Soc. Rev.* **2018**, *47*, 4981−5037.
(53) Alharbi, F. H.; Kais, S. Theoretical limits of photovoltaics efficiency and possible improvements by intuitive approaches learned from photosynthesis and quantum coherence. *Renewable Sustainable Energy Rev.* **2015**, *43*, 1073−1089.
(54) Palczewski, K. Chemistry and biology of vision. *J. Biol. Chem.* **2012**, *287*, 1612−1619.
(55) Tapavicza, E.; Tavernelli, I.; Rothlisberger, U. Trajectory Surface Hopping within Linear Response Time-Dependent Density-Functional Theory. *Phys. Rev. Lett.* **2007**, *98*, 023001.
(56) Cerullo, G.; Polli, D.; Lanzani, G.; De Silvestri, S.; Hashimoto, H.; Cogdell, R. J. Photosynthetic Light Harvesting by Carotenoids: Detection of an Intermediate Excited State. *Science* **2002**, *298*, 2395−2398.
(57) Hu, Z.; Engel, G. S.; Kais, S. Double-excitation manifolds effect on exciton transfer dynamics and the efficiency of coherent light harvesting. *Phys. Chem. Chem. Phys.* **2018**, *20*, 30032−30040.
(58) Rodgers, C. T.; Hore, P. J. Chemical magnetoreception in birds: The radical pair mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 353−360.
(59) Zhang, Y.; Berman, G. P.; Kais, S. Sensitivity and entanglement in the avian chemical compass. *Phys. Rev. E* **2014**, *90*, 042707.
(60) Chung, L. W.; Hayashi, S.; Lundberg, M.; Nakatsu, T.; Kato, H.; Morokuma, K. Mechanism of efficient firefly bioluminescence via adiabatic transition state and seam of sloped conical intersection. *J. Am. Chem. Soc.* **2008**, *130*, 12880−12881.
(61) Aleksandrowicz, G.; Alexander, T.; Barkoutsos, P.; Bello, L.; Ben-Haim, Y.; Bucher, D.; Cabrera-Hernández, F. J.; Carballo-Franquis, J.; Chen, A.; Chen, C. F. et al. Qiskit: An open-source framework for quantum computing. DOI: 10.5281/zenodo.2562111 (accessed on Mar 16, 2019).
(62) Stoudenmire, E. M.; White, S. R. Studying Two-Dimensional Systems with the Density Matrix Renormalization Group. *Annu. Rev. Condens. Matter Phys.* **2012**, *3*, 111−128.

https://doi.org/10.1021/jacs.1c06246
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

(63) Umrigar, C. J.; Wilson, K. G.; Wilkins, J. W. Optimized trial wave functions for quantum Monte Carlo calculations. *Phys. Rev. Lett.* **1988**, *60*, 1719−1722.
(64) Pathak, S.; Busemeyer, B.; Rodrigues, J. N. B.; Wagner, L. K. Excited states in variational Monte Carlo using a penalty method. *J. Chem. Phys.* **2021**, *154*, 034101.
(65) Kuroiwa, K.; Nakagawa, Y. O. Penalty methods for a variational quantum eigensolver. *Physical Review Research* **2021**, *3*, 013197.
(66) Ryabinkin, I. G.; Genin, S. N.; Izmaylov, A. F. Constrained Variational Quantum Eigensolver: Quantum Computer Search Engine in the Fock Space. *J. Chem. Theory Comput.* **2019**, *15*, 249−255.
(67) Greene-Diniz, G.; Muñoz Ramo, D. Generalized unitary coupled cluster excitations for multireference molecular states optimized by the variational quantum eigensolver. *Int. J. Quantum Chem.* **2021**, *121*, e26352.
(68) Suzuki, Y.; et al. Qulacs: a fast and versatile quantum circuit simulator for research purpose. *Quantum* **2021**, *5*, 559.
(69) Sureshbabu, S. H.; Sajjan, M.; Oh, S.; Kais, S. Implementation of Quantum Machine Learning for Electronic Structure Calculations of Periodic Systems on Quantum Computing Devices. *J. Chem. Inf. Model.* **2021**, *61*, 2667−2674.
(70) Miranda-Quintana, R. A.; González, M. M. Deflation techniques in quantum chemistry: Excited states from ground states. *Int. J. Quantum Chem.* **2013**, *113*, 2478−2488.
(71) Hemmatiyan, S.; Sajjan, M.; Schlimgen, A. W.; Mazziotti, D. A. Excited-State Spectra of Strongly Correlated Molecules from a Reduced-Density-Matrix Approach. *J. Phys. Chem. Lett.* **2018**, *9*, 5373−5378.
(72) Higgott, O.; Wang, D.; Brierley, S. Variational Quantum Computation of Excited States. *Quantum* **2019**, *3*, 156.
(73) Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U. S. A.* **1982**, *79*, 2554−2558.
(74) Ackley, D. H.; Hinton, G. E.; Sejnowski, T. J. A learning algorithm for boltzmann machines. *Cognitive Science* **1985**, *9*, 147−169.
(75) Hinton, G. E. Training Products of Experts by Minimizing Contrastive Divergence. *Neural Computation* **2002**, *14*, 1771−1800.
(76) Fischer, A.; Igel, C. Training restricted Boltzmann machines: An introduction. *Pattern Recognition* **2014**, *47*, 25−39.
(77) Hinton, G. E.; Salakhutdinov, R. R. Reducing the Dimensionality of Data with Neural Networks. *Science* **2006**, *313*, 504−507.
(78) Mehta, P.; Bukov, M.; Wang, C. H.; Day, A. G.; Richardson, C.; Fisher, C. K.; Schwab, D. J. A high-bias, low-variance introduction to Machine Learning for physicists. *Phys. Rep.* **2019**, *810*, 1−124.
(79) Nguyen, H. C.; Zecchina, R.; Berg, J. Inverse statistical problems: from the inverse Ising problem to data science. *Adv. Phys.* **2017**, *66*, 197−261.
(80) Zhang, N.; Sun, S. Multiview Graph Restricted Boltzmann Machines. *IEEE Trans. Cybern.* **2021**, 1−15.
(81) Chen, D.; Lv, J.; Yi, Z. Graph Regularized Restricted Boltzmann Machine. *IEEE Transactions on Neural Networks and Learning Systems* **2018**, *29*, 2651−2659.
(82) Cai, Z.; Liu, J. Approximating quantum many-body wave functions using artificial neural networks. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2018**, *97*, 035116.
(83) Tranter, A.; Love, P. J.; Mintert, F.; Coveney, P. V. A Comparison of the Bravyi-Kitaev and Jordan-Wigner Transformations for the Quantum Simulation of Quantum Chemistry. *J. Chem. Theory Comput.* **2018**, *14*, 5617−5630.
(84) Setia, K.; Chen, R.; Rice, J. E.; Mezzacapo, A.; Pistoia, M.; Whitfield, J. D. Reducing Qubit Requirements for Quantum Simulations Using Molecular Point Group Symmetries. *J. Chem. Theory Comput.* **2020**, *16*, 6091−6097.
(85) Takeshita, T.; Rubin, N. C.; Jiang, Z.; Lee, E.; Babbush, R.; McClean, J. R. Increasing the Representation Accuracy of Quantum Simulations of Chemistry without Extra Quantum Resources. *Phys. Rev. X* **2020**, *10*, 011004.
(86) Zhang, F.; Gomes, N.; Berthusen, N. F.; Orth, P. P.; Wang, C.-Z.; Ho, K.-M.; Yao, Y.-X. Shallow-circuit variational quantum eigensolver based on symmetry-inspired Hilbert space partitioning for quantum chemical calculations. *Phys. Rev. Research* **2021**, *3*, 013039.
(87) Kingma, D. P.; Ba, J. L. Adam: A method for stochastic optimization. *Proc. 3rd Int. Conf. Learn. Representations* **2015**, 1−15.
(88) Nekrasov, P.; Freeze, J.; Batista, V. Using Restricted Boltzmann Machines to Model Molecular Geometries. *arXiv (Chemical Physics)*, Dec. 13, 2020, 2012.06984. https://arxiv.org/pdf/2012.06984 (accessed 2021−11−10).
(89) Borin, A.; Abanin, D. A. Approximating power of machine-learning ansatz for quantum many-body states. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2020**, *101*, 195141.
(90) Long, P. M.; Servedio, R. A. Restricted Boltzmann Machines are hard to approximately evaluate or simulate. *Proc. 27th Int. Conf. Mach. Learn.* **2010**, 703−710.
(91) Harris, C. R.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357−362.
(92) 27-qubit backend: "IBMQ Team, IBM Q 27 Sydney backend specification V1.0.46", 2021. https://quantum-computing.ibm.com (accessed 2021−11−10).
(93) 27-qubit backend: "IBMQ Team, IBM Q 27 Toronto backend specification V1.4.25", 2021. https://quantum-computing.ibm.com (accessed 2021−11−10).
(94) Barron, G. S.; Wood, C. J. Measurement Error Mitigation for Variational Quantum Algorithms. *arXiv (Quantum Physics)*, Oct. 16, 2020, 2010.08520. https://arxiv.org/abs/2010.08520 (accessed 2021−11−10).
(95) Jattana, M. S.; Jin, F.; De Raedt, H.; Michielsen, K. General error mitigation for quantum circuits. *Quantum Inf. Process.* **2020**, *19*, 414.
(96) Yang, L.; Xie, C.; Jin, J.; Ali, R. N.; Feng, C.; Liu, P.; Xiang, B. Properties, Preparation and Applications of Low Dimensional Transition Metal Dichalcogenides. *Nanomaterials* **2018**, *8*, 463.
(97) Bao, W.; Cai, X.; Kim, D.; Sridhara, K.; Fuhrer, M. S. High mobility ambipolar MoS2 field-effect transistors: Substrate and dielectric effects. *Appl. Phys. Lett.* **2013**, *102*, 042104.
(98) Choudhary, N.; Patel, M. D.; Park, J.; Sirota, B.; Choi, W. Synthesis of large scale MoS2 for electronics and energy applications. *J. Mater. Res.* **2016**, *31*, 824−831.
(99) Ahn, E. C. 2D materials for spintronic devices. *npj 2D Materials and Applications* **2020**, *4*, 17.
(100) Mak, K. F.; Lee, C.; Hone, J.; Shan, J.; Heinz, T. F. Atomically Thin ${{MoS}}_{2}$: A New Direct-Gap Semiconductor. *Phys. Rev. Lett.* **2010**, *105*, 136805.
(101) Gong, C.; Zhang, Y.; Chen, W.; Chu, J.; Lei, T.; Pu, J.; Dai, L.; Wu, C.; Cheng, Y.; Zhai, T.; Li, L.; Xiong, J. Electronic and Optoelectronic Applications Based on 2D Novel Anisotropic Transition Metal Dichalcogenides. *Advanced Science* **2017**, *4*, 1700231.
(102) Mak, K. F.; Xiao, D.; Shan, J. Light-valley interactions in 2D semiconductors. *Nat. Photonics* **2018**, *12*, 451−460.
(103) Liu, G. B.; Shan, W. Y.; Yao, Y.; Yao, W.; Xiao, D. Three-band tight-binding model for monolayers of group-VIB transition metal dichalcogenides. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *88*, 085433.
(104) Ridolfi, E.; Le, D.; Rahman, T. S.; Mucciolo, E. R.; Lewenkopf, C. H. A tight-binding model for {MoS}2monolayers. *J. Phys.: Condens. Matter* **2015**, *27*, 365501.
(105) Zahid, F.; Liu, L.; Zhu, Y.; Wang, J.; Guo, H. A generic tight-binding model for monolayer, bilayer and bulk MoS2. *AIP Adv.* **2013**, *3*, 052111.
(106) Shahriari, M.; Ghalambor Dezfuli, A.; Sabaeian, M. Band structure and orbital character of monolayer MoS2 with eleven-band tight-binding model. *Superlattices Microstruct.* **2018**, *114*, 169−182.

https://doi.org/10.1021/jacs.1c06246
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

(107) Fang, S.; Kuate Defo, R.; Shirodkar, S. N.; Lieu, S.; Tritsaris, G. A.; Kaxiras, E. Ab initio tight-binding Hamiltonian for transition metal dichalcogenides. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2015**, *92*, 205108.
(108) Xiao, D.; Liu, G.-B.; Feng, W.; Xu, X.; Yao, W. Coupled Spin and Valley Physics in Monolayers of MoS2 and Other Group-VI Dichalcogenides. *Phys. Rev. Lett.* **2012**, *108*, 196802.
(109) Kormányos, A.; Zólyomi, V.; Drummond, N. D.; Rakyta, P.; Burkard, G.; Fal'Ko, V. I. Monolayer MoS2: Trigonal warping, the Γ valley, and spin-orbit coupling effects. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *88*, 045416.
(110) Berman, O. L.; Kezerashvili, R. Y. High-temperature superfluidity of the two-component Bose gas in a transition metal dichalcogenide bilayer. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *93*, 245410.
(111) Ominato, Y.; Fujimoto, J.; Matsuo, M. Valley-Dependent Spin Transport in Monolayer Transition-Metal Dichalcogenides. *Phys. Rev. Lett.* **2020**, *124*, 166803.
(112) Van Der Donck, M.; Zarenia, M.; Peeters, F. M. Excitons and trions in monolayer transition metal dichalcogenides: A comparative study between the multiband model and the quadratic single-band model. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2017**, *96*, 035131.
(113) Donck, M. V. Excitonic complexes in transition metal dichalcogenides and related materials. Ph.D. thesis, Universiteit Antwerpen, Belgium, 2019.
(114) Fang, H.; Bechtel, H. A.; Plis, E.; Martin, M. C.; Krishna, S.; Yablonovitch, E.; Javey, A. Quantum of optical absorption in two-dimensional semiconductors. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 11688−11691.
(115) Trolle, M. L.; Seifert, G.; Pedersen, T. G. Theory of excitonic second-harmonic generation in monolayer MoS2. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89*, 235410.
(116) Sengupta, P.; Tan, Y.; Klimeck, G.; Shi, J. Low-temperature thermal transport and thermopower of monolayer transition metal dichalcogenide semiconductors. *J. Phys.: Condens. Matter* **2017**, *29*, 405701.
(117) Burke, K. Perspective on density functional theory. *J. Chem. Phys.* **2012**, *136*, 150901.
(118) Maurer, R. J.; Freysoldt, C.; Reilly, A. M.; Brandenburg, J. G.; Hofmann, O. T.; Björkman, T.; Lebègue, S.; Tkatchenko, A. Advances in Density-Functional Calculations for Materials Modeling. *Annu. Rev. Mater. Res.* **2019**, *49*, 1−30.
(119) Gagliardi, L.; Truhlar, D. G.; Li Manni, G.; Carlson, R. K.; Hoyer, C. E.; Bao, J. L. Multiconfiguration Pair-Density Functional Theory: A New Way To Treat Strongly Correlated Systems. *Acc. Chem. Res.* **2017**, *50*, 66−73.
(120) Mazziotti, D. A. Large-Scale Semidefinite Programming for Many-Electron Quantum Mechanics. *Phys. Rev. Lett.* **2011**, *106*, 083001.
(121) Mazziotti, D. A. Two-Electron Reduced Density Matrix as the Basic Variable in Many-Electron Quantum Chemistry and Physics. *Chem. Rev.* **2012**, *112*, 244−262.
(122) Montgomery, J. M.; Mazziotti, D. A. Strong Electron Correlation in Nitrogenase Cofactor, FeMoco. *J. Phys. Chem. A* **2018**, *122*, 4988−4996.
(123) White, S. R. Density matrix formulation for quantum renormalization groups. *Phys. Rev. Lett.* **1992**, *69*, 2863−2866.
(124) Chan, G. K.-L.; Sharma, S. The Density Matrix Renormalization Group in Quantum Chemistry. *Annu. Rev. Phys. Chem.* **2011**, *62*, 465−481.
(125) Schollwöck, U. The density-matrix renormalization group. *Rev. Mod. Phys.* **2005**, *77*, 259−315.
(126) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. MP2 energy evaluation by direct methods. *Chem. Phys. Lett.* **1988**, *153*, 503−506.
(127) Raghavachari, K.; Anderson, J. B. Electron Correlation Effects in Molecules. *J. Phys. Chem.* **1996**, *100*, 12960−12973.
(128) Helgaker, T.; Jørgensen, P.; Olsen, J. Perturbation Theory. In *Molecular Electronic-Structure Theory*; John Wiley & Sons, Ltd, 2000; Chapter 14, pp 724−816.
(129) Helgaker, T.; Jørgensen, P.; Olsen, J. Configuration-Interaction Theory. In *Molecular Electronic-Structure Theory*; John Wiley & Sons, Ltd., 2000; Chapter 11, pp 523−597.
(130) Bartlett, R. J.; Musiał, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291−352.
(131) Li, J.; Otten, M.; Holmes, A. A.; Sharma, S.; Umrigar, C. J. Fast semistochastic heat-bath configuration interaction. *J. Chem. Phys.* **2018**, *149*, 214110.
(132) Holmes, A. A.; Umrigar, C. J.; Sharma, S. Excited states using semistochastic heat-bath configuration interaction. *J. Chem. Phys.* **2017**, *147*, 164111.
(133) Szalay, P. G.; Müller, T.; Gidofalvi, G.; Lischka, H.; Shepard, R. Multiconfiguration Self-Consistent Field and Multireference Configuration Interaction Methods and Applications. *Chem. Rev.* **2012**, *112*, 108−181.
(134) Helgaker, T.; Jørgensen, P.; Olsen, J. Coupled-Cluster Theory. In *Molecular Electronic-Structure Theory*; John Wiley & Sons, Ltd, 2000; Chapter 13, pp 648−723.
(135) Fan, P.-D.; Piecuch, P. The Usefulness of Exponential Wave Function Expansions Employing One- and Two-Body Cluster Operators in Electronic Structure Theory: The Extended and Generalized Coupled-Cluster Methods. *Adv. Quantum Chem.* **2006**, *51*, 1−57.
(136) Bulik, I. W.; Henderson, T. M.; Scuseria, G. E. Can Single-Reference Coupled Cluster Theory Describe Static Correlation? *J. Chem. Theory Comput.* **2015**, *11*, 3171−3179.
(137) Cao, Y.; Romero, J.; Olson, J. P.; Degroote, M.; Johnson, P. D.; Kieferová, M.; Kivlichan, I. D.; Menke, T.; Peropadre, B.; Sawaya, N. P. D.; Sim, S.; Veis, L.; Aspuru-Guzik, A. Quantum Chemistry in the Age of Quantum Computing. *Chem. Rev.* **2019**, *119*, 10856−10915.
(138) Lanyon, B. P.; Whitfield, J. D.; Gillett, G. G.; Goggin, M. E.; Almeida, M. P.; Kassal, I.; Biamonte, J. D.; Mohseni, M.; Powell, B. J.; Barbieri, M.; Aspuru-Guzik, A.; White, A. G. Towards quantum chemistry on a quantum computer. *Nat. Chem.* **2010**, *2*, 106.
(139) Veis, L.; Visnak, J.; Nishizawa, H.; Nakai, H.; Pittner, J. Quantum chemistry beyond Born-Oppenheimer approximation on a quantum computer: A simulated phase estimation study. *Int. J. Quantum Chem.* **2016**, *116*, 1328−1336.
(140) Abrams, D. S.; Lloyd, S. Quantum Algorithm Providing Exponential Speed Increase for Finding Eigenvalues and Eigenvectors. *Phys. Rev. Lett.* **1999**, *83*, 5162−5165.
(141) McClean, J. R.; Romero, J.; Babbush, R.; Aspuru-Guzik, A. The theory of variational hybrid quantum-classical algorithms. *New J. Phys.* **2016**, *18*, 023023.
(142) Romero, J.; Babbush, R.; McClean, J. R.; Hempel, C.; Love, P. J.; Aspuru-Guzik, A. Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz. *Quantum Science and Technology* **2019**, *4*, 014008.
(143) Yung, M.-H.; Casanova, J.; Mezzacapo, A.; McClean, J.; Lamata, L.; Aspuru-Guzik, A.; Solano, E. From transistor to trapped-ion computers for quantum chemistry. *Sci. Rep.* **2015**, *4*, 3589.
(144) Fedorov, D. A.; Peng, B.; Govind, N.; Alexeev, Y. VQE Method: A Short Survey and Recent Developments. *arXiv (Quantum Physics)*, Aug. 30, 2021, 2103.08505. https://arxiv.org/pdf/2103.08505 (accessed 2021−11−10).
(145) Grimsley, H. R.; Claudino, D.; Economou, S. E.; Barnes, E.; Mayhall, N. J. Is the Trotterized UCCSD Ansatz Chemically Well-Defined? *J. Chem. Theory Comput.* **2020**, *16*, 1−6.
(146) Bauer, B.; Bravyi, S.; Motta, M.; Chan, G. K.-L. Quantum Algorithms for Quantum Chemistry and Quantum Materials Science. *Chem. Rev.* **2020**, *120*, 12685−12717.
(147) Barkoutsos, P. K.; Gonthier, J. F.; Sokolov, I.; Moll, N.; Salis, G.; Fuhrer, A.; Ganzhorn, M.; Egger, D. J.; Troyer, M.; Mezzacapo, A.; et al. Quantum algorithms for electronic structure calculations: Particle-hole Hamiltonian and optimized wave-function expansions. *Phys. Rev. A: At., Mol., Opt. Phys.* **2018**, *98*, 022322.

https://doi.org/10.1021/jacs.1c06246
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

(148) McClean, J. R.; Boixo, S.; Smelyanskiy, V. N.; Babbush, R.; Neven, H. Barren plateaus in quantum neural network training landscapes. *Nat. Commun.* **2018**, *9*, 4812.
(149) Grimsley, H. R.; Economou, S. E.; Barnes, E.; Mayhall, N. J. An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nat. Commun.* **2019**, *10*, 3007.
(150) Nakanishi, K. M.; Mitarai, K.; Fujii, K. Subspace-search variational quantum eigensolver for excited states. *Phys. Rev. Research* **2019**, *1*, 033062.
(151) Oh, S. Quantum computational method of finding the ground-state energy and expectation values. *Phys. Rev. A: At., Mol., Opt. Phys.* **2008**, *77*, 012326.
(152) Yao, Q.-F.; Cai, J.; Tong, W.-Y.; Gong, S.-J.; Wang, J.-Q.; Wan, X.; Duan, C.-G.; Chu, J. H. Manipulation of the large Rashba spin splitting in polar two-dimensional transition-metal dichalcogenides. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2017**, *95*, 165401.
(153) Peng, Z.; Chen, X.; Fan, Y.; Srolovitz, D. J.; Lei, D. Strain engineering of 2D semiconductors and graphene: from strain fields to band-structure tuning and photonic applications. *Light: Sci. Appl.* **2020**, *9*, 190.
(154) Chen, Z.; et al. Exponential suppression of bit or phase errors with cyclic error correction. *Nature* **2021**, *595*, 383−387.
(155) Gambetta, J. IBM's Roadmap For Scaling Quantum Technology. 2020. https://www.ibm.com/blogs/research/2020/09/ibm-quantum-roadmap/ (accessed 2021−11−10).

https://doi.org/10.1021/jacs.1c06246
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX