

Scalable Neural Quantum State based Kernel Polynomial Method for Optical Properties from the First Principle

Wei Liu¹, Rui-Hao Bi¹, Wenjie Dou^{1,2,3*}

¹*Department of Chemistry, School of Science and Research Center for Industries of the Future, Westlake University, Hangzhou, Zhejiang 310030, China.

²Institute of Natural Sciences, Westlake Institute for Advanced Study, Hangzhou, Zhejiang 310024, China.

³Key Laboratory for Quantum Materials of Zhejiang Province, Department of Physics, School of Science and Research Center for Industries of the Future, Westlake University, Hangzhou, Zhejiang 310030, China.

*Corresponding author(s). E-mail(s): douwenjie@westlake.edu.cn;

Abstract

Variational optimization of neural-network quantum state representations has achieved FCI-level accuracy for ground state calculations, yet computing optical properties involving excited states remains challenging. In this work, we present a neural-network-based variational quantum Monte Carlo approach for ab-initio absorption spectra. We leverage parallel batch autoregressive sampling and GPU-supported local energy parallelism to efficiently compute ground states of complex systems. By integrating neural quantum ground states with the kernel polynomial method, our approach accurately calculates absorption spectra for large molecules with over 50 electrons, achieving FCI-level precision. The proposed algorithm demonstrates superior scalability and reduced runtime compared to FCI, marking a significant step forward in optical property calculations for large-scale quantum systems.

The central challenge of ab-initio quantum chemistry [1, 2] lies in accurately modeling optical properties of complex systems. These properties, such as absorption and emission spectra, depend on precise descriptions of excited states, which are significantly harder to compute than ground states. With the advent of ultra-fast pump-probe spectroscopy [3], experimental probing of excited states has become possible, increasing the demand for efficient and accurate electronic structure methods capable of handling systems with large numbers of electrons.

The absorption spectrum can be directly calculated using excited state information as [4]:

$$I(\omega) = \sum_n |\langle n | \hat{\mu} | 0 \rangle|^2 \delta(\omega_n - \omega), \quad (1)$$

where $|0\rangle$ is the ground state, ω_n and $|n\rangle$ are the n -th excited state energy and eigenvector, and $\hat{\mu}$ is the dipole moment operator. The commonly used Hamiltonian in quantum chemistry is the ab-initio electronic Hamiltonian under the Born-Oppenheimer approximation [5], given by:

$$\hat{H}_{\text{el}} = -\frac{1}{2} \sum_i \nabla_i^2 - \sum_{i,I} \frac{Z_{iI}}{|\mathbf{r}_i - \mathbf{R}_I|} + \frac{1}{2} \sum_{i \neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}. \quad (2)$$

This includes the kinetic energy of electrons, Coulomb attraction between electrons and nuclei, and Coulomb repulsion between electrons. Practical representations of \hat{H}_{el} use finite basis sets, introducing discretization errors, though stochastic methods like quantum Monte Carlo (QMC) [6] can reach the complete basis set limit. The task is to compute the ground and excited state energies and eigenvectors of \hat{H}_{el} . For a given basis set, the time-independent Schrödinger equation (TISE) [7] $\hat{H} |\Psi\rangle = E |\Psi\rangle$ exactly for \hat{H}_{el} can be solved exactly in the many-electron Hilbert space using full configuration interaction (FCI) [8] or exact diagonalization (ED) [9]. FCI provides exact absorption spectra via Eq. 1, but its exponential scaling limits applications to systems with 15 to 16 electrons, too small for most chemical systems. For larger systems, approximations like multi-reference configuration interaction (MRCI) [10] or time-dependent density functional theory (TDDFT) [11] are used, sacrificing accuracy.

Within linear response theory [12] and the Kubo formalism [13, 14], molecular absorption spectra are expressed through the Fourier transform of dipole-dipole correlation functions [15–17]:

$$I(\omega) \propto \int_{-\infty}^{\infty} C_{\hat{\mu}\hat{\mu}}(t) e^{i\omega t} dt, \quad (3)$$

where $C_{\hat{\mu}\hat{\mu}}(t) = \langle \hat{\mu}(t) \hat{\mu}(0) \rangle$ is the dipole autocorrelation function and $\hat{\mu}(t)$ is the dipole operator in Heisenberg picture. The autocorrelation function can be evaluated with the time-dependent quantum state $|\Psi(t)\rangle$, governed by the time-dependent Schrödinger equation (TDSE) [18],

$$i \frac{d}{dt} |\Psi(t)\rangle = \hat{H}(t) |\Psi(t)\rangle, \quad (4)$$

with the time-dependent Hamiltonian $\hat{H}(t)$. While ab-initio variational wave functions have recently succeeded in real-time TDSE evolution for systems like a quantum

dot with 18 electrons [19], limitations in accuracy and efficiency have restricted their broader application to many-electron systems, such as chemical molecules. By rescaling the Hamiltonian ($\hat{H} \rightarrow \tilde{H}$) and energy ($\omega \rightarrow \tilde{\omega}$), the absorption spectrum can be efficiently computed using the kernel polynomial method (KPM) [20],

$$\text{Im}[C_{\hat{\mu}\hat{\mu}}(\tilde{\omega})] = -\frac{1}{\sqrt{1-\tilde{\omega}^2}} \left(\tilde{\Omega}_0 + 2 \sum_{n=1}^{\infty} \tilde{\Omega}_n T_n(\tilde{\omega}) \right), \quad (5)$$

where $T_n(\tilde{\omega}) = \cos[n \arccos(\tilde{\omega})]$, $\tilde{\Omega}_n = g_n \Omega_n$, g_n is the Jackson kernel (Eq. 21) [21], and moments Ω_n are computed as:

$$\Omega_n = \langle 0 | \hat{\mu} T_n(\tilde{H}) \hat{\mu} | 0 \rangle. \quad (6)$$

Here, \hat{H} is the time-independent Hamiltonian, and $|0\rangle$ is its ground state of \hat{H} (see Methods). Importantly, calculating the absorption spectrum only requires solving for the ground state $|0\rangle$, which satisfies TISE. Achieving FCI-level accuracy for $|0\rangle$ directly translates to FCI-level accuracy for absorption spectra, providing significant improvements in efficiency and storage.

As discussed, FCI scales exponentially with system size, prompting modern quantum chemistry to focus on approximations that balance accuracy and computational feasibility. The Hartree–Fock (HF) method [22] treats electrons as independent particles influenced by the nuclei and a mean-field from other electrons, calculating expansion coefficients for atomic orbitals. Configuration interaction (CI) methods [23, 24] use a restricted active space built from spin orbitals to approximate electron configurations. Coupled cluster (CC) approaches [25] employ an exponential cluster operator to model electron correlation, though they occasionally yield unphysical solutions. The density matrix renormalization group (DMRG) method [26, 27] achieves FCI-level precision for molecular systems with up to 30 electrons [28, 29], but its efficiency is limited by the matrix product state ansatz, which only captures finitely correlated states [30]. Additionally, DMRG is poorly suited for large-scale parallelization. Neural quantum states (NQS) based on restricted Boltzmann machines (RBM) [31, 32] use neural networks to solve many-body quantum systems [33]. RBM has shown higher accuracy than CC with excitations up to second order CCSD and third order (CCSD(T)) for various molecular systems with up to 24 spin orbitals. Recently, autoregressive NQS architectures (NAQS) [34, 35] have been introduced for ab-initio quantum chemistry, enabling exact sampling of physical states through neural autoregressive models [36]. Subsequently, a parallelization scheme based on NAQS was implemented using a mask autoencoder based on distributed estimation (MADE) [37]. These advancements have extended NQS to systems with more than 30 electrons, surpassing RBM-based methods, but it is still limited in large-scale excited state applications.

To address these challenges, we propose a scalable NQS-KPM (sNQS-KPM) framework for ab-initio absorption spectra, enabling large-scale excited-state calculations. By combining the NQS with the KPM (Eqs. 5 and 6), sNQS-KPM achieves efficient and accurate parallel computations. Our method significantly improves computational

efficiency while maintaining accuracy comparable to ED. Using a parallel sampling strategy based on the autoregressive wavefunction ansatz, it ensures low sampling costs and balanced workloads. Additionally, we implement a batch-parallel local energy scheme with a compressed Hamiltonian representation, drastically reducing memory usage. Finally, we efficiently calculated the moments in KPM by using the recursive relationship, which avoided us from storing a large number of eigenvectors to calculate absorption spectra (Eq. 1), saving a lot of storage. These allow us to push the NQS towards the excited state era by calculating absorption spectra based on the neural quantum ground state. Experiments demonstrate its effectiveness on systems with up to 52 electrons and Hamiltonians containing millions of terms.

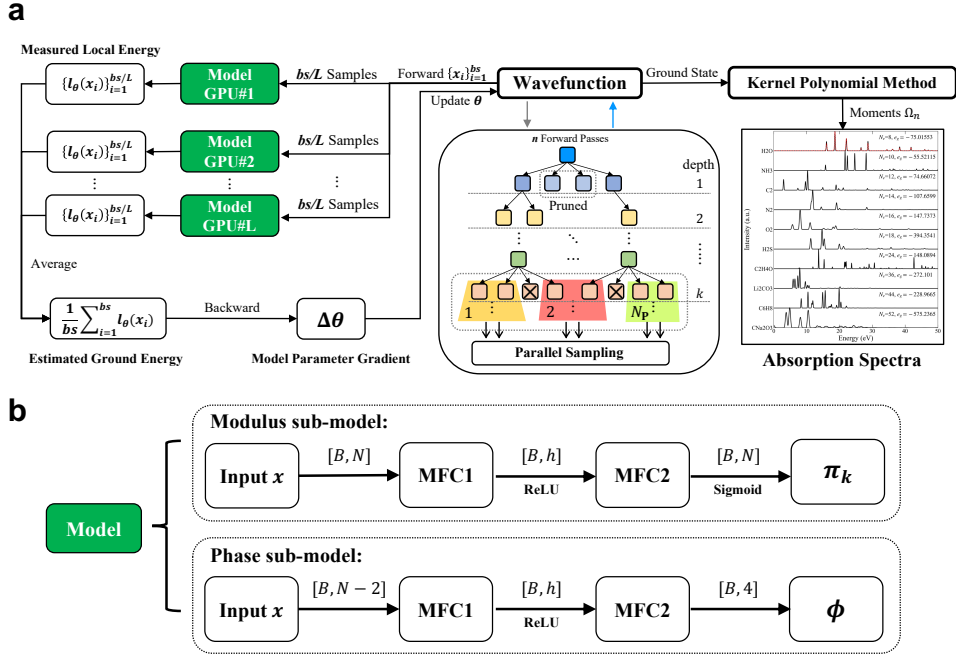


Fig. 1 Architecture of sNQS-KPM for ab-initio absorption spectra. a, Overview of the sNQS-KPM algorithm, showing parallel VQMC optimization (left), parallel autoregressive sampling (middle), and KPM (right). **b**, Neural quantum state model architecture, splitting the wavefunction output into modulus and phase components, with batch size B , dimensions N , hidden layer size h , and masked fully connected (MFC) layers. Activation functions include ReLU and Sigmoid.

1 Results

Neural-network wavefunction ansatz. The architecture of sNQS-KPM for ab-initio absorption spectra is shown in Fig. 1. We decompose the amplitude part and

phase part of the wavefunction ansatz (Fig. 1b)

$$\Psi(x, \theta) = \pi(x) e^{i\phi(x)}, \quad (7)$$

where the conditional probabilities $\pi(x) = \prod_{k=1}^N |\pi_k(x_k | x_{k-1}, \dots, x_1)|$ define a joint probability function $\pi : \{0, 1\}^N \rightarrow [0, \infty)$ for input configuration strings x . The phase $\phi : \{0, 1\}^N \rightarrow [0, 2\pi)$ of the wavefunction directly by feeding the input x into a two-layer multilayer perceptron (MLP). $\theta \in \mathbb{R}^d$ denotes the concatenation of parameters describing the neural networks π and ϕ . Our ansatz is autoregressive since only the amplitude contributes to the probability $|\pi(x)|^2$ and the amplitude part is autoregressive by design, therefore we can use very efficient sampling algorithms designed for autoregressive models.

The autoregressive sampling generates one sample per run, performing N local samplings in the process. To improve efficiency, Ref. [34] introduces the batch autoregressive sampling (BAS) algorithm, which generates a batch of samples at each local sampling step instead of a single sample. Specifically, in the first step, N_s samples (potentially as large as 10^{12}) are generated from the probability distribution $\pi(x)$, retaining only unique samples with their weights. At the k -th step, the conditional probability $\pi_k(x_k | x_{k-1}, \dots, x_1)$ is computed for each unique sample from the previous step (weighted by w_{x_{k-1}, \dots, x_1}), and local samplings are performed to generate exactly w_{x_{k-1}, \dots, x_1} samples. Intermediate samples with zero weights are pruned at each layer, ultimately producing N_s samples at the N -th step, stored as N_u unique samples with weights. The BAS algorithm, illustrated in Fig. 1a (middle), has a computational cost of $\sum_{n=1}^N k^2 N_{u,k}$, where $N_{u,k}$ is the number of unique samples at the k -th step. Assuming $N_{u,k} \approx N_u$ (typically $N_u < 10^6$ for molecular systems with $N > 100$), the total cost of BAS scales as $O(N_u N^3/3)$, independent of N_s .

Parallel sampling. The first challenge for NQS lies in the sampling process. Autoregressive neural networks enable exact sampling without pre-thermalization or discarding intermediate samples to avoid auto-correlations, making them more efficient than traditional Monte Carlo methods. The BAS process can be visualized as a tree, with each layer representing a local sampling step [35], as shown in Fig. 1a (middle). To parallelize BAS, we first perform serial sampling across all processes using the same random seed for the first k steps, ensuring identical samples on each process. At the k -th step, unique samples are divided among N_p processes, balancing the number of samples per process. The value of k is dynamically determined by setting a threshold N_u^* and choosing the first step where the number of unique samples ($N_{u,k}$) exceeds N_u^* . This parallel sampling scheme allows for the computation of molecules with more electrons compared to other NQS methods (see Table 1).

At each step, leaves with zero occurrence are pruned. Additionally, a number-conserving constraint reduces the sample space by preserving the total spin-up (N_\uparrow) and spin-down (N_\downarrow) electron counts separately. During the Jordan-Wigner transformation, the i -th spatial orbital maps to two qubits at positions $2i - 1$ and $2i$, representing spin-up and spin-down orbitals. For each sampled pair $x_i = s_{2i} s_{2i-1}$ ($s_i \in \{0, 1\}$), the constraint is imposed by regularizing the local conditional probabilities

$\pi_i(x_i|x_{i-1}, \dots, x_1)$ as [35, 37]:

$$\tilde{\pi}_i(x_i|x_{i-1}, \dots, x_1) = \begin{cases} 0, & \text{if } \sum_{j=1}^i s_{2j-1} > N_{\uparrow} \\ 0, & \text{if } \sum_{j=1}^i s_{2j} > N_{\downarrow} \\ \pi_i(x_i|x_{i-1}, \dots, x_1), & \text{o.w.} \end{cases} \quad (8)$$

The adjusted probabilities are normalized as $\tilde{\pi}_i(x_i|x_{i-1}, \dots, x_1) / \sum_i \tilde{\pi}_i(x_i|x_{i-1}, \dots, x_1)$.

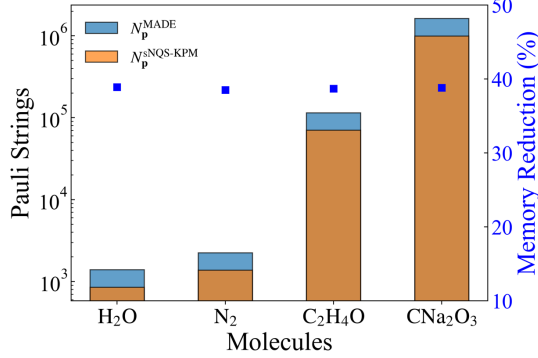


Fig. 2 Memory efficiency of sNQS-KPM. Comparison of memory requirements for storing the Hamiltonian between sNQS-KPM and MADE from Ref. [37] across various representative molecular systems in the STO-3G basis set.

GPU-supported local energy calculation. The second challenge in NQS for electronic structure calculations is evaluating the local energy in Eq. 18, which is computationally expensive for molecular systems due to the $O(N^4)$ scaling of Pauli strings and memory issues from $O(N^4)$ possible \mathbf{x}' coupled to an input \mathbf{x} . To address this, we implemented an efficient tensor representation of the second quantized spin Hamiltonian derived from chemical data, leveraging the sparsity of the matrix corresponding to a Pauli operator product $\hat{P}_{\mathbf{p}}$. Specifically, for each row index $x \in \{0, 1\}^N$, there is exactly one column index $x^{\mathbf{p}} \in \{0, 1\}^N$ with a nonzero matrix entry $\langle x | \hat{P}_{\mathbf{p}} | x^{\mathbf{p}} \rangle$. Denoting the subset of Pauli strings with nonzero coefficients as $S = \{\mathbf{p} \in \{I, x, y, z\} : h_{\mathbf{p}} \neq 0\}$, the local energy simplifies to:

$$l_{\theta}(x) = \sum_{\mathbf{p} \in S} h_{\mathbf{p}} \langle x | \hat{P}_{\mathbf{p}} | x^{\mathbf{p}} \rangle \frac{\Psi(x^{\mathbf{p}}, \theta)}{\Psi(x, \theta)}. \quad (9)$$

To efficiently compute this large summation $K = |S|$, we developed a GPU-compatible approach outlined in Algorithm 1. Key information from the molecular Hamiltonian, such as Pauli string indices and coefficients, is stored as tensors supporting GPU computation, as shown in Fig. 1a. A string parser constructs an incidence matrix with coefficients, tracking indices of Pauli operators ($\sigma_x, \sigma_y, \sigma_z$) as tensors for efficient computation of $x^{\mathbf{p}}$ and the matrix element $\langle x | \hat{P}_{\mathbf{p}} | x^{\mathbf{p}} \rangle$.

Algorithm 1 Parallel tensor computation of local energy (batch size = B).

- 1: **Input:** Batch of bit strings ($\{x_b\} \in \{0, 1\}^{B \times N}$), coefficient vector ($h_{\mathbf{p}}$), unique incidence matrix ($M_{\mathbf{u}}^{\{x,y\}} \in \{0, 1\}^{K \times N}$), incidence matrix ($M^{\{y,z\}} \in \{0, 1\}^{K \times N}$)
 - 2: **Output:** Batch of local energies $\{E_b\} \in \mathbb{C}^B$
 - 3: **Parallel for** $b = 1$ to B **in parallel**
 - 4: Compute $x_b^{\mathbf{p}} \in \{0, 1\}^{K \times N}$ as $x_b^{\mathbf{p}} = X_b \oplus M_{\mathbf{u}}^{\{x,y\}}$, where $X_b \in \{0, 1\}^{K \times N}$ (K -fold replication of x_b)
 - 5: Compute amplitudes $\Psi(x_b) \in \mathbb{C}$ and $\{\Psi(x_b^{\mathbf{p}})\} \in \mathbb{C}^K$
 - 6: Compute $\left\{ \left\langle x_b | \hat{P}_{\mathbf{p}} | x_b^{\mathbf{p}} \right\rangle \right\} \in \mathbb{C}^K$ using Eq. 10
 - 7: Evaluate the sum in Eq. 9 to get E_b
 - 8: **End parallel for**
-

In practice, $x^{\mathbf{p}}$ is computed by flipping the bits of x based on the positions of σ_x and σ_y in $P_{\mathbf{p}}$. To enhance efficiency, the indices of σ_x , σ_y operators ($M^{\{x,y\}}$) are pre-collected, storing only unique flips ($M_{\mathbf{u}}^{\{x,y\}}$) and reorganizing related indices ($M^{\{y,z\}}$) and their occurrence counts (r) [35]. The reorganization step is pre-computed and incurs negligible cost. This scheme reduces memory usage by approximately 40% compared to MADE (see Fig. 2), effectively shrinking the input size for the forward pass by eliminating redundant samples, which is especially beneficial for large-scale problems. The matrix element $\langle x | \hat{P}_{\mathbf{p}} | x^{\mathbf{p}} \rangle$ is given by:

$$\langle x | \hat{P}_{\mathbf{p}} | x^{\mathbf{p}} \rangle = \left\langle x \left| \prod_{i=1}^N \sigma_{p_i} \right| x^{\mathbf{p}} \right\rangle = (-i)^r \prod_{k:p_k \in \{y,z\}} (-1)^{x_k} \quad (10)$$

where r is the count of σ_y occurrences ($r = \sum_{k=1}^n \mathbf{1}_{\{p_k=y\}}$). Similar to before, the indices of σ_y , σ_z operators are pre-collected, and the product is efficiently computed via the Hadamard product of the indices with x , followed by a reduction of entries. All computations are parallelized for GPU execution [34, 35, 37].

Absorption spectra of different molecules. To validate the precision of our sNQS-KPM approach, we first compute the ground-state energies of several small-scale molecular systems and compare them with existing NQS results, as summarized in Table 1. NAQS can handle molecules with up to 18 electrons, while MADE extends this limit to 36 electrons. In comparison, sNQS-KPM significantly pushes the boundary, enabling calculations for molecules with up to 52 electrons. The mean absolute errors (MAE) for each method are reported. Our method achieves comparable accuracy to NAQS (and surpasses it for C_2 , N_2 , and O_2), while consistently outperforming MADE. This level of precision ensures FCI-level accuracy for the absorption spectra calculations, even in cases where direct FCI computation is infeasible.

We show the performance of sNQS-KPM in calculating absorption spectra for various molecules sorted by electron count, as shown in Fig. 3. For H_2O and NH_3 , FCI absorption spectra serve as baselines, derived by diagonalizing the smallest 1000 eigenvalues and eigenvectors of the FCI Hamiltonian matrix (out of 2^n , where n is

the number of orbitals) due to computational limits. As illustrated in Figs. 3a and 3b, sNQS-KPM accurately reproduces the peak positions of absorption spectra with high fidelity to FCI results. Minor discrepancies in relative peak heights arise from the limited number of eigenstates used in FCI, which decrease as more eigenstates are included.

Our method exhibits robust performance for molecules with 12 to 52 electrons, including C_2 , N_2 , O_2 , H_2S , C_2H_4O , Li_2CO_3 , C_6H_8 , and CNa_2O_3 . As molecule size increases, the rapidly growing number of terms in the electronic Hamiltonian poses challenges in computational cost and memory usage. Despite this, our algorithm efficiently scales to systems with up to 52 electrons, handling Hamiltonians with up to 1.6 million terms, while achieving state-of-the-art excited-state performance. Notably, in Fig. 3j, a peak near relative energy of 0 indicates a degenerate ground state for CNa_2O_3 .

Table 1 Ground state energies (in Hartree) calculated by our method (sNQS-KPM). The conventional HF and CCSD results, FCI results as well as existing NQS results including NAQS [34] and MADE [37] are shown for comparison. N_e the total number of electrons (including spin up and spin down). The mean absolute errors (MAE) for various methods compared with FCI are also listed.

Molecule	N_e	HF	CCSD	NAQS [34]	MADE [37]	Ours	FCI
H_2O	8	-74.9644	-75.0154	-75.0155	-75.0155	-75.0155	-75.0155
NH_3	10	-55.4548	-55.5209	-55.5212	-55.5212	-55.5212	-55.5212
C_2	12	-74.4209	-74.6745	-74.6889	-74.6876	-74.6908	-74.6908
N_2	14	-107.4990	-107.6561	-107.6595	-107.6567	-107.6601	-107.6602
O_2	16	-147.5513	-147.7027	-147.7127	-147.7106	-147.7133	-147.7133
H_2S	18	-349.3114	-394.3483	-394.3546	-394.3545	-394.3546	-394.3546
C_2H_4O	24	-148.0456	-148.0798	—	-148.0827	-148.0894	—
Li_2CO_3	36	-272.0976	-272.0998	—	-272.1003	-272.1011	—
C_6H_8	44	-228.9263	-228.9586	—	—	-228.9665	—
CNa_2O_3	52	-575.0161	-575.2295	—	—	-575.2365	—
MAE(Hartree)				5.3×10^{-4}	1.6×10^{-3}	1.7×10^{-4}	

Resolution of absorption spectra. Our sNQS-KPM calculates the absorption spectrum using Chebyshev polynomial expansion. The number of polynomials required for the expansion is directly related to the resolution of the absorption spectrum. As discussed in Methods, to expand the δ function using Chebyshev polynomials, the eigenvalues of the Hamiltonian are typically rescaled to the range $[-1, 1]$. This rescaling is necessary because Chebyshev polynomials within this range exhibit excellent properties such as orthogonality and recurrence relations, which enhance the stability and efficiency of the calculations. However, different molecules have different ranges of eigenvalues for their Hamiltonians, such as $[a, b]$. This necessitates different rescaling ranges. For instance, the eigenvalue range of the Hamiltonian of H_2O is much smaller than that of CNa_2O_3 . To achieve the same final absorption spectral resolution for both molecules, CNa_2O_3 requires a significantly larger number of moments (N_Ω , equivalent to the number of polynomials) compared to H_2O . In Figs. 3a and 3j, H_2O used

60,000 moments, while CNa_2O_3 used 200,000 moments. Despite this, the resolution of CNa_2O_3 appears slightly lower than that of H_2O .

We report the convergence of the absorption spectral resolution as a function of N_Ω in Fig. 4a. We gradually increased N_Ω from 10,000 to 200,000, and the absorption spectrum of H_2S evolved from a single peak to five distinct peaks within the energy range of 10 to 22 eV. To ensure convergence, we propose an empirical formula:

$$N_\Omega = 400|e_g|, \quad (11)$$

where e_g is the numerical value of the molecular ground state energy in atomic units, calculated using NQS. The computation time for the moments is approximately linearly related to the number of moments, as shown in Fig. 3b.

Computational complexity. We evaluated the time complexity of our method using H_2 chains, as shown in Fig. 3c. The tests were performed on a single V100 GPU, with 10,000 training steps using NQS and 10,000 moments calculated via KPM. By fitting the computation times for H_2 chains containing 5 to 15 H_2 molecules, we found that the NQS, KPM, and overall calculation times scale approximately $O(N^4)$ with the number of molecules in the chains. When the number of H_2 in the chain increases, the computational efficiency of sNQS-KPM is much higher than that of ED. Specifically, the time for ED to calculate the chain $(\text{H}_2)_5$ is already greater than that for sNQS-KPM to calculate the chain $(\text{H}_2)_{16}$.

To assess the effectiveness of our parallelization scheme, we examined the running time for the O_2 molecule using 10^3 iterations across varying batch sizes, as illustrated in Fig. 3d. The results show that the running time scales inversely with the number of GPUs, with doubling the GPUs reducing the time by roughly half. Furthermore, experiments with memory-saturated GPUs confirmed that the running time remains consistent across different GPU counts, demonstrating near-optimal weak scaling of our approach.

2 Discussion

This work presents a scalable neural quantum state-based kernel polynomial method (sNQS-KPM) for ab-initio absorption spectra, capable of achieving FCI-level accuracy even for systems with up to 52 electrons, where traditional FCI calculations become computationally infeasible. By integrating neural quantum states (NQS) with the kernel polynomial method (KPM), this approach leverages Chebyshev polynomial expansions to compute absorption spectra with high precision, while mitigating truncation errors through kernel smoothing techniques.

Key innovations of this work include embedding physical priors into the network architecture to enhance the accuracy of quantum system modeling and implementing an efficient sampling strategy that scales with the number of unique configurations rather than the total batch size, thereby reducing computational overhead. Additionally, parallel computation of local energy across multiple GPUs is employed, distributing term-wise calculations without inter-term interactions and leveraging embarrassingly parallel algorithms to improve scalability for Hamiltonians with many

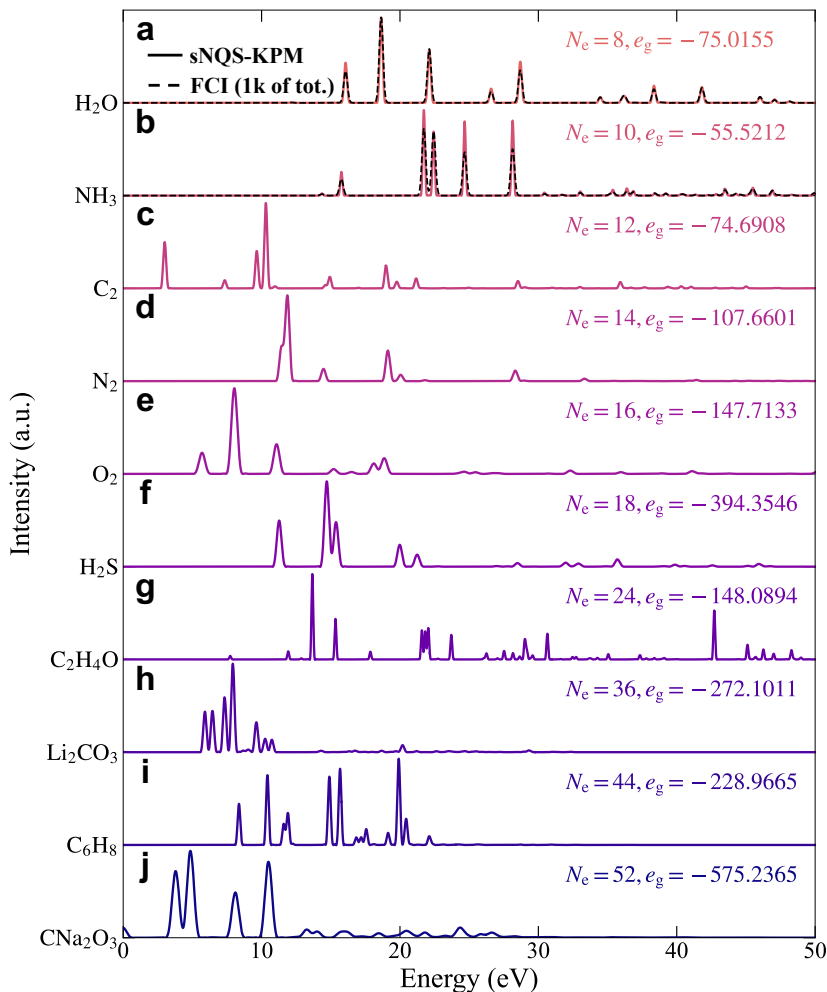


Fig. 3 Absorption spectra of various molecules computed using sNQS-KPM. Molecules are arranged in order of increasing electron count. The black dotted line represents the absorption spectra of H₂O and NH₃, obtained by diagonalizing the smallest 1000 eigenvalues and eigenvectors of the FCI Hamiltonian matrix (dimension 2^n , where n is the number of orbitals) on the STO-3G basis using the Lanczos algorithm.

terms. Furthermore, the moments in KPM are efficiently computed using recursive relationships, eliminating the need to store large numbers of eigenvectors for absorption spectra calculations, resulting in significant storage savings. Collectively, these advancements greatly extend the utility of ML-based methods for excited-state calculations in second-quantized quantum chemistry [32, 34, 35, 37].

Despite these advances, second-quantized methods are inherently constrained by the choice of basis set, posing challenges in scaling to larger basis sets akin to scaling to larger molecules. In contrast, first-quantized approaches, such as FermiNet [38]

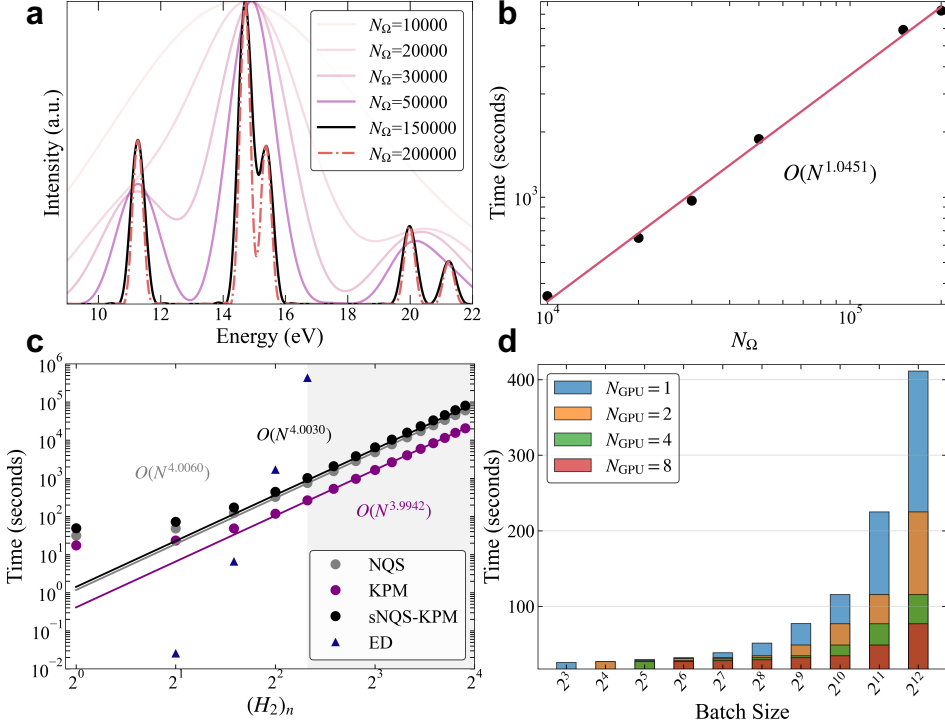


Fig. 4 Performance benchmark of sNQS-KPM. **a**, Convergence of absorption spectral resolution for H₂S as a function of the number of moments (N_{Ω}). **b**, Relationship between computation time and the number of moments (N_{Ω}) calculated. **c**, Comparison of computational time complexity for NQS, KPM, and the overall sNQS-KPM approach against exact diagonalization (ED) on extended H₂ chains. The benchmark is performed on a single V100 GPU, with NQS trained for 10^4 steps and KPM calculating 10^4 moments. **d**, Demonstration of near-optimal weak scaling for the O₂ molecule, running the algorithm for 10^3 iterations with varying batch sizes up to 2^{12} and reporting the elapsed time in seconds. Distributing batches across multiple GPUs significantly reduces runtime, while training with a fixed batch size per GPU maintains consistent runtimes across different configurations.

and PauliNet [39], employ flexible basis sets and fewer determinants, achieving unconstrained accuracies at the cost of increased computational complexity. These methods rely on Markov chain Monte Carlo sampling, but their compatibility with second-quantized frameworks remains an open avenue for future investigation.

3 Methods

Mapping fermionic Hamiltonians to the qubits. We consider the second-quantized molecular electronic Hamiltonian [40]:

$$\hat{H} = \sum_{pq} h_{pq} \hat{a}_p^\dagger \hat{a}_q + \frac{1}{2} \sum_{pqrs} h_{pqrs} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_s, \quad (12)$$

where \hat{a}_i^\dagger (\hat{a}_i) are fermionic creation (annihilation) operators for the i -th orbital, and h_{pq} , h_{pqrs} are the one- and two-body integrals, respectively. Using the Jordan-Wigner encoding [41], these operators are mapped to qubit operators:

$$\hat{a}_j^\dagger \rightarrow \left(\prod_{i=0}^{j-1} \sigma_i^z \right) \sigma_j^+, \quad \hat{a}_j \rightarrow \left(\prod_{i=0}^{j-1} \sigma_i^z \right) \sigma_j^-, \quad (13)$$

where the occupation state of the j -th fermionic mode is stored locally in the j -th qubit using $\sigma_j^\pm = \sigma_j^x \pm i\sigma_j^y$, while the parity is stored non-locally. Alternative mappings, such as parity [42] or Bravyi–Kitaev [43] encodings, would require specific adjustments to encode physical priors but do not change the general form of the qubit Hamiltonian:

$$\hat{H}_Q = \sum_{\mathbf{p} \in \{I,x,y,z\}^N} h_{\mathbf{p}} \hat{P}_{\mathbf{p}}, \quad (14)$$

where $\hat{P}_{\mathbf{p}} := \prod_{i=1}^N \sigma_{p_i}$ is a tensor product of Pauli matrices, and the coefficients $h_{\mathbf{p}} = \frac{1}{2^N} \text{Tr}(\hat{H}_Q \hat{P}_{\mathbf{p}})$ are real-valued.

Variational Monte Carlo optimization. Following the standard neural-network variational Monte Carlo approach [32], we define trial wavefunctions whose complex amplitudes relative to the standard basis are computed by a neural network, parametrized by $\theta \in \mathbb{R}^d$. Given a function

$$\Psi : \{0, 1\}^N \times \mathbb{R}^d \rightarrow \mathbb{C}, \quad (15)$$

which is differentiable in θ , the associated family of trial quantum states is expressed as:

$$|\Psi_\theta\rangle = \sum_{x \in \{0,1\}^N} \Psi(x, \theta) |x\rangle, \quad (16)$$

where $|x\rangle := \prod_{i=1}^N |x_i\rangle$ represents the standard basis vectors. The wavefunction $\Psi(x, \theta)$ is normalized for all θ , and sampling from the probability distribution $\pi_\theta := |\Psi(x, \theta)|^2$ is computationally feasible.

Using $|\Psi_\theta\rangle$ as a trial vector in the Rayleigh quotient for \hat{H}_Q , we obtain a differentiable objective function:

$$\mathcal{E}(\theta) = \frac{1}{2} \langle \Psi_\theta | \hat{H}_Q | \Psi_\theta \rangle, \quad (17)$$

which upper bounds the minimal eigenvalue $\lambda_{\min}(\hat{H}_Q)$ per the Rayleigh–Ritz principle. The objective can be estimated via Monte Carlo sampling:

$$\mathcal{E}(\theta) = \frac{1}{2} \mathbb{E}[l_\theta(x)], \quad l_\theta(x) = \frac{\langle x | \hat{H}_Q | \Psi_\theta \rangle}{\Psi(x, \theta)}, \quad x \sim \pi_\theta, \quad (18)$$

where $l_\theta(x)$, the local energy, is crucial for both updating θ and estimating $\mathcal{E}(\theta)$. Stochastic gradient optimization minimizes \mathcal{E} , with gradients estimated as:

$$\nabla\mathcal{E}(\theta) = \text{Re}\mathbb{E}[(l_\theta(x)\mathbf{1} - \mathbf{b})\bar{\sigma}_\theta(x)], \quad \sigma_\theta(x) = \frac{\nabla_\theta\Psi(x, \theta)}{\Psi(x, \theta)}, \quad x \sim \pi_\theta, \quad (19)$$

where \mathbf{b} is a baseline matrix, often chosen as $\mathbf{b} = \mathbb{E}[l_\theta(x)]$.

Computing $x \rightarrow l_\theta(x)$ for a minibatch of size B has complexity $O(BK)$, where K is the number of terms in the Hamiltonian expansion (Eq. 14). Although $K \ll 4^N$, the sparsity does not prevent out-of-memory (OOM) issues for large molecules, as 4^N grows exponentially with N . For instance, Sodium Carbonate ($N = 76$) has $4^N \approx 5.7 \times 10^{45}$ terms, while $K = 1,625,991$. With a modest batch size $B = 1024$, computing local energy involves a forward pass of 1.7 billion samples with input size 76, creating a significant computational bottleneck. Large batch sizes are essential for accurate Monte Carlo approximations of $\mathcal{E}(\theta)$, directly impacting algorithm performance. Thus, efficient sampling and local energy computation are critical for scaling neural-network variational methods to large molecules.

Parallel algorithm implementation. In quantum chemistry applications, even small molecules can have Hamiltonians with thousands of terms, causing severe OOM issues on existing VMC platforms. To address this, our pipeline tensorizes the molecular Hamiltonian to optimize memory usage. Local energy is computed term-wise with no inter-term interaction, enabling embarrassingly parallel algorithms for Hamiltonians with many terms. We tackle this bottleneck by applying a sampling parallelization strategy where identical model copies on multiple GPUs generate a few samples per unit. These independent samples are then combined to construct an accurate expectation estimate. Additionally, duplicate configurations are removed locally on each GPU before the forward pass to further reduce memory usage.

The energy expectation is approximated as:

$$\frac{\langle \Psi_\theta | \hat{H}_Q | \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle} \approx \frac{1}{B} \sum_{i=1}^B \sum_{\mathbf{p} \in S} h_{\mathbf{p}} \langle x_i | P_{\mathbf{p}} | x_i^{\mathbf{p}} \rangle \frac{\Psi(x_i^{\mathbf{p}}, \theta)}{\Psi(x_i, \theta)}. \quad (20)$$

We distribute the BK summands across L GPUs, each handling a partial sum of size BK/L . Each GPU computes local gradients using forward and backward passes, and model parameters are updated by averaging these local gradients. To avoid OOM issues for larger molecules, we use gradient accumulation [44], splitting batches into smaller mini-batches before each update.

KPM for absorption spectra. The KPM approximates physical quantities using polynomial expansions, with Chebyshev polynomials being the most commonly applied due to their efficiency. To mitigate truncation errors, modified kernels like the Jackson kernel [20] are used, ensuring smooth approximations of functions such as the Dirac delta:

$$g_n = \frac{1}{N+1} \left[(N-n-1) \cos \frac{\pi n}{N+1} + \sin \frac{\pi n}{N+1} \cot \frac{\pi}{N+1} \right]. \quad (21)$$

For two operators A and B , the dynamical correlation function is defined as:

$$\text{Im}[C_{AB}(\omega)] = \lim_{\epsilon \rightarrow 0} \langle 0 | A \frac{1}{\omega + i\epsilon + H} B | 0 \rangle, \quad (22)$$

which, assuming real product $\langle 0 | A | k \rangle \langle k | B | 0 \rangle$, simplifies to:

$$C_{AB}(\omega) = -\pi \sum_{k=0}^{D-1} \langle 0 | A | k \rangle \langle k | B | 0 \rangle \delta(\omega + E_k), \quad (23)$$

where $|k\rangle$ and E_k are the eigenstates and eigenvalues of H , and $|0\rangle$ is the ground state. After rescaling $H \rightarrow \tilde{H}$ and $\omega \rightarrow \tilde{\omega}$, $\text{Im}[C_{AB}(\omega)]$ is expanded in Chebyshev polynomials:

$$\text{Im}[C_{AB}(\omega)] = -\frac{1}{\sqrt{1 - \tilde{\omega}^2}} \left(\tilde{\Omega}_0 + 2 \sum_{n=1}^{\infty} \tilde{\Omega}_n T_n(\tilde{\omega}) \right), \quad (24)$$

where $\tilde{\Omega}_n = g_n \Omega_n$ and moments Ω_n are computed as:

$$\Omega_n = \langle 0 | A T_n(\tilde{H}) B | 0 \rangle. \quad (25)$$

For dipole autocorrelation functions, with $\hat{\mu} = \hat{\mu}^\dagger$, moments are:

$$\Omega_n = \langle 0 | \hat{\mu}^\dagger T_n(\tilde{H}) \hat{\mu} | 0 \rangle = \langle \alpha | T_n(\tilde{H}) | \alpha \rangle, \quad (26)$$

where $|\alpha\rangle = \hat{\mu} | 0 \rangle$. The dipole operator $\hat{\mu}$, in second-quantized form, is:

$$\hat{\mu} = - \sum_{pq} d_{pq} [\hat{c}_{p\uparrow}^\dagger \hat{c}_{q\uparrow} + \hat{c}_{p\downarrow}^\dagger \hat{c}_{q\downarrow}] - \hat{\mu}_c, \quad (27)$$

with d_{pq} derived from the integral of the position operator $\hat{\mathbf{r}}$ over molecular orbitals ϕ :

$$d_{pq} = \int dr \phi_p^*(r) \hat{\mathbf{r}} \phi_q(r), \quad (28)$$

and core contributions $\hat{\mu}_c$:

$$\hat{\mu}_c = 2 \sum_{i=1}^{N_{\text{core}}} d_{ii}. \quad (29)$$

Starting from $|\alpha\rangle$, states $|\alpha_n\rangle = T_n(\tilde{H}) |\alpha\rangle$ are iteratively constructed using the recursion:

$$\begin{aligned} |\alpha_0\rangle &= |\alpha\rangle, \\ |\alpha_1\rangle &= \tilde{H} |\alpha_0\rangle, \\ |\alpha_{n+1}\rangle &= 2\tilde{H} |\alpha_n\rangle - |\alpha_{n-1}\rangle. \end{aligned} \quad (30)$$

Moments are then computed as:

$$\begin{aligned}\Omega_{2n} &= 2 \langle \alpha_n | \alpha_n \rangle - \langle \alpha_0 | \alpha_0 \rangle, \\ \Omega_{2n+1} &= 2 \langle \alpha_{n+1} | \alpha_n \rangle - \langle \alpha_1 | \alpha_0 \rangle.\end{aligned}\tag{31}$$

Training details. Training follows the standard variational Monte Carlo approach described in the main text. Parameters are optimized using the Adam algorithm [45] with an initial learning rate of 5×10^{-3} , reduced to 5×10^{-4} halfway through. The batch size, initially set to 10^6 , is dynamically adjusted to maintain 10,000–100,000 unique, physically valid samples, with a maximum batch size of 10^{12} [34].

4 Data availability

The exact molecular data generated, along with a notebook to reproduce these steps, can be found in the supporting code at <https://github.com/Weithesgmt/sNQS-KPM>.

5 Code availability

Source code to reproduce the reported results can be found at <https://github.com/Weithesgmt/sNQS-KPM>.

6 Competing interests

The authors declare no competing interests.

7 Acknowledgments

W.D. acknowledges the support from National Natural Science Foundation of China (No. 22361142829 and No. 22273075) and Zhejiang Provincial Natural Science Foundation (No. XHD24B0301). W.L. thanks Yu Wang for useful discussions.

References

- [1] Friesner, R.A.: Ab initio quantum chemistry: Methodology and applications. *Proceedings of the National Academy of Sciences* **102**(19), 6648–6653 (2005)
- [2] Hermann, J., Spencer, J., Choo, K., Mezzacapo, A., Foulkes, W.M.C., Pfau, D., Carleo, G., Noé, F.: Ab initio quantum chemistry with neural-network wavefunctions. *Nature Reviews Chemistry* **7**(10), 692–709 (2023)
- [3] Bv, D.C., Hydraulics, D.: Absorption spectroscopy. *Methods in Enzymology* vol **5002011** (2000)
- [4] Zuehlsdorff, T.J., Isborn, C.M.: Modeling absorption spectra of molecules in solution. *International Journal of Quantum Chemistry* **119**(1), 25719 (2019)

- [5] Born, M., Oppenheimer, R.: Zur quantentheorie der molekeln. *Annalen der Physik* **389**, 457–484 (1927)
- [6] Foulkes, W.M., Mitas, L., Needs, R., Rajagopal, G.: Quantum monte carlo simulations of solids. *Reviews of Modern Physics* **73**(1), 33 (2001)
- [7] Szabo, A., Ostlund, N.S.: *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Courier Corporation, New York (1996)
- [8] Knowles, P.J., Handy, N.C.: A new determinant-based full configuration interaction method. *Chemical physics letters* **111**(4-5), 315–321 (1984)
- [9] Weiße, A., Fehske, H.: Exact diagonalization techniques. *Computational many-particle physics*, 529–544 (2008)
- [10] Szalay, P.G., Muller, T., Gidofalvi, G., Lischka, H., Shepard, R.: Multiconfiguration self-consistent field and multireference configuration interaction methods and applications. *Chemical reviews* **112**(1), 108–181 (2012)
- [11] Marques, M.A., Gross, E.K.: Time-dependent density functional theory. *Annu. Rev. Phys. Chem.* **55**(1), 427–455 (2004)
- [12] Kubo, R., Toda, M., Hashitsume, N., Kubo, R., Toda, M., Hashitsume, N.: Statistical mechanics of linear response. *Statistical Physics II: Nonequilibrium Statistical Mechanics*, 146–202 (1991)
- [13] Kubo, R.: Statistical-mechanical theory of irreversible processes. i. general theory and simple applications to magnetic and conduction problems. *Journal of the physical society of Japan* **12**(6), 570–586 (1957)
- [14] Kubo, R., Yokota, M., Nakajima, S.: Statistical-mechanical theory of irreversible processes. ii. response to thermal disturbance. *Journal of the Physical Society of Japan* **12**(11), 1203–1211 (1957)
- [15] Kubo, R.: The fluctuation-dissipation theorem. *Reports on progress in physics* **29**(1), 255 (1966)
- [16] Liu, W., Chen, Z.-H., Su, Y., Wang, Y., Dou, W.: Predicting rate kernels via dynamic mode decomposition. *The Journal of Chemical Physics* **159**(14) (2023)
- [17] Liu, W., Su, Y., Wang, Y., Dou, W.: Memory kernel coupling theory: Obtain time correlation function from higher-order moments. *arXiv preprint arXiv:2407.01923* (2024)
- [18] Itaka, T.: Solving the time-dependent schrödinger equation numerically. *Physical Review E* **49**(5), 4684 (1994)
- [19] Nys, J., Pescia, G., Sinibaldi, A., Carleo, G.: Ab-initio variational wave functions

- for the time-dependent many-electron schrödinger equation. *Nature communications* **15**(1), 9404 (2024)
- [20] Weiße, A., Wellein, G., Alvermann, A., Fehske, H.: The kernel polynomial method. *Reviews of modern physics* **78**(1), 275–306 (2006)
- [21] Vyazovskaya, M., Pupashenko, N.: On the normalizing multiplier of the generalized jackson kernel. *Mathematical Notes* **80** (2006)
- [22] Lykos, P., Pratt, G.: Discussion on the hartree-fock approximation. *Reviews of Modern Physics* **35**(3), 496 (1963)
- [23] Shavitt, I.: The method of configuration interaction. In: *Methods of Electronic Structure Theory*, pp. 189–275. Springer, New York (1977)
- [24] Sherrill, C.D., Schaefer III, H.F.: The configuration interaction method: Advances in highly correlated approaches **34**, 143–269 (1999)
- [25] Bartlett, R.J., Musiał, M.: Coupled-cluster theory in quantum chemistry. *Reviews of Modern Physics* **79**(1), 291–352 (2007)
- [26] White, S.R.: Density matrix formulation for quantum renormalization groups. *Physical review letters* **69**(19), 2863 (1992)
- [27] White, S.R.: Density-matrix algorithms for quantum renormalization groups. *Physical review b* **48**(14), 10345 (1993)
- [28] Brabec, J., Brandejs, J., Kowalski, K., Xantheas, S., Legeza, Ö., Veis, L.: Massively parallel quantum chemical density matrix renormalization group method. *Journal of Computational Chemistry* **42**(8), 534–544 (2021)
- [29] Larsson, H.R., Zhai, H., Umrigar, C.J., Chan, G.K.-L.: The chromium dimer: closing a chapter of quantum chemistry. *Journal of the American Chemical Society* **144**(35), 15932–15937 (2022)
- [30] Perez-Garcia, D., Verstraete, F., Wolf, M.M., Cirac, J.I.: Matrix product state representations. *arXiv preprint quant-ph/0608197* (2006)
- [31] Fischer, A., Igel, C.: An introduction to restricted boltzmann machines. In: *Iberoamerican Congress on Pattern Recognition*, pp. 14–36 (2012). Springer
- [32] Choo, K., Mezzacapo, A., Carleo, G.: Fermionic neural-network states for ab-initio electronic structure. *Nature communications* **11**(1), 2368 (2020)
- [33] Carleo, G., Troyer, M.: Solving the quantum many-body problem with artificial neural networks. *Science* **355**(6325), 602–606 (2017)

- [34] Barrett, T.D., Malyshev, A., Lvovsky, A.: Autoregressive neural-network wavefunctions for ab initio quantum chemistry. *Nature Machine Intelligence* **4**(4), 351–358 (2022)
- [35] Wu, Y., Guo, C., Fan, Y., Zhou, P., Shang, H.: Nnqs-transformer: an efficient and scalable neural network quantum states approach for ab initio quantum chemistry. In: *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 1–13 (2023)
- [36] Larochelle, H., Murray, I.: The neural autoregressive distribution estimator. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 29–37 (2011). *JMLR Workshop and Conference Proceedings*
- [37] Zhao, T., Stokes, J., Veerapaneni, S.: Scalable neural quantum states architecture for quantum chemistry. *Machine Learning: Science and Technology* **4**(2), 025034 (2023)
- [38] Pfau, D., Spencer, J.S., Matthews, A.G., Foulkes, W.M.C.: Ab initio solution of the many-electron schrödinger equation with deep neural networks. *Physical review research* **2**(3), 033429 (2020)
- [39] Hermann, J., Schätzle, Z., Noé, F.: Deep-neural-network solution of the electronic schrödinger equation. *Nature Chemistry* **12**(10), 891–897 (2020)
- [40] McArdle, S., Endo, S., Aspuru-Guzik, A., Benjamin, S.C., Yuan, X.: Quantum computational chemistry. *Reviews of Modern Physics* **92**(1), 015003 (2020)
- [41] Jordan, P., Wigner, E.P.: Über Das Paulische Äquivalenzverbot. Springer, Verlag Berlin Heidelberg (1993)
- [42] Seeley, J.T., Richard, M.J., Love, P.J.: The bravyi-kitaev transformation for quantum computation of electronic structure. *The Journal of chemical physics* **137**(22) (2012)
- [43] Bravyi, S.B., Kitaev, A.Y.: Fermionic quantum computation. *Annals of Physics* **298**(1), 210–226 (2002)
- [44] Lin, Y., Han, S., Mao, H., Wang, Y., Dally, W.J.: Deep gradient compression: Reducing the communication bandwidth for distributed training. arXiv preprint arXiv:1712.01887 (2017)
- [45] Kingma, D.P.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)