

Modal Backflow Neural Quantum States for Anharmonic Vibrational Calculations

Lexin Ding^{1,*} and Markus Reiher^{1,†}

¹*ETH Zürich, Department of Chemistry and Applied Biosciences,*

Vladimir-Prelog-Weg 2, CH-8093 Zürich, Switzerland

Abstract

Neural quantum states (NQS) are a promising ansatz for solving many-body quantum problems due to their inherent expressiveness. Yet, this expressiveness can only be harnessed efficiently for treating identical particles if the suitable physical knowledge is hardwired into the neural network itself. For electronic structure, NQS based on backflow determinants has been shown to be a powerful ansatz for capturing strong correlation. By contrast, the analogue for bosons, backflow permanents, is unpractical due to the steep cost of computing the matrix permanent and due to the lack of particle conservation in common bosonic problems. To circumvent these obstacles, we introduce a modal backflow (MBF) NQS design and demonstrate its efficacy by solving the anharmonic vibrational problem. To accommodate the demand of high accuracy in spectroscopic calculations, we implement a selected-configuration scheme for evaluating physical observables and gradients, replacing the standard stochastic approach based on Monte Carlo sampling. A vibrational self-consistent field calculation is conveniently carried out within the MBF network, which serves as a pretraining step to accelerate and stabilize the optimization. In applications to both artificial and *ab initio* Hamiltonians, we find that the MBF network is capable of delivering spectroscopically accurate zero-point energies and vibrational transitions in all anharmonic regimes.

* lexin.ding@phys.chem.ethz.ch

† mreiher@ethz.ch

I. INTRODUCTION

A central endeavor in quantum chemistry is the efficient solution of the many-body Schrödinger equation. In electronic structure theory, this involves solving the electronic Hamiltonian with the Coulomb interaction in an exponentially large Hilbert space. Since exact diagonalization is only feasible for around two dozens of half-filled spatial orbitals,[1, 2] tailored wavefunction ansätze and optimization schemes have been developed to bypass the exponential scaling. Examples are selected configuration interaction (see, e.g., [3–8]), tensor network states (TNSs),[9–11] full configuration interaction quantum Monte Carlo,[12, 13] and auxiliary field quantum Monte Carlo.[14, 15] Despite the tremendous improvements in the past few decades, common wavefunction methods still suffer from at least one of the following shortcomings: The failure to include both static and dynamical correlation, steep scaling in the number of parameters to be optimized for strongly correlated systems, and the lack of efficient optimization methods.

In an effort to overcome these shortcomings, Carleo and Troyer proposed a novel type of quantum state ansatz using artificial neural networks (ANN).[16] Their idea was to leverage the expressiveness and flexible designs of ANNs,[17–19] as well as the knowledge on their optimization.[20–23] After all, a quantum state is nothing but a function that maps particle positions to complex numbers, which, according to the universal approximation theorem, is formally learnable to arbitrary accuracy by feedforward neural networks.[24, 25] We note that deep feedforward neural networks are more expressive than tensor networks[26] and that neural networks can even simulate volume-law entanglement with only polynomially many degrees of freedom.[27, 28] These seminal results quickly ignited a new type of approximation, now known as neural quantum states (NQSs).[29–36] For the electronic ground state problem in small molecules, NQSs can even reach the accuracy obtained with full configuration interaction. [32, 37, 38]

Compared to the large body of work on electronic NQSs, application to the bosonic vibrational problem is somewhat less explored and limited to the first-quantized form.[39] In contrast to lattice Hamiltonians, the vibrational problem has much less structure, in that every pair of vibrational modes is connected by the Hamiltonian. In analogy to correlated wavefunction ansätze in electronic structure theory, there exist vibrational counterparts, including vibrational configuration interaction methods[40–43] and vibrational cou-

pled cluster theories.[44–47] Recently, the vibrational density matrix renormalization group (vDMRG) optimizing a TNS ansatz has become an alternative for calculating accurate molecular spectra.[48–52] However, matrix product states (MPSs) and tree tensor network states (TTNSs) assume certain topologies of the correlation of the vibrational modes, which conceptually clashes with the structureless character of the vibrational Hamiltonian. By contrast, fully connected neural networks assume no such topologies.

Under the Born-Oppenheimer approximation, the vibrational Hamiltonian is comprised by a kinetic term and by a representation of the potential energy surface (PES, given by the electronic energy at different nuclear positions), which determines the achievable accuracy crucially. At first sight, the vibrational problem seems easier than the electronic one, due to the absence of long-range Coulomb interactions. However, for a solution to be predictive for vibrational spectroscopy, one must be able to resolve closely lying excited states while dealing with highly many-body coupling terms (typically up to six-body in practice, whereas the Coulomb interaction is only two-body) in the anharmonic vibrational Hamiltonian.

Although for a moderate number of normal modes (about two dozens) the spectroscopic accuracy of 1 cm^{-1} [53] is routinely reachable with multilayer multiconfigurational time-dependent Hartree [54–56] and TNSs methods [48–50, 52, 57], a larger number of modes can only be handled for lattice Hamiltonians with a limited connectivity of interactions. [56] Hence, in view of the properties of NQSs, it is natural to explore and assess their capabilities for the vibrational structure problem. We note that the solution of the vibrational Hamiltonian is a fundamental building block for more complex problems in quantum chemistry, such as vibronic effects [58, 59] or vibrational dynamics inside cavities.[60, 61]

That being said, the universal approximation theorem [24, 25] alone does not guarantee that any neural network can be efficiently trained to solve a given problem. Successful implementations of NQSs for the electronic Hamiltonians showed that it is crucial to incorporate physical knowledge, such as the antisymmetrization of electronic wavefunctions and cusp conditions, into the networks.[32] Another important insight from the earlier work is that using ANN to learn indirect features, such as Jastrow factors and backflow determinants, is more efficient than directly learning the wavefunction.[28, 29, 31, 62] In particular, backflow determinants (Slater determinants computed from backflow transformed orbitals) plays a crucial role in compressing exponential degrees of freedom in highly entangled fermionic systems.[28, 63] The starting point of this work is therefore to identify the physical knowl-

edge in the bosonic vibrational problem that can be incorporated into the network design. Specifically, we set out to develop the bosonic analog of backflow transformed orbitals, which turns out to be not the vibrational modes, but rather basis functions of the single-mode Fock spaces, i.e., modals. Based on this insight, we propose the modal backflow (MBF) NQs and demonstrate their capability of solving the vibrational problem at different degrees of anharmonicity.

This paper is structured as follows. In Section II we briefly review the Watson Hamiltonian in second-quantized form, which will serve as our vibrational model Hamiltonian here. In Section III, we introduce the MBF NQs as an ansatz to solve the vibrational problem, at the example of the Watson Hamiltonian. In Section IV, we investigate MBF NQs on a set of randomly generated Watson Hamiltonian at different levels of anharmonicity. And finally in Section V, we apply the MBF NQ wavefunction to ab initio vibrational Hamiltonians of molecules and solve for the zero-point energy and low-lying transition energies.

II. ANHARMONIC VIBRATIONAL HAMILTONIAN

A general PES may be approximated by a finite-order Taylor expansion around the equilibrium structure of a molecule. In the normal coordinates $\hat{q}_i$ where the Hessian of the PES is diagonal, the Watson Hamiltonian with L degrees of freedom takes the following form[64]

$$\begin{aligned}
H_{\text{vib}} = & \frac{1}{2} \sum_{i=1}^L w_i (\hat{p}_i^2 + \hat{q}_i^2) \\
& + \frac{1}{6} \sum_{i,j,k=1}^L \Phi_{ijk}^{(3)} \hat{Q}_{ijk} + \frac{1}{24} \sum_{i,j,k,l=1}^L \Phi_{ijkl}^{(4)} \hat{Q}_{ijkl} + \dots \\
& + \sum_{i,j,k,l=1}^L \sum_{\tau=x,y,z} B^\tau \xi_{ij}^\tau \xi_{kl}^\tau \sqrt{\frac{w_j w_l}{w_i w_k}} \hat{q}_i \hat{p}_j \hat{q}_k \hat{p}_l.
\end{aligned} \tag{1}$$

Here, w_i 's are harmonic frequencies of the normal modes, and we use the shorthand notation $\hat{Q}_{ijk} = \hat{q}_i \hat{q}_j \hat{q}_k$ and $\hat{Q}_{ijkl} = \hat{q}_i \hat{q}_j \hat{q}_k \hat{q}_l$. $\Phi^{(3)}$ and $\Phi^{(4)}$ are the third- and fourth-order reduced force constants, defined as

$$\Phi_{ijk}^{(3)} = \frac{\kappa_{ijk}^{(3)}}{\sqrt{w_i w_j w_k}}, \quad \Phi_{ijkl}^{(4)} = \frac{\kappa_{ijkl}^{(4)}}{\sqrt{w_i w_j w_k w_l}}, \tag{2}$$

where $\kappa^{(3)}$ and $\kappa^{(4)}$ are the third-and fourth-order partial derivatives of the PES evaluated at the local minimum of the equilibrium structure, respectively. They provide anharmonic corrections to the harmonic approximation of the vibrational Hamiltonian. Additionally, the Watson Hamiltonian includes the Coriolis terms that couple the positions $\hat{q}_i$ and momenta $\hat{p}_i$ defined by the rotational constants B^τ and Coriolis coupling constants ξ_{ij}^τ . [65]

To arrive at the second-quantized form of the anharmonic Hamiltonian, we substitute the position and momentum operators with

$$\hat{p}_i = \frac{1}{\sqrt{2}}(b_i^\dagger + b_i), \quad \hat{q}_i = \frac{1}{\sqrt{2}}(b_i^\dagger - b_i), \quad (3)$$

where b_i ($b_i^\dagger$) are bosonic annihilation (creation) operators of mode i with actions on the occupation number vector (ONV) as follows

$$b_i |n_1 \cdots n_i \cdots n_L\rangle = \begin{cases} \sqrt{n_i} |n_1 \cdots n_i - 1 \cdots n_L\rangle, & n_i > 0 \\ 0, & n_i = 0 \end{cases}, \quad (4)$$

$$b_i^\dagger |n_1 \cdots n_i \cdots n_L\rangle = \sqrt{n_i + 1} |n_1 \cdots n_i + 1 \cdots n_L\rangle.$$

The substitution above then allows us to rewrite the Hamiltonian in second-quantized form,

$$\begin{aligned} \hat{H}_{\text{vib}} = & \sum_{i=1}^L w_i \left(b_i^\dagger b_i + \frac{1}{2} \right) \\ & + \frac{1}{12\sqrt{2}} \sum_{i,j,k=1}^L \Phi_{ijk}^{(3)} \prod_{r=i,j,k} (b_r^\dagger + b_r) \\ & + \frac{1}{96} \sum_{i,j,k,l=1}^L \Phi_{ijkl}^{(4)} \prod_{r=i,j,k,l} (b_r^\dagger + b_r) + \cdots \\ & + \frac{1}{4} \sum_{i,j,k,l=1}^L \sum_{\tau=x,y,z} B^\tau \xi_{ij}^\tau \xi_{kl}^\tau \sqrt{\frac{w_j w_l}{w_i w_k}} \\ & \times (b_i^\dagger + b_i)(b_j^\dagger - b_j)(b_k^\dagger + b_k)(b_l^\dagger - b_l). \end{aligned} \quad (5)$$

The Hilbert space $\mathcal{H}$ on which $\hat{H}_{\text{vib}}$ acts is, in principle, infinite. In practice, however, we impose a cutoff N_{max} as the maximum number of particles in each mode, or more generally $N_{\text{modal}} = N_{\text{max}} + 1$ as the number of basis states, also known as modal functions. Typically, N_{modal} is higher than the number of basis states in an electronic orbital, which is 4. This leads to a steeper exponential scaling $(N_{\text{modal}})^L$ in the Hilbert space dimension in terms of the number of vibrational modes L , compared to 4^L in the electronic case, where L now refers to the number of electronic orbitals.

The cutoff $N_{\max}$ for the mode occupancies (or N_{modal} for the number of modal basis functions) also leads to restrictions of bosonic operators to the truncated Hilbert space. We denote such restrictions as $[\cdot]_{N_{\max}}$. It is straightforward to write down the restricted annihilation (creation) operator $[b_i]_{N_{\max}}$ ($[b_i^\dagger]_{N_{\max}}$)

$$\begin{aligned}
[b_i]_{N_{\max}} &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & 0 & \sqrt{N_{\max}} & 0 \end{pmatrix}, \\
[b_i^\dagger]_{N_{\max}} &= \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \sqrt{2} & \cdots & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \cdots & \sqrt{N_{\max}} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.
\end{aligned} \tag{6}$$

However, the restriction of strings of creation/annihilation operators, such as $\hat{N}_i = b_i^\dagger b_i$, cannot be constructed by simply multiplying $[b_i^\dagger]_{N_{\max}}$ and $[b_i]_{N_{\max}}$, as one can verify that

$$[b_i^\dagger]_{N_{\max}} [b_i]_{N_{\max}} \neq [\hat{N}_i]_{N_{\max}} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & N_{\max} \end{pmatrix}. \tag{7}$$

To consistently construct the restrictions of bosonic operator strings, $[b_i^\dagger]_{N'_{\max}}$ and $[b_i]_{N'_{\max}}$ should be first constructed in a larger Hilbert space with $N'_{\max} > N_{\max}$, then multiply to form the bosonic operator string, and finally truncated to the desired dimension $N_{\max}$. The difference $\Delta N_{\max} = N'_{\max} - N_{\max}$ should be greater or equal to $l - 1$ where l is the length of the bosonic operator string. For example, in the case of $b_i^\dagger b_i$ we have $l = 2$, and the following relation holds

$$\left[[b_i^\dagger]_{N_{\max}+1} [b_i]_{N_{\max}+1} \right]_{N_{\max}} = [\hat{N}_i]_{N_{\max}}. \tag{8}$$

III. MODAL BACKFLOW NEURAL QUANTUM STATES

A. Network Ansatz

The concept of a backflow transformation dates back to Feynmann and Cohen in the 1950s.[66] It adds a multi-particle component to the single-particle coordinates[66, 67] to introduce a reverse flow of particle current, hence the term backflow. As the concept of backflow becomes more general, multi-particle dependency can also be mediated using occupation numbers,[68] while the backflow transformed quantities can also be single-particle functions (orbitals) instead of positions.[29, 62] In particular, neural backflow (NBF) ansätze for fermions have been shown to be significantly more efficient in capturing the correlation of indistinguishable fermions compared to a standard feedforward neural network (FNN).[28, 29, 62, 63] We therefore briefly recall the formulation of the NBF ansatz for N spinless fermions. Let $\mathbf{M}$ be an $L \times N$ matrix,

$$\mathbf{M} = \begin{pmatrix} \varphi_{1,1} & \varphi_{2,1} & \cdots & \varphi_{N,1} \\ \varphi_{1,2} & \varphi_{2,2} & \cdots & \varphi_{N,2} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{1,L} & \varphi_{2,L} & \cdots & \varphi_{N,L} \end{pmatrix}. \quad (9)$$

consisting of the molecular orbital coefficients of N occupied molecular orbitals $\varphi_i = \sum_{j=1}^L \varphi_{i,j} \chi_j$ expanded in an orbital basis of L one-electron functions $\{\chi_j\}_{j=1}^L$. An N -electron Slater determinant (SD) state where $\varphi_1, \varphi_2, \dots, \varphi_N$ are occupied can be written as

$$\begin{aligned} |\Psi^{\text{SD}}\rangle &= \prod_{i=1}^N f_{\varphi_i}^\dagger |0\rangle \\ &= \prod_{i=1}^N \left(\sum_{j=1}^L \varphi_{i,j} f_{\chi_j}^\dagger \right) |0\rangle \\ &= \sum_{1 \leq j_1 < j_2 < \dots < j_N \leq N} \left(\sum_{\sigma \in S_N} \text{sgn}(\sigma) \prod_{n=1}^N \varphi_{n, j_{\sigma(n)}} \right) f_{\chi_{j_1}}^\dagger f_{\chi_{j_2}}^\dagger \cdots f_{\chi_{j_N}}^\dagger |0\rangle. \end{aligned} \quad (10)$$

where $f_{\varphi_i/\chi_j}^\dagger$ are fermionic creation operators which create electrons in orbitals φ_i/χ_j from the vacuum state $|0\rangle$, and σ 's are elements of the permutation group S_N . The last line of Eq. (10) reveals that the corresponding coefficient $\Psi^{\text{SD}}(\mathbf{n})$ of an ONV state $|\mathbf{n}\rangle$ in the basis

$$\{\chi_j\}_{j=1}^L$$

$$|\mathbf{n}\rangle = |n_1, n_2, \dots, n_L\rangle \equiv (f_{\chi_1}^\dagger)^{n_1} (f_{\chi_2}^\dagger)^{n_2} \cdots (f_{\chi_L}^\dagger)^{n_L} |0\rangle, \quad (11)$$

is given by

$$\Psi^{\text{SD}}(\mathbf{n}) = \langle \mathbf{n} | \Psi^{\text{SD}} \rangle = \det \mathbf{M}(\mathbf{n}), \quad (12)$$

where $\mathbf{M}(\mathbf{n})$ is an $N \times N$ matrix consisting only of the N rows of $\mathbf{M}$ with indices i 's such that $n_i = 1$. The coefficients $\Psi^{\text{SD}}(\mathbf{n})$ are related by the same orbital coefficient matrix $\mathbf{M}$ such that $|\Psi^{\text{SD}}\rangle$ is indeed a Slater determinant state. The key idea behind the NBF state $|\Psi^{\text{NBF}}\rangle$ is that the coefficients of $|\mathbf{n}\rangle$ are instead computed from an *ONV-dependent* set of occupied orbitals $(\mathbf{M}^{(\mathbf{n})})_{ij} = \varphi_{i,j}^{(\mathbf{n})}$, namely,

$$\Psi^{\text{NBF}}(\mathbf{n}) = \langle \mathbf{n} | \Psi^{\text{NBF}} \rangle = \det \mathbf{M}^{(\mathbf{n})}(\mathbf{n}). \quad (13)$$

As a result, the coefficients $\Psi^{\text{NBF}}(\mathbf{n})$ can vary independently to introduce correlation, leading to a higher level of expressiveness. Additionally, by calculating the coefficients using backflowed determinants, NBF is able to impose a specific structure on the scalar output that gives a compact encoding of the target wavefunction.

However, the generalization of NBF to the case of bosonic modes is not straightforward. The direct analogue of the backflow determinant for bosons would be the backflow *permanent*, because bosonic wavefunctions must be fully symmetric. Although the matrix determinant can be computed with polynomial cost, no such efficient algorithm is available to exactly compute the matrix permanent that is needed for the symmetry of bosons.[69] Moreover, in many bosonic problems of interest, the number of bosons is not a good quantum number. This means that the number of bosons involved in each configuration and, therefore, the size of the matrix of which the permanent is to be computed is not fixed.

To avoid the issues above, we propose an alternative formulation of neural backflow ansatz using modals.[70] Modals are elements of the Fock space of a single mode i , the canonical example being the eigenstates $|n_i\rangle = (b_i^\dagger)^{n_i} |0\rangle$ of the particle number operator $\hat{N}_i = b_i^\dagger b_i$, where $n_i = 0, 1, \dots, N_{\text{max}}$. In general, a modal wavefunction $|\phi_i\rangle$ is an arbitrary superposition

$$|\phi_i\rangle = \sum_{n_i=0}^{N_{\text{max}}} \phi_{i,n_i+1} |n_i\rangle, \quad (14)$$

where $\boldsymbol{\phi}_i = (\phi_{i,j})_{j=1}^{N_{\text{modal}}}$ (recall that $N_{\text{modal}} = N_{\text{max}} + 1$) is a vector of modal coefficients. An L -mode state without any mode-mode correlation, such as the vibrational self-consistent

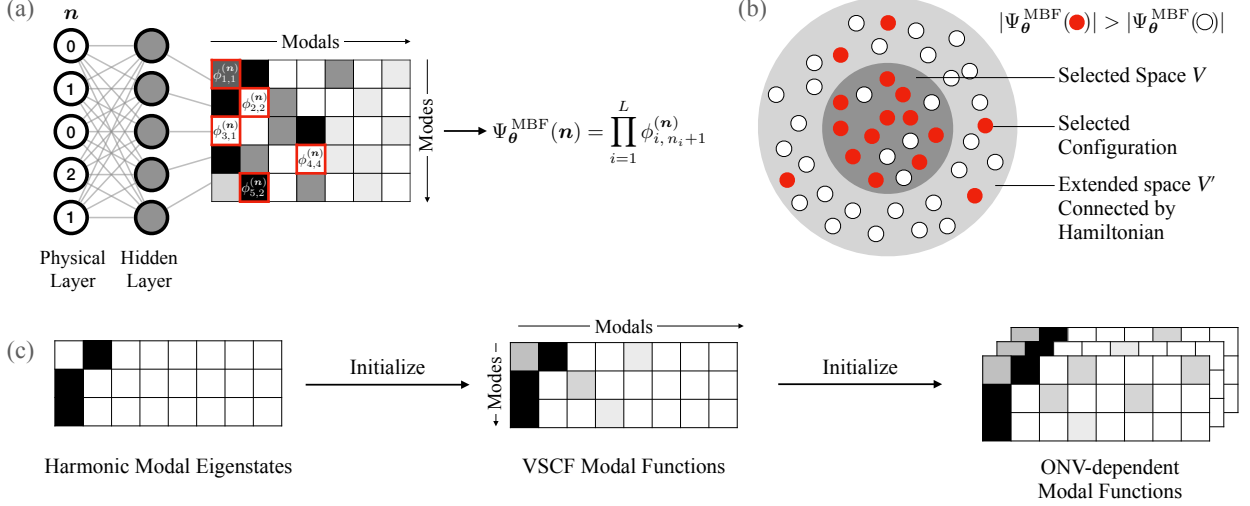


FIG. 1. (a) Architecture of the MBF network: ONVs of a set of bosonic modes are fed through a shallow FNN and output to a set of modal functions, from which the wavefunction value is calculated based on the occupations of each mode. (b) Selected-configuration scheme: For a given selected space V , a new selection of configurations can be proposed by selecting from the extended space V' connected by the Hamiltonian, based on comparing the wavefunction amplitudes given by the current MBF wavefunction. (c) Multi-step initialization scheme for bootstrapping the optimization of the MBF network: An eigenstate of the harmonic part of the Watson Hamiltonian is used to initialize a VSCF calculation. The resulting VSCF solution is used as the fixed part of the ONV-dependent modal functions, while the actual ONV-dependent deviations are learned by the MBF network.

field (VSCF) [71, 72] solution, is a modal product (MP) state,

$$|\Psi^{\text{MP}}\rangle = \bigotimes_{i=1}^L |\phi_i\rangle = \bigotimes_{i=1}^L \left(\sum_{n_i=0}^{N_{\max}} \phi_{i,n_i+1} |n_i\rangle \right) = \sum_{\mathbf{n}} \left(\prod_{i=1}^L \phi_{i,n_i+1} \right) |\mathbf{n}\rangle, \quad (15)$$

where

$$|\mathbf{n}\rangle = |n_1, n_2, \dots, n_L\rangle \equiv (b_1^\dagger)^{n_1} (b_2^\dagger)^{n_2} \dots (b_L^\dagger)^{n_L} |0\rangle. \quad (16)$$

The coefficient $\Psi^{\text{MP}}(\mathbf{n})$ of a bosonic ONV state $|\mathbf{n}\rangle$ is then given by

$$\Psi^{\text{MP}}(\mathbf{n}) = \langle \mathbf{n} | \Psi^{\text{MP}} \rangle = \prod_{i=1}^L \phi_{i,n_i+1}, \quad (17)$$

without the need for computing the permanent of a matrix. Analogously, to incorporate correlation into the wavefunction, we introduce the modal backflow (MBF) state $|\Psi^{\text{MBF}}\rangle$ to

be such that the modal functions ϕ_i 's are ONV-dependent. Accordingly, the coefficients of the ONV states become

$$\Psi^{\text{MBF}}(\mathbf{n}) = \langle \mathbf{n} | \Psi^{\text{MBF}} \rangle = \prod_{i=1}^L \phi_{i, n_i+1}^{(n)}. \quad (18)$$

The L ONV-dependent modal functions $\phi_i^{(n)}$ form an $L \times N_{\text{modal}}$ matrix which we will learn with a fully connected neural network. The MBF state naturally covers all bosonic particle number sectors, and correlates both within and among different sectors freely, without incurring the exponential computational cost of evaluating the bosonic permanent. Moreover, the modal representation allows us to work with a matrix of fixed dimensions consisting of L modal functions. We summarize the MBF network architecture in Figure 1 (a).

The parameters $\theta = (\mathbf{W}, \mathbf{b})$ of the MBF wavefunctions are the weights $\mathbf{W}$ and biases $\mathbf{b}$ that propagate the information through the network according to

$$\begin{aligned} \mathbf{h} &= \tanh(\mathbf{W}^{(0)} \mathbf{n} + \mathbf{b}^{(0)}), \\ \mathbf{o} &= \tanh(\mathbf{W}^{(1)} \mathbf{h} + \mathbf{b}^{(1)}), \end{aligned} \quad (19)$$

where $\mathbf{h}$ and $\mathbf{o}$ are the hidden and output layers, respectively. The vector $\mathbf{o}$ is flattened from the output matrix consisting of L modal functions of length N_{modal} depicted in Figure 1 (a). The number of hidden neurons N_{hidden} can be regulated using the hidden neuron density $\alpha = N_{\text{hidden}}/L$, which also acts as a tuning parameter for the complexity of the network. In practice we break down the ONV-dependent modal functions into a fixed part (which corresponds to a mode product state), and the variable part that is actually ONV-dependent

$$\phi_i^{(n)} = \phi_i^{(0)} + \delta_i^{(n)}. \quad (20)$$

During training, $\delta_i^{(n)}$ is reshaped from the vector output $\mathbf{o}$ of the network. In a later section, we will describe the way the fixed modal functions $\phi_i^{(0)}$ are chosen.

B. Optimization

In this section, we discuss several aspects of the optimization of the MBF network that are crucial to achieving the target accuracy. The implementation of MBF is carried out using the NQS package Netket.[73, 74]

Markov Chain Monte Carlo. The most common method of extracting observables from a neural quantum state Ψ is by Monte Carlo sampling.[75, 76] The expectation of an observable is mathematically equivalent to a weighted sum of the so-called local energy O_{loc} ,

$$\langle \hat{O} \rangle_{\Psi} = \frac{\langle \Psi | \hat{O} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \sum_{\mathbf{n}} P_{\Psi}(\mathbf{n}) O_{\text{loc}}(\mathbf{n}), \quad (21)$$

with

$$P_{\Psi}(\mathbf{n}) \equiv \frac{|\Psi(\mathbf{n})|^2}{\sum_{\mathbf{n}'} |\Psi(\mathbf{n}')|^2}, \quad (22)$$

and

$$O_{\text{loc}}(\mathbf{n}) \equiv \sum_{\mathbf{n}'} \langle \mathbf{n} | \hat{O} | \mathbf{n}' \rangle \frac{\Psi(\mathbf{n}')}{\Psi(\mathbf{n})}, \quad (23)$$

where $P_{\Psi}(\mathbf{n})$ is the probability distribution given by the quantum state Ψ . In Monte Carlo sampling, the exact weighted sum over all configurations $\mathbf{n}$ is approximated by a Markov chain $\mathcal{C}$ of configuration samples drawn according to probability distribution $P_{\Psi}(\mathbf{n})$. In other words,

$$\sum_{\mathbf{n}} P_{\Psi}(\mathbf{n}) O_{\text{loc}}(\mathbf{n}) \approx \sum_{\mathbf{n} \in \mathcal{C}} O_{\text{loc}}(\mathbf{n}). \quad (24)$$

Although this is a simple and powerful method, its stochastic nature does not align well with the requirement of high precision in spectroscopic calculations. Moreover, in most scenarios anharmonic eigenstates are dominated by a single configuration, resulting in a sharp peak in the probability distribution which heavily skews the sampling. One needs an exceedingly large number of samples in order to reach enough configurations beside the dominating one.

Stochastically Selected Configuration. To avoid the above issues, we use a selected-configuration scheme introduced in Ref. [77], which is illustrated in Figure 1 (b). Instead of replacing the exact weighted sum over all configurations with Monte Carlo sampling, we restrict the weighted sum to only a small number of distinct configurations with the highest weights according to the full wavefunction, collected in the set V . Mathematically, this is equivalent to an asymmetric evaluation of the expectation with the truncated state $|\Psi_V\rangle = \sum_{\mathbf{n} \in V} \Psi(\mathbf{n}) |\mathbf{n}\rangle$ and the exact state

$$\begin{aligned} \sum_{\mathbf{n}} P_{\Psi}(\mathbf{n}) O_{\text{loc}}(\mathbf{n}) &\approx \sum_{\mathbf{n} \in V} P_{\Psi_V}(\mathbf{n}) O_{\text{loc}}(\mathbf{n}) \\ &= \frac{\langle \Psi_V | \hat{O} | \Psi \rangle}{\langle \Psi_V | \Psi_V \rangle} \\ &\equiv \langle \hat{O} \rangle_V. \end{aligned} \quad (25)$$

Note that the evaluation $\langle \cdot \rangle_V$ is no longer strictly variational. However, we will find that the variational condition is well maintained as long as all configurations with significant weights are selected in V .

Optimizing an NQS $\Psi_{\boldsymbol{\theta}}$ with network parameters $\boldsymbol{\theta}$ using the selected-configuration scheme consists of the following steps:

1. Initialize $\Psi_{\boldsymbol{\theta}^{(0)}}$ and $V^{(0)}$ as $\{\mathbf{n}_0\}$ where $\mathbf{n}_0 = (0, 0, \dots, 0)$, or another seed state of choice (depending on the level of excitation).
2. In step t , compute the set $V' = \hat{H}V^{(t)}$ consisting of configurations connected by $\hat{H}$ to $V^{(t)}$.
3. Select N_s configurations from $V^{(t)} \cup V'$ with the highest amplitudes $|\Psi_{\boldsymbol{\theta}}(\mathbf{n})|$ to form the new set of selected states $V^{(t+1)}$.
4. Evaluate the energy gradients $\partial E / \partial \theta_m$ and update the network parameters $\boldsymbol{\theta}^{(t)} \leftarrow \boldsymbol{\theta}^{(t+1)}$.
5. Repeat from step 2 until the maximum number of iterations is reached.

The above scheme was first implemented to solve the electronic Hamiltonian.[77] In our case, the vibrational Hamiltonian contains long strings of creation and annihilation operators that go beyond the two-body terms. The implication is that the connected subspace V' in the vibrational case can be prohibitively large, making the step of identifying the N_s states with the highest contributing weights an expensive task. To avoid this, we restrict the full extended space V' to a smaller subset consisting of only $K \times N_s$ randomly sampled connected states, where K can be tuned as a parameter of the degree of exploration. This randomized restriction not only eases the computational effort, but also introduces some level of stochasticity.

Within the selected-configuration scheme, the gradients of the expectation value $\langle \hat{O} \rangle_V$ with respect to the network parameters θ_m 's are defined as

$$F_m = \frac{\partial E}{\partial \theta_m} = \sum_{\mathbf{n} \in V} P_{\Psi}(\mathbf{n}) D_m(\mathbf{n})^* (O_{\text{loc}}(\mathbf{n}) - \langle \hat{O} \rangle_V), \quad (26)$$

where

$$D_m(\mathbf{n}) = \frac{1}{\Psi_{\boldsymbol{\theta}}^{\text{MBF}}(\mathbf{n})} \frac{\partial \Psi_{\boldsymbol{\theta}}^{\text{MBF}}(\mathbf{n})}{\partial \theta_m}, \quad (27)$$

is the derivative of the logarithm of the wavefunction. At each iteration, we update the network parameters $\boldsymbol{\theta}$ using the continuous resilient (CoRe) optimizer [78, 79]

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \eta \mathbf{G}^{(t)}, \quad (28)$$

where η is the learning rate and $\mathbf{G}^{(t)}$ is the scaled gradient

$$\begin{aligned} G_m^{(t)} &= \frac{g_m^{(t)}}{1 - (\beta_1^{(t)})^t} \left(\sqrt{\frac{h_m^{(t)}}{1 - \beta_2^t}} + \epsilon \right)^{-1} \\ g_m^{(t)} &= \beta_1^{(t)} g_m^{(t-1)} + (1 - \beta_1^{(t)}) F_m^{(t-1)}, \\ h_m^{(t)} &= \beta_2 h_m^{(t-1)} + (1 - \beta_2) (F_m^{(t-1)})^2, \\ \beta_1^{(t)} &= \beta_1^b + (\beta_1^a - \beta_1^b) \exp \left[- \left(\frac{t-1}{\beta_1^c} \right)^2 \right]. \end{aligned} \quad (29)$$

The default hyperparameters used in this work are $\eta = 0.05$, $\beta_1^a = 0.9$, $\beta_1^b = 0.5$, $\beta_1^c = 100$, $\beta_2 = 0.99$.

VSCF pretraining. Pretraining the network parameters is a proven technique for accelerating and stabilizing the optimization of the network.[32, 62] Here, we can pretrain the backflow-free part $\phi_i^{(0)}$ of the ONV-dependent modal functions $\phi_i^{(n)}$ in Eq. (20) to match the solution of a vibrational self-consistent field (VSCF) theory.[71, 72] The VSCF theory assumes a modal product (MP) state, where each mode i is described by a single modal function ϕ_i ,

$$|\Psi^{\text{MP}}\rangle = \bigotimes_{i=1}^L |\phi_i\rangle_i. \quad (30)$$

The mean-field solution to the Watson Hamiltonian is then obtained by the minimization

$$\begin{aligned} E^{\text{VSCF}} &= \min_{|\Psi^{\text{MP}}\rangle} \frac{\langle \Psi^{\text{MP}} | \hat{H}_{\text{vib}} | \Psi^{\text{MP}} \rangle}{\langle \Psi^{\text{MP}} | \Psi^{\text{MP}} \rangle} \\ &= \frac{\langle \Psi^{\text{VSCF}} | \hat{H}_{\text{vib}} | \Psi^{\text{VSCF}} \rangle}{\langle \Psi^{\text{VSCF}} | \Psi^{\text{VSCF}} \rangle}. \end{aligned} \quad (31)$$

This minimization is equivalent to solving for the optimal set of $\phi_i^{(0)}$, which is, in turn, equivalent to optimizing the MBF network with all weights and biases set to zero (meaning $\boldsymbol{\delta}_i^{(n)} = \mathbf{0}$). The VSCF modal functions already capture a considerable amount of anharmonic correction to the energy, allowing us to focus on the remaining part of the energy minimization by optimizing only the ONV-dependent modal corrections $\boldsymbol{\delta}_i^{(n)}$. The optimization for

$\phi_i^{(0)}$ is itself initialized with the eigenstates of the harmonic part of the vibrational Hamiltonian, which are ONV states (e.g., the vacuum for the ground state). The assumption behind this initialization scheme is that the harmonic solutions are in the vicinity of the VSCF solutions, which are themselves in the vicinity of the anharmonic solutions. We illustrate this nested pretraining procedure in Figure 1 (c). The VSCF modal functions are optimized using the same selected-configuration procedure, and the final set of selected-configurations are fed to the MBF network together with the VSCF modal functions. We will see that the VSCF pretraining significantly speeds up the optimization of the MBF wavefunction, and is crucial to avoiding getting stuck in higher lying states when targeting excited states.

Targeting excited states. There are two main approaches to calculate excited states: state-average/ensemble methods [80, 81] and state-specific calculations using an orthogonality penalty.[82–84] In this work we choose the second route for its simple implementation. Specifically, we use the shifted Hamiltonian and the corresponding implementation of the excited state solver using the NQS package NetKet [73, 74] given in Ref. [84]. The penalty-modified Hamiltonian for the n -th excited state reads

$$\hat{H}_{\text{vib}}^{(n)}(z) = \hat{H}_{\text{vib}} + z \sum_{j=1}^{n-1} \frac{|\Psi_j\rangle\langle\Psi_j|}{\langle\Psi_j|\Psi_j\rangle} - \text{ZPE}, \quad n \geq 1, \quad (32)$$

where the zero-point energy (ZPE) is subtracted for convenience. Mathematically, the shift constant z can be chosen to be any value higher than the target vibrational transitions. For example, setting $z = 1000 \text{ cm}^{-1}$ allows us to target all vibrational transitions from the ZPE below 1000 cm^{-1} . In practice, however, a z value that is too high can skew the optimization to prioritize maintaining orthogonality rather than minimizing the excited state energy. Therefore, z should ideally be level-dependent and only slightly above the excited state energy, which naturally requires some *a priori* knowledge of the target energy level. Since the n -th exact eigenstate of $\hat{H}_{\text{vib}}$ is also the variational ground state of $\hat{H}_{\text{vib}}^{(n)}$ regardless of the value of z , we can estimate z to be the energy of the VSCF solution of $\hat{H}_{\text{vib}}^{(n)}$

$$\begin{aligned} z_n &= \frac{\langle\Psi_n^{\text{VSCF}}|\hat{H}_{\text{vib}}|\Psi_n^{\text{VSCF}}\rangle}{\langle\Psi_n^{\text{VSCF}}|\Psi_n^{\text{VSCF}}\rangle} + z' \sum_{j=1}^{n-1} \frac{|\langle\Psi_j|\Psi_n^{\text{VSCF}}\rangle|^2}{\langle\Psi_j|\Psi_j\rangle\langle\Psi_n^{\text{VSCF}}|\Psi_n^{\text{VSCF}}\rangle} - \text{ZPE} \\ &\geq \frac{\langle\Psi_n^{\text{exact}}|\hat{H}_{\text{vib}}|\Psi_n^{\text{exact}}\rangle}{\langle\Psi_n^{\text{exact}}|\Psi_n^{\text{exact}}\rangle} - \text{ZPE} = E_n - \text{ZPE}, \end{aligned} \quad (33)$$

where z' is an initial estimate of the shift constant. Typically, the overlap $|\langle\Psi_j|\Psi_n^{\text{VSCF}}\rangle|$ is

already small due to the fact that they are dominated by different harmonic configurations, and z_n is considerably improved compared to z' .

IV. NUMERICAL EXPERIMENT

In this section, we investigate different degrees of anharmonicity to be described by the MBF network. In order to allow for a systematic investigation, we decided on creating an artificial Hamiltonian for a 4-mode system with randomly sampled third- and fourth-order partial derivatives $\kappa_{i_1, i_2, \dots, i_\nu}^{(\nu)}$ ($\nu = 3, 4$) of the PES that emulate weak, moderate, and strong anharmonicity. The harmonic frequencies are random numbers uniformly sampled from the interval $[1500, 3000]$ cm^{-1} and the partial derivatives of the PES are sampled according to normal distributions

$$|\kappa_{i_1, i_2, \dots, i_\nu}^{(\nu)}| = \mathcal{N}(\lambda_\nu, \lambda_\nu/5), \quad \nu = 3, 4, \quad (34)$$

where the mean λ and standard deviation $\lambda_\nu/5$ for the normal distribution $\mathcal{N}$ are controlled by the parameter (in Hartree atomic unit, a.u.)

$$\lambda_\nu = \begin{cases} (50 \text{ cm}^{-1})_{\text{a.u.}} \times \bar{w}^{\nu/2}, & \text{weak} \\ (150 \text{ cm}^{-1})_{\text{a.u.}} \times \bar{w}^{\nu/2}, & \text{moderate} \\ (500 \text{ cm}^{-1})_{\text{a.u.}} \times \bar{w}^{\nu/2}, & \text{strong} \end{cases} \quad (35)$$

for weak, moderate, and strong anharmonicity. Here, $\bar{w}$ was taken to be the midpoint of the sampling interval $(2250 \text{ cm}^{-1})_{\text{a.u.}}$. Furthermore, to mimic a typical PES, we took the absolute values of the fourth-order constants and set an off-diagonal decay factor $\eta = 1, 0.1, 0.01$ for fully- (all indices identical), semi- (only some indices identical), and off-diagonal (no indices identical) tensor elements, respectively. Throughout the numerical experiment, only cubic and quartic force constants were used for the Watson Hamiltonian, and no Coriolis terms were included.

We quantified the anharmonicity of the sampled Hamiltonians by the anharmonic correction in their ZPE, defined as the difference between the ground state energy of $\hat{H}_{\text{harm}} = \sum_{i=1}^L w_i(\hat{n}_i + \frac{1}{2})$ and that of $\hat{H}_{\text{vib}}$. In Figure 2 we plotted the distribution anharmonic correction of 10^3 sampled Hamiltonian, with $\kappa_{ijk}^{(3)}$ and $\kappa_{ijkl}^{(4)}$ both set to represent weak, moderate, and strong anharmonicity. For each Hamiltonian, the ground state was calculated with exact diagonalization and N_{max} was set to 9. We found that the three settings indeed

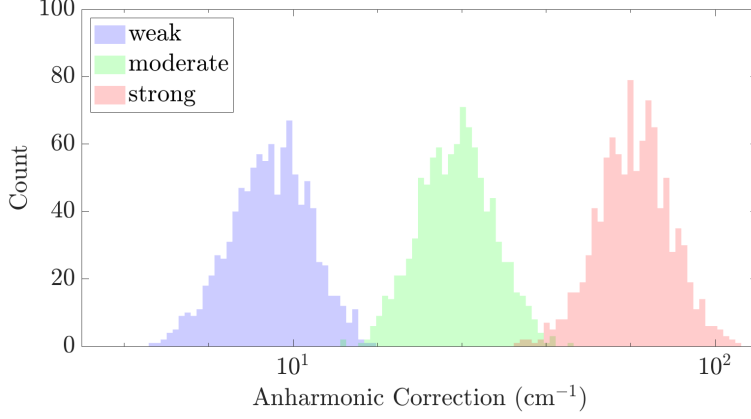


FIG. 2. Distribution of the anharmonic correction of sampled anharmonic 4-mode Hamiltonians with $N_{\max} = 9$. In each distribution, 10^3 Hamiltonians were sampled, and both $\kappa_{ijk}^{(3)}$ and $\kappa_{ijkl}^{(4)}$ were set to the weak, moderate, and strong regime of anharmonicity.

cover different ranges of the anharmonic correction with very little overlap. The setting for strong anharmonicity reaches the order of a 100 cm^{-1} anharmonic correction, which already surpasses molecules considered to be strongly anharmonic.

First, we investigate the effect of learning ONV-dependent modal functions instead of directly learning the wavefunction values through a shallow FNN. For this purpose, we chose the moderate setting for both the third- and fourth-order reduced force constants, and set to $N_{\max} = 6$. To isolate the effect of the network architecture for our comparison, we fixed $N_s = 128$ and did not apply any pretraining. The size of the extended subspace V' was set at $5N_s$. Furthermore, we imposed an exponential decay in the learning rate $\eta = \exp(-rt/T)\eta_0$, where t and T are the current and total step counts, respectively, and the decay rate r was set to 0.1.

In the top panel of Figure 3, we show a comparison between the accuracy reached by the FNN and MBF networks after 2000 iteration steps, both with a hidden neuron density $\alpha = 1$. The energy of the MBF wavefunction improved significantly faster than that of FNN. At the end of the optimization, the error of the MBF energy reached around 0.1 cm^{-1} , which is two orders of magnitude smaller than that of the FNN energy. In the bottom panel of Figure 3, we compared the final errors of the energy reached by both networks with different values of hidden neuron density α . As α increases, the error of the MBF energy decreases roughly according to a power law, indicated by the approximately linear curve in

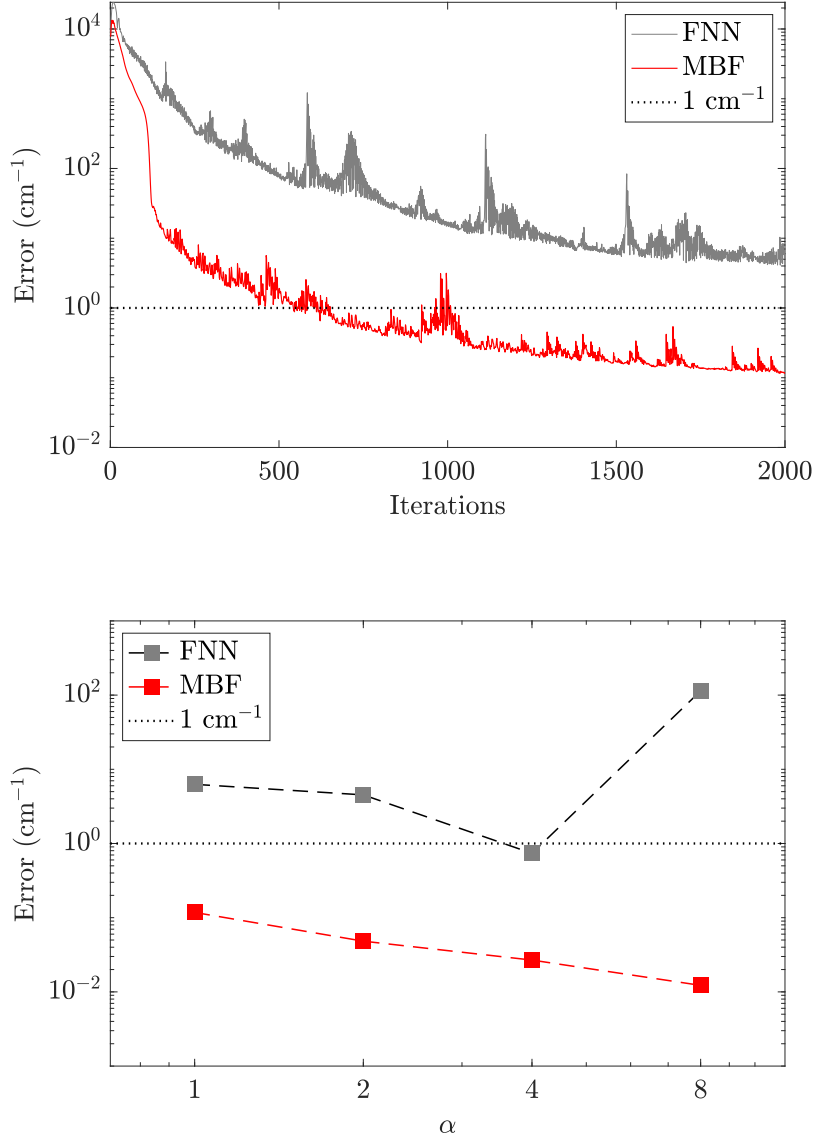


FIG. 3. (Top) Comparison of the optimization between FNN and MBF for targeting the ground state of a randomly generated 4-mode Watson Hamiltonian with the moderate anharmonicity setting. For both networks we set $\alpha = 1$. (Bottom) Comparison of the final error between FNN and MBF after 2000 iterations, for $\alpha = 1, 2, 4$, and 8. $N_s = 128$ and $N_{\max} = 6$ for all calculations.

the logarithmic scale. We first found a similar improvement in FNN. At $\alpha = 4$, the FNN was able to reach spectroscopic accuracy of 1 cm⁻¹ and lower. However, increasing α further made the optimization unstable and did not produce an improved error. Together, both plots in Figure 3 demonstrate the clear advantage of the MBF network over FNN in terms of expressiveness and systematic improvability.

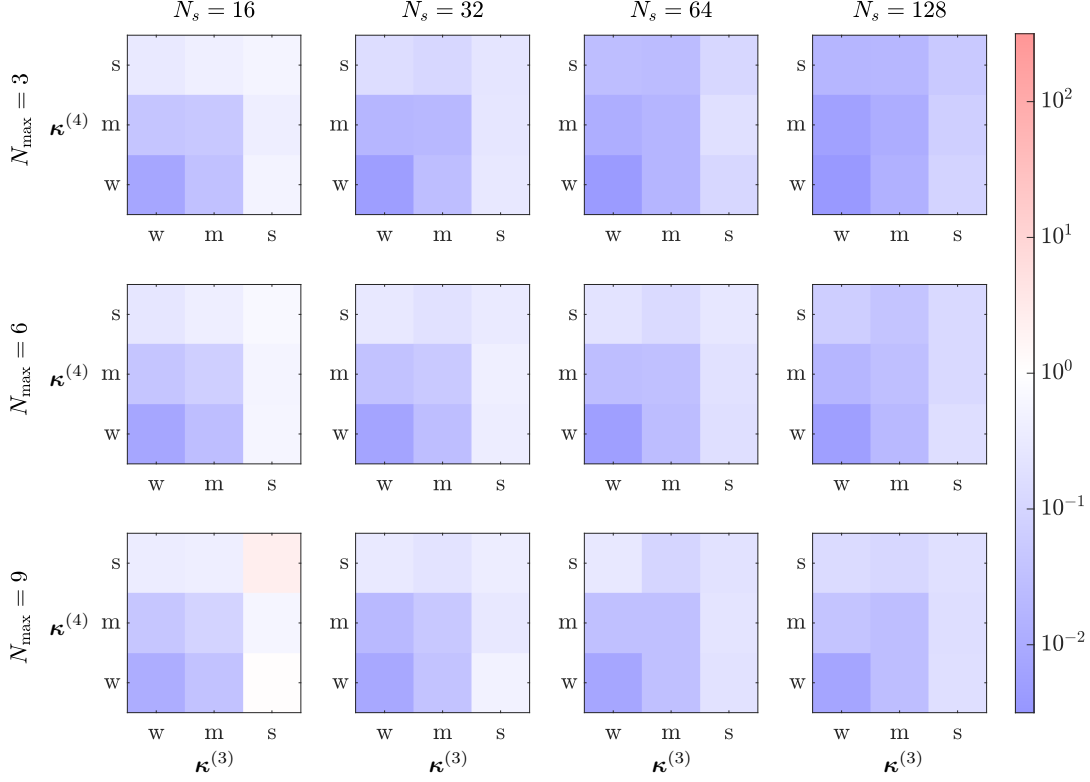


FIG. 4. Error (cm^{-1}) of the ground state optimization with MBF wavefunctions. Each 3×3 block correspond to a combination of N_{max} and selected space dimension N_s . Within each block, different combinations of anharmonic strength (w: weak, m: moderate, s: strong) of third- ($\kappa^{(3)}$) and fourth-order partial derivatives ($\kappa^{(4)}$) of the PES are used to generate 100 4-mode Watson Hamiltonians, over which the final errors of the MBF energy are averaged and color-coded on a log-scale, with red, white, and blue to represent above, at, and below spectroscopic accuracy of 1 cm^{-1} , respectively. All calculations used the same hyperparameters for the optimization and ran for 1000 iterations. A VSCF pretraining was used before each optimization.

Next, we analyzed the reliability of MBF for different combinations of anharmonic strength of the third- ($\kappa^{(3)}$) and fourth-order partial derivatives ($\kappa^{(4)}$) and for different N_{max} . For each combination, we used an MBF network with hidden neuron density $\alpha = 1$, and the sizes of the selected subspaces were $N_s = 16, 32, 64$, and 128 . In Figure 4, we showed the errors of the MBF ground state energy for different anharmonic strengths, grouped by N_{max} and N_s . The errors were measured against exact diagonalization results and color coded on a logarithmic scale in red, white, and blue to represent above, on,

and below the spectroscopic accuracy of 1 cm^{-1} , respectively. For each combination of $(\boldsymbol{\kappa}^{(3)}, \boldsymbol{\kappa}^{(4)}, N_{\text{max}}, N_s)$, 100 Hamiltonians were sampled and solved, and the final errors were averaged. A VSCF pretraining was used for all calculations. Almost all errors were below 1 cm^{-1} (blue), except for the top right corner of the grid of $N_{\text{max}} = 9$ and $N_s = 16$, which corresponds to solving the largest and most anharmonic system with the smallest number of selected-configurations. Moreover, we observed the following trends: (1) increasing N_s systematically reduces the final error; (2) the stronger the anharmonicity, the more configurations should be taken into the selected space to increase accuracy; (3) a larger N_{max} typically also calls for a larger selected subspace to reach spectroscopic accuracy of 1 cm^{-1} .

In this section, using randomly generated Watson Hamiltonians, we systematically tested the performance of the MBF network in terms of its advantages over FNNs, its range of applicability, and the effect of the choices of parameters such as the hidden neuron density α and size of selected space N_s . The insights gained provide valuable guidance as we now apply the MBF network to solve for the ground and excited states of *ab initio* anharmonic vibrational Hamiltonians of molecules.

V. AB INITIO ANHARMONIC VIBRATIONAL HAMILTONIANS

We applied the MBF network to target the ZPE and low-lying vibrational transitions of three molecules, ClO_2 , H_2CO , and CH_3CN , increasing both the number of modes and the degree of anharmonicity. For ClO_2 and H_2CO , we used the sextic force field and Corioli constants from the library PyPES [85]. For the CH_3CN molecule, although only quartic force fields are available (e.g., see Refs. [86, 87]), its strong anharmonic character makes it a suitable subject for benchmarking tensor network methods.[48, 52]. Here, we used the PES reported in Ref. [87], which was directly taken from the Supporting Information of Ref. [52]. The reference energies were computed using the program QCMAquis [88] with the vibrational density matrix renormalization group (vDMRG) calculations. We chose MPSs of maximum bond dimensions 100, which is sufficiently large for the largest system CH_3CN according to Ref. [48]. For the MBF calculations, the size of the extended subspace V' was set at $2N_s$ and the exponential decay rate r of the learning rate η is set to 0.5.

The triatomic ClO_2 is with its three normal modes weakly anharmonic. First, we demonstrate the ability of the VSCF pretraining step in accelerating and stabilizing excited state

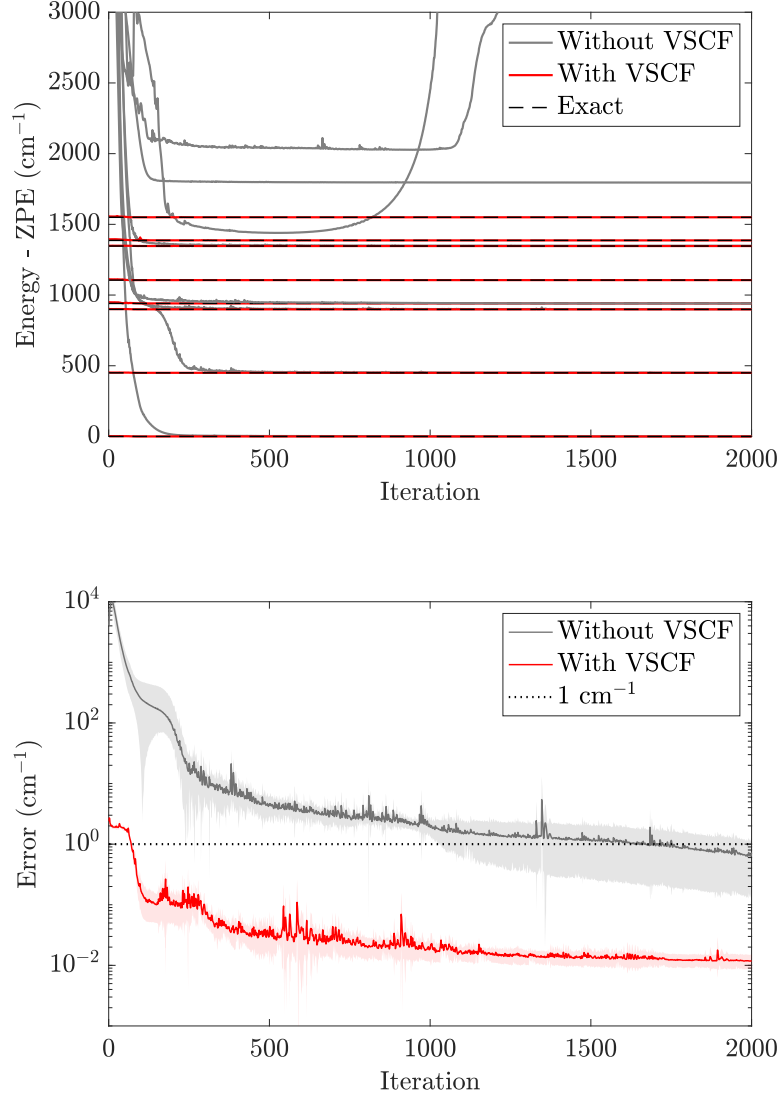


FIG. 5. (Top) Optimization of the lowest eight vibrational levels (shifted by the ZPE) of ClO_2 using an MBF with $\alpha = 2$ and $N_s = 64$. Energy levels obtained by vDMRG are plotted as reference. $N_{\text{max}} = 9$. (Bottom) Comparison of the average error of the three lowest eigenstates between optimizations with and without VSCF pretraining. The shaded area marks the range of the standard deviation of the three states.

optimizations. In the top panel of Figure 5, we show the optimization of an MBF ($\alpha = 2$ and $N_s = 64$) targeting the lowest eight eigenstates, both with and without VSCF pretraining. When the optimization starts from a random initial guess without VSCF pretraining, the optimization for the ground and first two excited states still converged, albeit slowly. However, from the third excited state onwards, the optimization tends to get trapped in

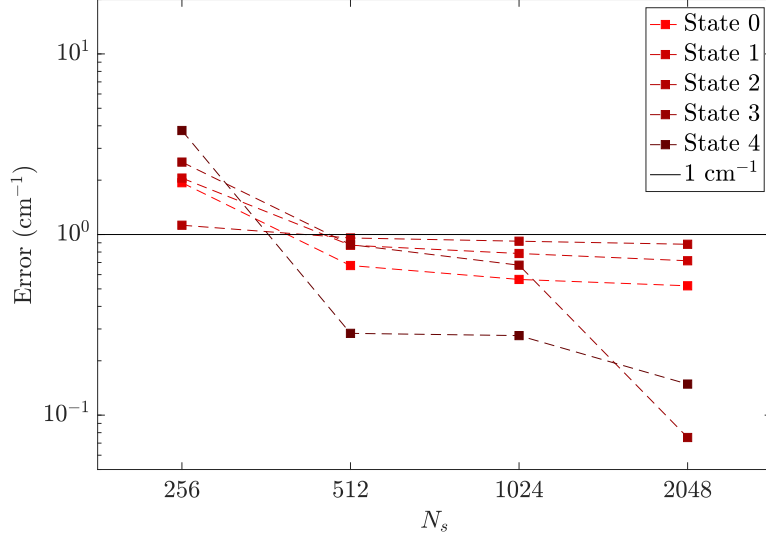


FIG. 6. Error (cm^{-1}) of the eigenstate energies of H_2CO calculated with MBF and compared to vDMRG results as N_s increases ($N_{\text{max}} = 6$).

other local minima, eventually arriving at the wrong eigenstates. The third excited state, for example, was only obtained in the fifth calculation. Some optimizations even became unstable and failed to reach any of the eight low-lying states.

When VSCF pretraining was employed, the optimization started already around the exact eigenenergies, and its improvement is hardly visible on the scale chosen. A superior set of initial modal functions given by the VSCF pretraining also ensured that we obtain each eigenstate in the correct order, which is a crucial feature for identifying all transition energies below the target threshold. For the lowest three eigenstates, for which optimizations both with and without VSCF pretraining converged, we compared the decaying behavior of the error averaged over three states in the bottom panel of Figure 5. We found that optimizing with VSCF pretraining allowed the error to reach 1 cm^{-1} accuracy much faster than in the case without pretraining, and eventually this reduced the error by about two orders of magnitude. The accuracy was well maintained from ground to excited states when VSCF pretraining was in place, indicated by the small standard deviation (pink shaded area) compared to the case without pretraining (gray shaded area).

For the larger molecules H_2CO (6 modes) and CH_3CN (12 modes) with $N_{\text{max}} = 6$ and increased anharmonicity, indicated by the anharmonic corrections of about 77 cm^{-1} for H_2CO and 67 cm^{-1} for CH_3CN , we increased the hidden neuron density α to 4 for H_2CO

and 8 for CH₃CN. We used a sequential optimization scheme that increases N_s at each stage. There were four stages in total with step counts 500, 200, 100, and 100. After each stage, the number of selected configurations N_s doubles. The initial N_s was set at 256 for H₂CO and 256 for CH₃CN. The learning rate η was set to 0.005 for both molecules. VSCF pretraining was used in all calculations.

In Figure 6, we present the absolute errors of five optimized energy levels compared to the vDMRG results for H₂CO as functions of N_s . We found that a moderate $N_s = 512$ was able to reduce the errors to less than 1 cm⁻¹ for all targeted states. The improvement in the error for the first three states is small for $N_s > 512$, while for states 3 and 4 the energy error dropped noticeably at $N_s = 2048$. In Table I, we list the ZPE ($n = 0$) and the four lowest vibrational transitions from the ground state ($n = 1, 2, 3, 4$) for both molecules. Due to the relatively uniform errors across all states and error cancellation effects, we found the transition energies for both molecules to be in good agreement with the vDMRG reference results (≤ 0.51 cm⁻¹). It is also promising to see that, although the Hilbert space dimension increases quadratically going from H₂CO to CH₃CN, doubling N_s is sufficient to achieve similar accuracy for the larger molecule CH₃CN.

In this section, we understood that, while the performance of the MBF network is satisfactory for spectroscopic accuracy, it does not exceed the accuracy or computational efficiency achieved by TNSs. A recent benchmarking study showed that TTNSs deliver a ZPE and vibrational transitions for CH₃CN with error estimates below 10⁻³ cm⁻¹. In terms of efficiency, the discrepancy between MBF and TNSs can be attributed to two issues. First, evaluating the energy of MBF is a global action. Although the selected-configuration scheme is more efficient in terms of the number of sampled states compared to the routinely used Monte Carlo method, it still requires collecting configurations in the total Fock space and evaluating their wavefunction amplitudes and local energies. By contrast, vDMRG reduces computational cost by optimizing the parameters of only one or two sites in each step. Second, gradient-based energy minimizations can be prone to problems. Although MBF is much better at learning the vibrational energies than FNN, we still had to employ various techniques (such as pretraining and additional learning rate scheduling on top of the CoRe optimizer) to ensure stable optimization. Yet, the expressiveness of the MBF network could not fully be exploited by the current optimization scheme for the two larger molecules H₂CO and CH₃CN. However, it should be emphasized that these issues in optimization

	n	MBF	DMRG	Error
H ₂ CO	0	5773.69	5773.17	+0.52
	1	1164.44	1164.25	+0.19
	2	1245.58	1245.22	+0.36
	3	1497.87	1498.31	-0.44
	4	1744.20	1743.83	+0.37
CH ₃ CN	0	9838.26	9837.41	+0.85
	1	361.14	360.99	+0.15
	2	361.14	360.99	+0.15
	3	723.28	723.18	+0.10
	4	723.69	723.18	+0.51

TABLE I. ZPE ($n = 0$) and the four lowest vibrational transitions ($n = 1, 2, 3, 4$) for H₂CO (6 modes) and CH₃CN (12 modes) calculated with MBF and compared to vDMRG reference results. The maximal N_s for H₂CO and CH₃CN were 2048 and 4096, respectively. All energies are in cm⁻¹. $N_{\max} = 6$. The hidden neuron density was $\alpha = 4$ for H₂CO and $\alpha = 8$ for CH₃CN.

are common to NQS methods and do not diminish the value of the modal backflow ansatz. Rather, the physically motivated MBF ansatz establishes a solid foundation for NQS-based vibrational structure calculations, positioning it to benefit from the ongoing advances in optimization techniques.

Another aspect for comparison is the number of free parameters: A vibrational MPS with bond dimension m contains $\mathcal{O}(LN_{\max}m^2)$ parameters, while for the MBF ansatz with a single hidden layer, the number of free parameters scales as $\mathcal{O}(\alpha L^2 N_{\max})$. Although the scaling of the number of parameters for MBF is less favorable with respect to the number of modes L , it is unclear how α or m would scale with L for a given target accuracy. For example, for the 12-mode CH₃CN a fully-converged vDMRG calculation of various eigenstates requires a bond dimension around 100,[48] while the largest hidden layer density α we used in this work was 8, which makes the ratio between the number of parameters of the two approaches (ignoring the effect of prefactors) $m^2/(\alpha L)$ quite large. Still, for the systems we studied, increasing α beyond 8 led to only marginal improvements in energy, indicating that future advances in optimization methods will be crucial to fully exploit the expressiveness of the

MBF ansatz and settle this comparison.

VI. CONCLUSIONS AND OUTLOOK

Neural quantum states (NQSs) have emerged as a versatile ansatz for solving quantum many-body Hamiltonians. While there exist recent applications of NQS in electronic structure theory, implementations of NQS for the vibrational part of the molecular problem have so far been lacking. In this work, we explored the feasibility of a tailored NQS to solve for the low-lying eigenstates of anharmonic vibrational problems.

The centerpiece of our theory is the modal backflow (MBF) NQS design, which uses modal functions that depend on the input occupation number vectors to capture anharmonicity. This tailored network significantly improved both the expressiveness and trainability compared to a conventional feedforward neural network (FNN) when applied for a Watson Hamiltonian. We implemented a selected-configuration method for the calculation of expectation values and gradients, in place of the more commonly used Monte Carlo approach, to accommodate the highly peaked amplitude distributions of typical anharmonic eigenstates.

We incorporated a pretraining step using vibrational self-consistent field calculations, which can conveniently be carried out within the MBF framework. This pretraining step was found to be instrumental in promoting robustness of the optimization, especially when targeting excited states. First, we investigated the MBF ansatz with randomly sampled force constants to mimic different levels of anharmonicity and obtained accurate zero-point vibrational energies (ZPE) across all regimes of anharmonicity. Second, we applied the MBF ansatz in calculations with Watson Hamiltonians of three molecules, including the strongly anharmonic CH_3CN . We were able to resolve both the ZPE and the low-lying vibrational transitions to spectroscopic accuracy.

Our work extends the scope of NQSs to vibrational calculations, demonstrating their potential in quantum chemistry beyond the electronic problem. Using a comparison with FNNs, we showcased the clear advantages of the modal backflow formalism and underscored the merit of embedding the features of indistinguishable particles directly into the network architecture. Despite the universal approximation theorem, the approximating power of FNNs for anharmonic vibrational problems can be harnessed only with the inclusion of the essential MBF output layer. Moreover, the effectiveness of using occupation number

vector dependent modals revealed a distinctive structure of the correlation in anharmonic vibrational eigenstates, offering potentially transferable insights for other bosonic systems.

We noted several aspects that present practical challenges for network optimization, such as the global nature of evaluating physical observables and limitations in current gradient-based optimization schemes. However, these challenges are not unique to the MBF ansatz, but rather general to the optimization of NQSs, an area of rapid development. Importantly, the physically motivated MBF ansatz provides the necessary theoretical groundwork that can be readily combined with future advances in optimization techniques, thus paving the way toward an accurate description of the vibrational structure of large molecular systems.

We envision several directions for future work. First, we can extend the modal basis from the harmonic eigenstates to general modal basis functions. The benefit of this is a potential reduction in the entanglement of the wavefunction and, hence, the number of important configurations to be selected. Second, the current paradigm of gradient-based optimization schemes struggles with the energy landscape of high-complexity networks. To fully exploit the expressiveness of NQS, more robust optimization schemes, such as second-order methods,[89] need to be explored. Finally, one can extend the current scope to include also pre-Born-Oppenheimer Hamiltonians, where fermionic and bosonic degrees of freedom of the molecule are treated on an equal footing (see, e.g., Ref. [90]) by combining the MBF network with existing NQS ansätze for electrons.

ACKNOWLEDGEMENT

This project was supported by an ETH Postdoctoral Fellowship.

-
- [1] K. D. Vogiatzis, D. Ma, J. Olsen, L. Gagliardi, and W. A. De Jong, Pushing configuration-interaction to the limit: Towards massively parallel MCSCF calculations, *J. Chem. Phys.* **147**, 184111 (2017).
 - [2] H. Gao, S. Imamura, A. Kasagi, and E. Yoshida, Distributed implementation of full configuration interaction for one trillion determinants, *J. Chem. Theory Comput.* **20**, 1185 (2024).
 - [3] C. F. Bender and E. R. Davidson, Studies in configuration interaction: The first-row diatomic hydrides, *Phys. Rev.* **183**, 23 (1969).

- [4] J. B. Schriber and F. A. Evangelista, Communication: An adaptive configuration interaction approach for strongly correlated electrons with tunable accuracy, *J. Chem. Phys.* **144**, 161106 (2016).
- [5] A. A. Holmes, N. M. Tubman, and C. Umrigar, Heat-bath configuration interaction: An efficient selected configuration interaction algorithm inspired by heat-bath sampling, *J. Chem. Theory Comput.* **12**, 3674 (2016).
- [6] W. Liu and M. R. Hoffmann, iCI: Iterative CI toward full CI, *J. Chem. Theory Comput.* **12**, 1169 (2016).
- [7] P. M. Zimmerman, Incremental full configuration interaction, *J. Chem. Phys.* **146**, 224104 (2017).
- [8] V. G. Chilkuri and F. Neese, Comparison of many-particle representations for selected configuration interaction: II. Numerical benchmark calculations, *J. Chem. Theory Comput.* **17**, 2868 (2021).
- [9] S. R. White, Density matrix formulation for quantum renormalization groups, *Phys. Rev. Lett.* **69**, 2863 (1992).
- [10] A. Baiardi and M. Reiher, The density matrix renormalization group in chemistry and molecular physics: Recent developments and new challenges, *J. Chem. Phys.* **152**, 040903 (2020).
- [11] H. Ma, U. Schollwöck, and Z. Shuai, *Density Matrix Renormalization Group (DMRG)-Based Approaches in Computational Chemistry* (Elsevier, 2022).
- [12] G. H. Booth, A. J. Thom, and A. Alavi, Fermion Monte Carlo without fixed nodes: A game of life, death, and annihilation in Slater determinant space, *J. Chem. Phys.* **131**, 054106 (2009).
- [13] J. J. Eriksen, The shape of full configuration interaction to come, *J. Phys. Chem. Lett.* **12**, 418 (2020).
- [14] S. Zhang and H. Krakauer, Quantum Monte Carlo Method using Phase-Free Random Walks with Slater Determinants, *Phys. Rev. Lett.* **90**, 136401 (2003).
- [15] J. Lee, H. Q. Pham, and D. R. Reichman, Twenty years of auxiliary-field quantum Monte Carlo in quantum chemistry: An overview and assessment on main group chemistry and bond-breaking, *J. Chem. Theory Comput.* **18**, 7024 (2022).
- [16] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017).

- [17] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning representations by back-propagating errors, *Nature* **323**, 533 (1986).
- [18] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, Backpropagation applied to handwritten zip code recognition, *Neural Comput.* **1**, 541 (1989).
- [19] D. F. Specht, A general regression neural network, *IEEE Trans. Neural Netw. Learn. Syst.* **2**, 568 (1991).
- [20] H. Robbins and S. Monro, A stochastic approximation method, *Ann. Math. Stat* **22**, 400 (1951).
- [21] B. T. Polyak, Some methods of speeding up the convergence of iteration methods, *Zh. Vych. Mat.* **4**, 1 (1964).
- [22] J. Duchi, E. Hazan, and Y. Singer, Adaptive subgradient methods for online learning and stochastic optimization, *J. Mach. Learn. Res.* **12**, 2121 (2011).
- [23] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, *arXiv:1412.6980* (2014).
- [24] G. Cybenko, Approximation by superpositions of a sigmoidal function, *Math. Control Signals Syst.* **2**, 303 (1989).
- [25] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, *Neural Netw.* **2**, 359 (1989).
- [26] O. Sharir, A. Shashua, and G. Carleo, Neural tensor contractions and the expressive power of deep neural quantum states, *Phys. Rev. B* **106**, 205136 (2022).
- [27] D.-L. Deng, X. Li, and S. Das Sarma, Quantum entanglement in neural network states, *Phys. Rev. X* **7**, 021021 (2017).
- [28] Z. Denis, A. Sinibaldi, and G. Carleo, Comment on “Can Neural Quantum States Learn Volume-Law Ground States?”, *Phys. Rev. Lett.* **134**, 079701 (2025).
- [29] D. Luo and B. K. Clark, Backflow transformations via neural networks for quantum many-body wave functions, *Phys. Rev. Lett.* **122**, 226401 (2019).
- [30] S. Manzhos, Machine learning for the solution of the Schrödinger equation, *Mach. Learn.: Sci. Technol.* **1**, 013002 (2020).
- [31] J. Hermann, Z. Schätzle, and F. Noé, Deep-neural-network solution of the electronic schrödinger equation, *Nat. Chem.* **12**, 891 (2020).

- [32] D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, Ab initio solution of the many-electron Schrödinger equation with deep neural networks, *Phys. Rev. Res.* **2**, 033429 (2020).
- [33] J. Robledo Moreno, G. Carleo, A. Georges, and J. Stokes, Fermionic wave functions from neural-network constrained hidden states, *Proc. Natl. Acad. Sci.* **119**, e2122059119 (2022).
- [34] M. Scherbela, R. Reisenhofer, L. Gerard, P. Marquetand, and P. Grohs, Solving the electronic Schrödinger equation for multiple nuclear geometries with weight-sharing deep neural networks, *Nat. Comput. Sci.* **2**, 331 (2022).
- [35] J. Hermann, J. Spencer, K. Choo, A. Mezzacapo, W. M. C. Foulkes, D. Pfau, G. Carleo, and F. Noé, Ab initio quantum chemistry with neural-network wavefunctions, *Nat. Rev. Chem.* **7**, 692 (2023).
- [36] T. Zhao, J. Stokes, and S. Veerapaneni, Scalable neural quantum states architecture for quantum chemistry, *Mach. Learn.: Sci. Technol.* **4**, 025034 (2023).
- [37] K. Choo, A. Mezzacapo, and G. Carleo, Fermionic neural-network states for ab-initio electronic structure, *Nat. Comm.* **11**, 2368 (2020).
- [38] N. Yoshioka, W. Mizukami, and F. Nori, Solving quasiparticle band spectra of real solids using neural-network quantum states, *Comm. Phys.* **4**, 106 (2021).
- [39] Q. Zhang, R.-S. Wang, and L. Wang, Neural canonical transformations for vibrational spectra of molecules, *J. Chem. Phys.* **161**, 024103 (2024).
- [40] J. M. Bowman, K. Christoffel, and F. Tobin, Application of SCF-SI theory to vibrational motion in polyatomic molecules, *J. Phys. Chem.* **83**, 905 (1979).
- [41] R. Garnier, M. Odunlami, V. Le Bris, D. Bégue, I. Baraille, and O. Coulaud, Adaptive vibrational configuration interaction (A-VCI): A posteriori error estimation to efficiently compute anharmonic IR spectra, *J. Chem. Phys.* **144**, 204123 (2016).
- [42] B. Schröder and G. Rauhut, Incremental vibrational configuration interaction theory, iVCI: Implementation and benchmark calculations, *J. Chem. Phys.* **154**, 124114 (2021).
- [43] B. Schröder and G. Rauhut, Vibrational configuration interaction theory, in *Vibrational Dynamics of Molecules* (World Scientific Publishing Company, 2022) Chap. 1, pp. 1–40.
- [44] O. Christiansen, Vibrational coupled cluster theory, *J. Chem. Phys.* **120**, 2149 (2004).
- [45] J. A. Fauchaux and S. Hirata, Higher-order diagrammatic vibrational coupled-cluster theory, *J. Chem. Phys.* **143**, 134105 (2015).

- [46] N. K. Madsen, I. H. Godtliebsen, S. A. Losilla, and O. Christiansen, Tensor-decomposed vibrational coupled-cluster theory: Enabling large-scale, highly accurate vibrational-structure calculations, *The Journal of Chemical Physics* **148**, 024103 (2018).
- [47] O. Christiansen, Vibrational coupled cluster theory, in *Vibrational Dynamics of Molecules* (World Scientific Publishing Company, 2022) Chap. 2, pp. 41–79.
- [48] A. Baiardi, C. J. Stein, V. Barone, and M. Reiher, Vibrational density matrix renormalization group, *J. Chem. Theory Comput.* **13**, 3764 (2017).
- [49] A. Baiardi, C. J. Stein, V. Barone, and M. Reiher, Optimization of highly excited matrix product states with an application to vibrational spectroscopy, *J. Chem. Phys.* **150**, 094113 (2019).
- [50] H. R. Larsson, Computing vibrational eigenstates with tree tensor network states (TTNS), *J. Chem. Phys.* **151**, 204102 (2019).
- [51] N. Glaser, A. Baiardi, and M. Reiher, Tensor network states for vibrational spectroscopy, in *Vibrational Dynamics of Molecules* (World Scientific Publishing Company, 2022) Chap. 3, pp. 80–144.
- [52] H. R. Larsson, Benchmarking vibrational spectra: 5000 accurate eigenstates of acetonitrile using tree tensor network states, *J. Phys. Chem. Lett.* **16**, 3991 (2025).
- [53] C. Puzzarini, J. Bloino, N. Tasinato, and V. Barone, Accuracy and interpretability: The devil and the holy grail. New routes across old boundaries in computational spectroscopy, *Chem. Rev.* **119**, 8131 (2019).
- [54] H. Wang and M. Thoss, Multilayer formulation of the multiconfiguration time-dependent Hartree theory, *J. Chem. Phys.* **119**, 1289 (2003).
- [55] U. Manthe, A multilayer multiconfigurational time-dependent Hartree approach for quantum dynamics on general potential energy surfaces, *J. Chem. Phys.* **128**, 164116 (2008).
- [56] O. Vendrell and H.-D. Meyer, Multilayer multiconfiguration time-dependent Hartree method: Implementation and applications to a Henon–Heiles Hamiltonian and to pyrazine, *J. Chem. Phys.* **134**, 044135 (2011).
- [57] N. Glaser, A. Baiardi, and M. Reiher, Flexible DMRG-based framework for anharmonic vibrational calculations, *J. Chem. Theory Comput.* **19**, 9329 (2023).
- [58] R. L. Fulton and M. Gouterman, Vibronic coupling. I. Mathematical treatment for two electronic states, *J. Chem. Phys.* **35**, 1059 (1961).

- [59] Q. Meng and H.-D. Meyer, A multilayer MCTDH study on the full dimensional vibronic dynamics of naphthalene and anthracene cations, *J. Chem. Phys.* **138**, 014313 (2013).
- [60] D. S. Wang and S. F. Yelin, A roadmap toward the theory of vibrational polariton chemistry, *ACS Photonics* **8**, 2818 (2021).
- [61] A. D. Dunkelberger, B. S. Simpkins, I. Vurgaftman, and J. C. Owrutsky, Vibration-cavity polariton chemistry and dynamics, *Annu. Rev. Phys. Chem.* **73**, 429 (2022).
- [62] A.-J. Liu and B. K. Clark, Neural network backflow for ab initio quantum chemistry, *Phys. Rev. B* **110**, 115137 (2024).
- [63] G. Passetti, D. Hofmann, P. Neiteimeier, L. Grunwald, M. A. Sentef, and D. M. Kennes, Can neural quantum states learn volume-law ground states?, *Phys. Rev. Lett.* **131**, 036502 (2023).
- [64] J. K. Watson, Simplification of the molecular vibration-rotation hamiltonian, *Mol. Phys.* **15**, 479 (1968).
- [65] P. Carbonniere and V. Barone, Coriolis couplings in variational computations of vibrational spectra beyond the harmonic approximation: implementation and validation, *Chem. Phys. Lett.* **392**, 365 (2004).
- [66] R. P. Feynman and M. Cohen, Energy spectrum of the excitations in liquid helium, *Phys. Rev.* **102**, 1189 (1956).
- [67] Y. Kwon, D. M. Ceperley, and R. M. Martin, Effects of three-body and backflow correlations in the two-dimensional electron gas, *Phys. Rev. B* **48**, 12037 (1993).
- [68] L. F. Tocchio, F. Becca, A. Parola, and S. Sorella, Role of backflow correlations for the nonmagnetic phase of the t - t' hubbard model, *Phys. Rev. B* **78**, 041101 (2008).
- [69] L. G. Valiant, The complexity of computing the permanent, *Theor. Comput. Sci.* **8**, 189 (1979).
- [70] O. Christiansen, A second quantization formulation of multimode dynamics, *J. Chem. Phys.* **120**, 2140 (2004).
- [71] J. M. Bowman, The self-consistent-field approach to polyatomic vibrations, *Acc. Chem. Res.* **19**, 202 (1986).
- [72] R. Gerber and M. A. Ratner, Self-consistent-field methods for vibrational excitations in polyatomic systems, *Adv. Chem. Phys.* **70**, 97 (1988).
- [73] G. Carleo, K. Choo, D. Hofmann, J. E. T. Smith, T. Westerhout, F. Alet, E. J. Davis, S. Efthymiou, I. Glasser, S.-H. Lin, M. Mauri, G. Mazzola, C. B. Mendl, E. van Nieuwenburg, O. O'Reilly, H. Th  veniaut, G. Torlai, F. Vicentini, and A. Wietek, Netket: A machine learning

- toolkit for many-body quantum systems, *SoftwareX* , 100311 (2019).
- [74] F. Vicentini, D. Hofmann, A. Szabó, D. Wu, C. Roth, C. Giuliani, G. Pescia, J. Nys, V. Vargas-Calderón, N. Astrakhantsev, and G. Carleo, NetKet 3: Machine Learning Toolbox for Many-Body Quantum Systems, *SciPost Phys. Codebases* , 7 (2022).
 - [75] W. L. McMillan, Ground state of liquid He₄, *Phys. Rev.* **138**, A442 (1965).
 - [76] W. M. Foulkes, L. Mitas, R. Needs, and G. Rajagopal, Quantum Monte Carlo simulations of solids, *Rev. Mod. Phys.* **73**, 33 (2001).
 - [77] X. Li, J.-C. Huang, G.-Z. Zhang, H.-E. Li, C.-S. Cao, D. Lv, and H.-S. Hu, A nonstochastic optimization algorithm for neural-network quantum states, *J. Chem. Theory Comput.* **19**, 8156 (2023).
 - [78] M. Eckhoff and M. Reiher, Lifelong machine learning potentials, *J. Chem. Theory Comput.* **19**, 3509 (2023).
 - [79] M. Eckhoff and M. Reiher, CoRe optimizer: an all-in-one solution for machine learning, *Mach. Learn.: Sci. Technol* **5**, 015018 (2024).
 - [80] M. T. Entwistle, Z. Schätzle, P. A. Erdman, J. Hermann, and F. Noé, Electronic excited states in deep variational Monte Carlo, *Nat. Commun.* **14**, 274 (2023).
 - [81] D. Pfau, S. Axelrod, H. Sutterud, I. von Glehn, and J. S. Spencer, Accurate computation of quantum excited states with neural networks, *Science* **385**, eadn0137 (2024).
 - [82] K. Choo, G. Carleo, N. Regnault, and T. Neupert, Symmetries and many-body excitations with neural-network quantum states, *Phys. Rev. Lett.* **121**, 167204 (2018).
 - [83] S. Pathak, B. Busemeyer, J. N. Rodrigues, and L. K. Wagner, Excited states in variational Monte Carlo using a penalty method, *J. Chem. Phys.* **154**, 034101 (2021).
 - [84] V. Wei, A. Orfi, F. Fehse, and W. A. Coish, Finding the dynamics of an integrable quantum many-body system via machine learning, *Adv. Phys. Res.* **3**, 2300078 (2024).
 - [85] M. Sibaeve and D. L. Crittenden, The PyPES library of high quality semi-global potential energy surfaces, *J. Comput. Chem.* **36**, 2200 (2015).
 - [86] D. Bégué, P. Carbonniere, and C. Pouchan, Calculations of vibrational energy levels by using a hybrid ab initio and DFT quartic force field: Application to acetonitrile, *J. Phys. Chem. A* **109**, 4611 (2005).
 - [87] G. Avila and T. Carrington, Using nonproduct quadrature grids to solve the vibrational Schrödinger equation in 12D, *J. Chem. Phys.* **134**, 054126 (2011).

- [88] K. Szenes, N. Glaser, M. Erakovic, V. Barandun, M. Mörchen, R. Feldmann, S. Battaglia, A. Baiardi, and M. Reiher, QCMAquis 4.0: Multipurpose Electronic, Vibrational, and Vibronic Structure and Dynamics Calculations with the Density Matrix Renormalization Group, *J. Phys. Chem. A* **129**, 7549 (2025).
- [89] M. Drissi, J. T. Keeble, J. Rozalén Sarmiento, and A. Rios, Second-order optimization strategies for neural network quantum states, *Philos. Trans. R. Soc. A* **382**, 20240057 (2024).
- [90] A. Muolo, A. Baiardi, R. Feldmann, and M. Reiher, Nuclear-electronic all-particle density matrix renormalization group, *J. Chem. Phys.* **152**, 204103 (2020).