

Review article



Ab initio quantum chemistry with neural-network wavefunctions

Jan Hermann ^{1,2,11}, James Spencer ^{3,11}, Kenny Choo ^{4,5,11}, Antonio Mezzacapo ⁶, W. M. C. Foulkes ⁷, David Pfau ^{3,7} , Giuseppe Carleo ⁸ & Frank Noé ^{1,2,9,10}

Abstract

Deep learning methods outperform human capabilities in pattern recognition and data processing problems and now have an increasingly important role in scientific discovery. A key application of machine learning in molecular science is to learn potential energy surfaces or force fields from ab initio solutions of the electronic Schrödinger equation using data sets obtained with density functional theory, coupled cluster or other quantum chemistry (QC) methods. In this Review, we discuss a complementary approach using machine learning to aid the direct solution of QC problems from first principles. Specifically, we focus on quantum Monte Carlo methods that use neural-network ansatzes to solve the electronic Schrödinger equation, in first and second quantization, computing ground and excited states and generalizing over multiple nuclear configurations. Although still at their infancy, these methods can already generate virtually exact solutions of the electronic Schrödinger equation for small systems and rival advanced conventional QC methods for systems with up to a few dozen electrons.

Sections

Introduction

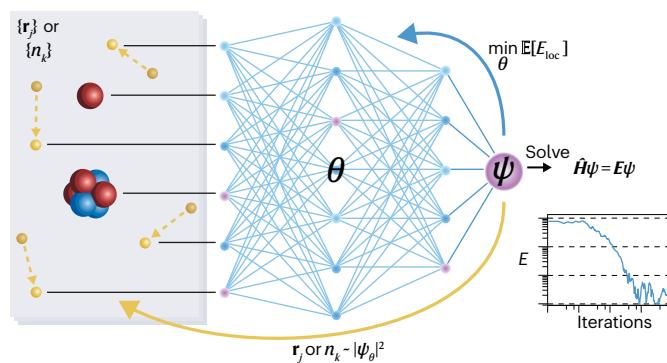
Electronic structure

Machine learning for electronic Schrödinger equation

Electrons in first quantization

Electrons in second quantization

Challenges and outlook



A full list of affiliations appears at the end of the paper. e-mail: pfau@google.com; giuseppe.carleo@epfl.ch; franknoe@microsoft.com

Introduction

Since the early 2010s, machine learning (ML) has made inroads into many areas of the physical sciences¹, often outperforming more conventional computational methods^{2,3} or offering entirely new approaches to solve scientific problems^{4,5}. Quantum chemistry (QC) has been among the first fields to have been affected by this revolution^{6–8}. Most applications of ML in QC have been concerned with supervised learning of molecular properties from molecular structure⁹, either across conformational¹⁰ or chemical space¹¹, as well as with unsupervised learning for the generation of novel molecules¹². These methods all require a pre-existing data set of molecules and their properties for training, typically obtained with standard methods of QC such as density functional theory (DFT)¹³ or coupled cluster¹⁴. In these scenarios, ML accurately approximates a given method of QC at vastly increased computational efficiency. This approach has already been reviewed in other works cited earlier. In this Review, we focus on the complementary use of ML as an ab initio technique in QC, which requires no external data, but recovers molecular properties from first principles. Here, ML is ‘integrated’ into QC, with the goal of arriving at ab initio methods with a more favourable accuracy–efficiency trade-off than conventional QC methods.

The goal of computational chemistry is to predict uncharacterized properties of molecules with a given structure and to design new molecules with desired properties. Most molecular properties are determined by the behaviour of the electrons, so QC methods attempt to approximate the Schrödinger equation for electrons in molecules. QC methods can be divided into ab initio and semi-empirical methods, in which the former have no fitted parameters determined from external data, whereas the latter do. Methods that do not use quantum mechanics at all (such as force fields) are called empirical and are typically not considered part of QC, although this view might be changing with the advent of principled and accurate ML-based empirical methods. Here we use ‘empirical’ to refer to the use of external data, be it experimental, calculated, microscopic or macroscopic. It is useful to cast these three categories of methods in the light of ML terminology (Fig. 1a).

ML can be roughly divided into supervised, unsupervised and reinforcement learning. In supervised learning, the ML model learns to predict the labels in the data given the corresponding features so as to minimize the difference between the predicted and reference labels. By identifying the features (inputs) with molecular structures and the labels (outputs) with molecular properties, all empirical – and to various degrees also semi-empirical – QC methods fit into supervised learning, mostly using relatively simple and physically motivated functional forms rather than the more general and highly flexible functions typical for ML. Conversely, the many successful supervised ML models that predict energies or other molecular properties on the basis of QC training data can be classified as empirical methods^{10,15–17}. Along this line, ML models of density functionals or approximate Hamiltonians that are embedded into some QC frameworks and trained end to end in a supervised fashion on QC data can be considered semi-empirical methods^{18–23}. Unsupervised learning is concerned with unlabelled data, and the general task is to learn the underlying probability distribution that would generate a given data set. Examples in chemistry include generative models for structural formulas²⁴ as well as full 3D structures of molecules^{4,25} and in physics include the estimation of quantum states from measurements, known as quantum tomography²⁶. Finally, in reinforcement learning, the ML model (also referred to as an ‘agent’) is able to interact directly with its environment, rather than to just passively receive data. Here, the aim is for the agent to learn a ‘policy’ for how to interact with the environment so as to maximize a long-term reward

such as predicting molecular properties²⁷. Reinforcement learning is behind some of the most prominent successes of ML such as playing games at a superhuman level^{28–30} or the control of plasma in tokamaks³¹. In certain settings, the agent can self-generate data by treating its own policy as the environment. This is known as ‘self-play’ and has been the basis for many advances in symmetric games such as chess^{32,33}. Although there are many key differences, this is the branch of ML – conceptually – most similar to ab initio QC, in the sense that no external data other than the rules of the system or game are required for either.

In our case, the rules of the system are encoded in the Schrödinger equation, which is an eigenvalue problem that can be equivalently formulated via various variational principles – its solutions, the eigenstate wavefunctions (a glossary of quantum and ML terms used throughout this Review has been included in Box 1) and energies can be found by searching for stationary points of certain functionals over the space of all physically admissible wavefunctions. Importantly, the ground state of a molecule can be found by minimizing the energy expectation value of a wavefunction. This principle underlies many ab initio QC methods, and also the methods in this Review, as such a variational principle naturally defines an ML problem – the eigenstates (such as the ground state) are represented as a neural network, and the parameters of that network are obtained by minimizing the variational electronic energy. The various reviewed methods then differ in the particular form of the neural-network ansatz used. A tutorial introduction into this topic has been published previously³⁴, and ref. 35 is a complementary review of neural-network approaches for solving the electronic Schrödinger equation.

Although one obvious contribution of the methods reviewed subsequently is extending the scope of application of ML into new areas, these methods must be ultimately evaluated within the context of QC with its existing advanced numerical methods and established criteria for what constitutes a practically useful method. A central concept in such an evaluation is the trade-off between the accuracy of a method and its computational efficiency, which in practice translates into the largest system size that can be feasibly modelled (Fig. 1b). Ab initio methods are in general costly – virtually exact results can be obtained for systems with at most one to two dozen electrons, and one order of magnitude larger systems can be routinely treated with highly accurate approximate methods, which can nevertheless fail even qualitatively for systems with complicated electronic structure (referred to as ‘strong correlation’ or ‘multireference character’). In this context, the aspiration of ab initio QC with neural networks is to restore the high accuracy of the approximate QC methods even for such complicated cases, without reducing the accessible system sizes too much. Although none of the methods reviewed here is yet mature enough to routinely model systems with more than a few dozen electrons with guaranteed accuracy, the results obtained so far and reviewed subsequently indicate that with further effort and progress, this aspiration might be achievable.

Electronic structure

The theoretical basis of QC is electronic-structure theory, and here we outline its fundamentals necessary for understanding the later sections that involve deep learning.

Schrödinger equation

QC aims at finding approximate solutions of the electronic Schrödinger equation that strike a good balance between accuracy and efficiency³⁶ (Fig. 1b). The non-relativistic electronic Schrödinger equation within

the Born–Oppenheimer approximation for a given molecule specified by Hamiltonian \hat{H} through the charges (Z_i) and coordinates (\mathbf{R}_i) of the nuclei is a second-order differential equation for the wavefunction, $\psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$, which is a function of the coordinates of N electrons (Fig. 2a):

$$\hat{H}\psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = E\psi(\mathbf{r}_1, \dots, \mathbf{r}_N), \quad (1)$$

$$\hat{H} := \sum_i \left(-\frac{1}{2} \nabla_{\mathbf{r}_i}^2 - \sum_j \frac{Z_j}{|\mathbf{r}_i - \mathbf{R}_j|} \right) + \sum_{i < j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}. \quad (2)$$

An alternative formulation of the Schrödinger equation uses the notion of an expectation value,

$$\langle \hat{H} \rangle_\psi = \frac{\int d\mathbf{r}_1 \cdots d\mathbf{r}_N \psi(\mathbf{r}_1, \dots, \mathbf{r}_N) \hat{H} \psi(\mathbf{r}_1, \dots, \mathbf{r}_N)}{\int d\mathbf{r}_1 \cdots d\mathbf{r}_N |\psi(\mathbf{r}_1, \dots, \mathbf{r}_N)|^2}. \quad (3)$$

Instead of solving equation (1), the ground-state (lowest-energy) solution can be found by minimizing this energy expectation value with respect to all possible wavefunctions (variational principle):

$$E = \min_{\psi} \langle \hat{H} \rangle_\psi. \quad (4)$$

Antisymmetric wavefunctions

Electrons are fermions, and as such their wavefunction must be antisymmetric with respect to exchange of any two electrons. This cardinal feature of electronic wavefunctions permeates the whole of QC. In general, electrons also possess spin coordinates, $s_i \in \{\uparrow, \downarrow\}$, but the non-relativistic Hamiltonian does not operate on spin, so the spin coordinate of each electron can be considered fixed. In this scenario (for full treatment, see Sec. IV.E of ref. 37), the spatial wavefunction must be antisymmetric only with respect to the exchange of same-spin electrons, for example, when $s_i = s_j$ (Fig. 2b):

$$\psi(\dots, \mathbf{r}_i, \dots, \mathbf{r}_j, \dots) = -\psi(\dots, \mathbf{r}_j, \dots, \mathbf{r}_i, \dots). \quad (5)$$

By far, the most common way to form antisymmetric wavefunctions in QC is as antisymmetrized products of single-electron functions (orbitals), $\phi_i(\mathbf{r})$. These products can be written as determinants, D , of an $N \times N$ matrix, $\phi_i(\mathbf{r}_i)$, formed by putting N electrons into N orbitals, and are referred to as Slater determinants (Fig. 2c):

$$D_\phi(\mathbf{r}_1, \dots, \mathbf{r}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\mathbf{r}_1) & \phi_1(\mathbf{r}_2) & \cdots & \phi_1(\mathbf{r}_N) \\ \phi_2(\mathbf{r}_1) & \phi_2(\mathbf{r}_2) & \cdots & \phi_2(\mathbf{r}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_N(\mathbf{r}_1) & \phi_N(\mathbf{r}_2) & \cdots & \phi_N(\mathbf{r}_N) \end{vmatrix}. \quad (6)$$

When interpreting $\phi_i(\mathbf{r}_i)$ as the i th component of an N -dimensional feature vector for the i th electron (using ML parlance), $\phi(\mathbf{r}_i)$, a Slater determinant is in fact the only antisymmetric function of N feature vectors that is linear in every one of the vectors, making it a natural choice. Alternative antisymmetric forms exist, such as the Pfaffian³⁸ or the Vandermonde determinant and its generalizations^{39,40}, but these are far less common and are not discussed.

Slater determinants formed from different orbitals can be further mixed in a linear combination without breaking the antisymmetry (Fig. 2c). This simple technique is the powerhouse behind all the

high-accuracy methods of QC, yet it is also its bane, because the number of Slater determinants required to achieve a given accuracy rises exponentially with the number of atoms in most cases. For fermionic

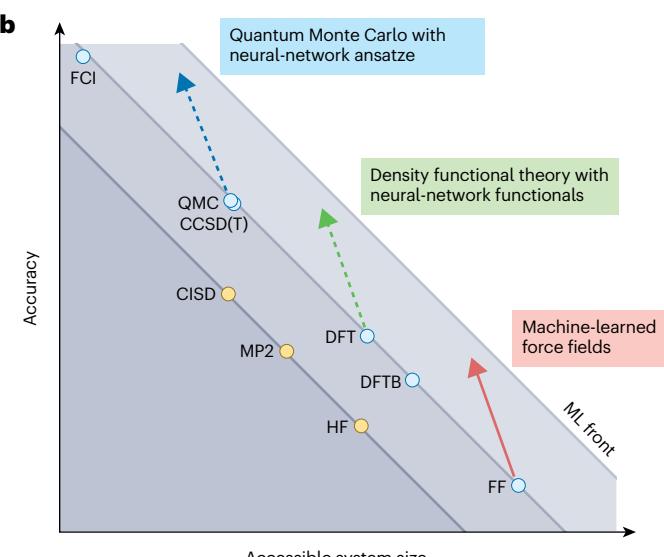
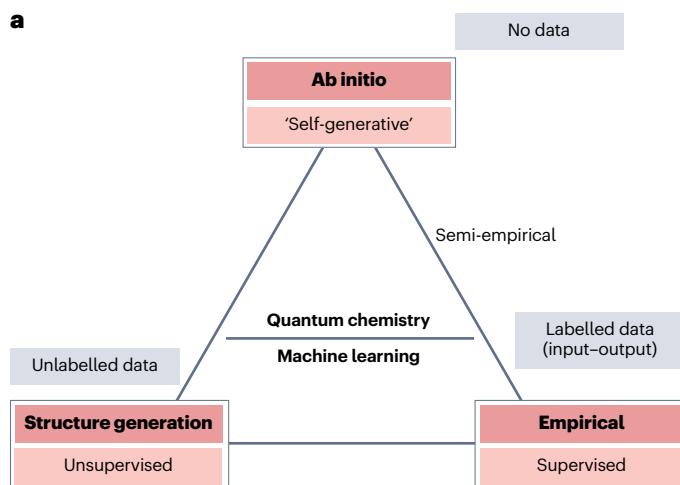


Fig. 1 | Quantum chemistry and machine learning. **a**, Machine learning (ML) disciplines and their dependence on data can be mapped to disciplines in quantum chemistry (QC). This work reviews the use of ML in ab initio QC, in which the only input to ML is the Schrödinger equation itself. This approach uses self-generated data, rather than relying on external data. The closest analogue in ML is reinforcement learning with self-play, which substitutes data from an external environment with data generated by the agent, although in many other respects the two approaches are distinct. **b**, Trade-off between the accuracy and computational efficiency, resulting in practically accessible molecular system size for QC methods. New technologies can lead to better trade-offs, such as pushing the front (diagonal lines) to better accuracy and/or system sizes. ML methods are being developed to achieve higher accuracy within existing QC frameworks. Whereas supervised ML force fields are already established (orange), this is work in progress for the ML of density functionals (green) and quantum Monte Carlo (QMC) (blue, this Review). CCSD(T), coupled cluster with singlets, doublets and perturbative triplets; CISD, configuration interaction with singlets and doublets; DFT, density functional theory; DFTB, density-functional-based tight binding; FCI, full configuration interaction; FF, force field; HF, Hartree–Fock; MP2, Møller–Plesset perturbation theory to second order.

Box 1

Glossary of quantum and machine learning terms

Quantum terms

Wavefunction, ψ

Eigenfunction of the Schrödinger equation. Describes the electronic structure of a quantum state, its square amplitude $|\psi|^2$ is the probability density to find N electrons anywhere in $3N$ -dimensional space.

Energy, E

Eigenvalue of the Schrödinger equation. The lowest-lying (ground-state) eigenvalue quantifies the stability of the nuclear configuration. When evaluated for different nuclear coordinates (\mathbf{R}), $E(\mathbf{R})$ is the potential energy surface used in geometry optimization and molecular dynamics.

Born–Oppenheimer approximation

Separating motion of electrons and nuclei, owing to their vastly different weight.

Spin

A property of quantum particles. Electrons are fermions with spin quantum numbers $\{+\frac{1}{2}, -\frac{1}{2}\}$. The wavefunction must be antisymmetric (switch sign) when exchanging electrons of the same spin.

Slater determinant

Popular ansatz to model electronic wavefunctions that are antisymmetric with respect to electron exchange.

Variational principle

Makes it possible to solve the Schrödinger equation by minimizing E over all antisymmetric ψ .

Basis set and orbitals

A set of fixed functions used to construct solutions of the Schrödinger equation. For molecules, atomic orbitals are single functions from the basis set suitable to locate an electron around a nucleus. Molecular

orbitals are weighted combinations of atomic orbitals, describing the location of an electron in a molecule.

Hartree–Fock

A foundational method for solving the Schrödinger equation with one Slater determinant composed of N single-electron orbitals.

Variational Monte Carlo

Numerical approximation method to solving the Schrödinger equation. Uses the variational principle and can work with flexible ansatz functions for ψ .

Machine learning terms

Deep learning

Machine learning involving multiple layers of nonlinear functions.

Backpropagation

Principle used for training deep neural networks by implementing the chain rule of differentiation.

Graph neural network

Artificial neural network in which nodes are connected by edges. Nodes have states that are updated by computations involving the connected edges and nodes.

Invariance and equivariance

Functions whose output does not change (invariance) or changes in the same way (equivariance) in response to group transformation of the inputs, such as rotations, translations and permutations.

Markov chain Monte Carlo

Numerical method to sample a probability distribution by generating a sequence of random variables (RVs) in which each RV is generated by perturbing the previous RV.

wavefunctions, there is no known general approach to effectively reduce the search space from this exponential regime without sacrificing accuracy. However, QC has produced many methods that achieve excellent approximations for specific molecules and materials of practical interest. The cost of these highly accurate methods is generally less than exponential, but nevertheless increases rapidly with system size (Fig. 1b).

Variational wavefunction methods

An important class of QC methods derives directly from the variational principle (equation (4)), by assuming a certain wavefunction ‘ansatz’, $\psi(\cdot; \boldsymbol{\theta})$, parametrized by $\boldsymbol{\theta}$. Minimizing the energy of this ansatz with respect to $\boldsymbol{\theta}$ then always yields an upper bound for the exact ground-state energy,

$$E = \min_{\psi} \langle \hat{H} \rangle_{\psi} \leq \min_{\boldsymbol{\theta}} \langle \hat{H} \rangle_{\psi(\cdot; \boldsymbol{\theta})}. \quad (7)$$

The bound becomes tighter as the expressiveness of the ansatz is improved.

One can discern two strategies to construct the ansatzes. First, traditional QC uses relatively simple forms, such that the integral of equation (3) can be evaluated analytically, which drastically simplifies the minimization problem^{36,41}. Second, quantum Monte Carlo (QMC) enables the use of ansatzes with arbitrary analytical forms at the cost of having to do the integral evaluation and minimization stochastically⁴². The latter is a natural framework to incorporate neural networks, and we introduce it in more detail in the next section. Here we introduce three ansatzes for electronic wavefunctions of the first (traditional) kind, as they serve as scaffolding for the neural-network ansatzes introduced in later sections (Fig. 2d–f).

Hartree–Fock. Perhaps, the simplest non-trivial ansatz in QC is the single Slater determinant of equation (6), in which the orbitals $\phi_i(\mathbf{r})$

are varied. Optimized variationally, this ansatz leads to the so-called Hartree–Fock (HF) method. In practice, the orbitals are linearly expanded in a fixed finite one-electron basis, $\varphi_k(\mathbf{r})$, $k = 1, \dots, M$, with $M \sim N$ in most cases:

$$E_{\text{HF}} = \min_{\phi_j} \langle \hat{H} \rangle_{\det \phi_j(\mathbf{r}_j)} \approx \min_{C_{kj}} \langle \hat{H} \rangle_{\det \sum_k C_{kj} \varphi_k(\mathbf{r}_j)}. \quad (8)$$

The use of a finite basis set turns the functional optimization problem of equation (8) into a computational problem whose cost scales with the fourth power of the number of basis functions, $O(K^4)$, assuming a naive implementation. On its own, the HF ansatz is expressive enough to describe much of chemistry qualitatively, but not always, and certainly not quantitatively. However, it can be considered a starting point for most wavefunction-based QC methods.

Dynamical DFT is quite a different approach and relies instead on an in-principle exact mapping of the ab initio Hamiltonian (equation (2)) to a mean-field-like problem. For a given energy functional, this problem can be solved exactly with a single Slater determinant^{13,43}. However, the variational principle does not hold in DFT because the exchange-correlation contributions to the energy functional are not known exactly and must be approximated in practice. From here on, we will stay within the variational principle and instead focus on increasing the expressiveness of the HF ansatz.

Configuration interaction. The HF ansatz can be straightforwardly extended by forming multiple Slater determinants from different sets of orbitals and considering their linear combination (Fig. 2c),

$$\psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = \sum_p c_p D_{\phi_p}(\mathbf{r}_1, \dots, \mathbf{r}_N). \quad (9)$$

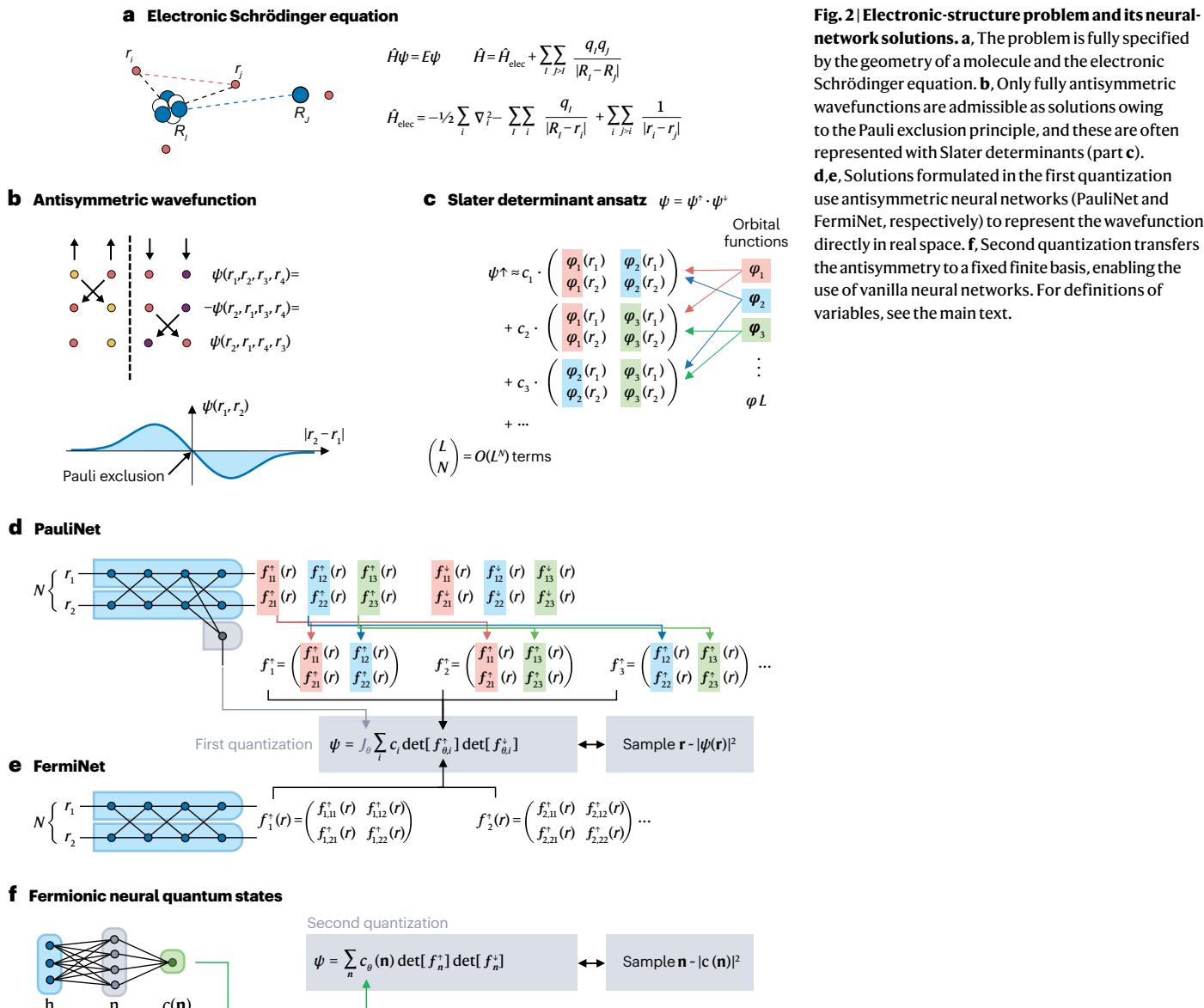


Fig. 2 | Electronic-structure problem and its neural-network solutions. **a**, The problem is fully specified by the geometry of a molecule and the electronic Schrödinger equation. **b**, Only fully antisymmetric wavefunctions are admissible as solutions owing to the Pauli exclusion principle, and these are often represented with Slater determinants (part **c**). **d,e**, Solutions formulated in the first quantization use antisymmetric neural networks (PauliNet and FermiNet, respectively) to represent the wavefunction directly in real space. **f**, Second quantization transfers the antisymmetry to a fixed finite basis, enabling the use of vanilla neural networks. For definitions of variables, see the main text.

Box 2

First and second quantization

Computational methods for the electronic Schrödinger equation can be divided to first-quantized approaches in real space and second-quantized approaches in a discrete basis (see 1st ed., Vol. 2, Appx. U of ref. 36). In first quantization, one works with the individual electrons and their coordinates directly in real space ($\mathbf{r}_i \in \mathbb{R}^3$, $i=1, \dots, N$) as in equation (1). Here, the wavefunction $\psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$ must be an antisymmetric function that specifies which electrons occupy which coordinates.

In second quantization, one has first to introduce a discrete one-electron basis (in practice finite), labelled by k , which then enables one to work with preformed antisymmetric many-electron basis states (Slater determinants). Within this formalism, rather than specifying which electrons occupy which one-electron states, the occupation numbers ($n_k \in \{0, 1\}$, $\sum_k n_k = N$) specify which one-electron states are occupied without any reference to a particular electron. Here, the wavefunction $\psi_{n_1, n_2, \dots} \equiv \psi(n_1, n_2, \dots)$ can be an arbitrary tensor (function of the discrete indices n_k) without any prescribed (anti)symmetry. The antisymmetry is instead fully encoded in the Slater basis states.

This ability to push the antisymmetry from the wavefunction object to the many-electron basis is the main advantage of second quantization, at the cost of having to commit to a particular discrete basis. But, regardless of the computational framework, either the wavefunction object itself (in first quantization) or the many-electron basis (in second quantization) consists of Slater determinants. In high-accuracy methods, their number grows rapidly with system size.



First and second quantization. Illustration on $N=3$ electrons in one dimension and in a finite basis of size 5. $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$.

When the orbitals of each determinant are pooled from a larger superset of (mutually orthogonal) fixed orbitals of size $M > N$, and the only free parameters are the linear coefficients of the determinants, the ansatz is called configuration interaction (CI). One of the appeals of the CI ansatz is that its Slater determinants can be considered a many-electron antisymmetric basis and labelled using the occupation numbers of the one-electron states. This so-called second quantized formalism has many convenient properties for computation (Box 2). The simplest version of CI, called full CI (FCI), considers all $(M \choose N)$ possible Slater determinants and is exact within the chosen finite one-electron basis. In the usual case when $M \sim N$, however, the computational effort scales exponentially with N , which makes FCI applicable only to the smallest molecules. Ways to tackle the exponential scaling include fixed truncation of the CI expansion or its ‘compression’ through analytical means (coupled cluster theory⁴⁴ and matrix product states⁴⁴), deterministic pruning (selected CI⁴⁵) or stochastic sampling (FCI-QMC⁴⁶). Neural-network states described later explore a novel way of ‘compressing’ this CI expansion.

Beyond fixed bases. The effectiveness of the CI ansatz depends on the choice of the fixed molecular orbitals $\phi_j(\mathbf{r})$ from which the Slater determinants $D_{\phi_p}(\mathbf{r}_1, \dots, \mathbf{r}_N)$ are built. A natural extension of CI allows both the orbitals and the CI expansion coefficients c_p to vary during the variational minimization. Such an ansatz of two stacked linear combinations (equations (8) and (9)) is harder to optimize but much more expressive. The most common variant is to consider all $(M' \choose N')$ Slater determinants formed by letting $N' < N$ electrons occupy a space of $M' < M$ orbitals, whereas the remaining $N - N'$ electrons occupy a fixed set of inactive orbitals. This is called the complete active space self-consistent field (CASSCF) method⁴⁷. Owing to the larger variational freedom, a CASSCF ansatz typically requires many fewer determinants than a CI ansatz of comparable accuracy.

But CASSCF and even FCI are still limited by the fixed one-electron basis used to form the molecular orbitals (equation (8)): FCI is only exact in the complete basis set limit, which in practice cannot be reached for any but the smallest molecular systems. An extension of the CAS-SCF ansatz would allow not only the one-electron orbitals but also the one-electron basis functions to vary. The stacked structure of such an ansatz would be reminiscent of deep neural networks, and the section on first quantization explores the culmination of this line of thought by incorporating actual deep neural networks into the ansatz. This removes any a priori limitations on the expressiveness. By making each individual determinant maximally expressive, such ansatzes further reduce the number of determinants required to reach a given accuracy.

Machine learning for electronic Schrödinger equation

Electronic-structure theory and deep learning are seemingly very different disciplines, and it might not be a priori clear how they can be connected. This section aims to build that bridge.

Mapping quantum mechanics to machine learning

An ML problem and its solution are specified by the model, its inputs and outputs, the data and the optimization criterion (loss function). In this regard, solving the Schrödinger equation with the variational principle amounts to the following ML problem (Fig. 3). The neural network represents a wavefunction, which accepts electron coordinates (first quantization) or occupation numbers (second quantization) as input and outputs the wavefunction value. The loss function is the energy expectation value corresponding to this wavefunction. The inputs are sampled from the probability distribution given by the square of the wavefunction represented by the current neural network. The Hamiltonian is used to estimate the loss function from the samples. The parameters of the network, and thus the wavefunction, are then modified to minimize the loss function. Except for the representation of the wavefunction as a network, this is the regular variational Monte Carlo (VMC) framework (Box 3). The optimization methods used (Box 4) are also fairly conventional, although adapted to a neural-network context.

The applicability of deep learning for representing many-body wavefunctions was first realized and exploited by Carleo and Troyer⁴⁸ for the case of spin lattices in one and two dimensions. Their approach, known as neural quantum states (NQS), has since been applied to many different quantum systems^{49–52}. In essence, this Review is concerned with the extension of this approach to electrons in molecules.

Deep learning

The standard practice in ab initio QC today is in some ways analogous to the state of computer vision – a field of research in computer

science – before the rise of deep learning. Before 2012, the best pipelines for large-scale image recognition consisted of a combination of hand-designed features and simple ML models⁵³. A single deep convolutional neural network trained end to end was able to cut the recognition error in half relative to these systems⁵⁴, and since then deep neural networks have dominated computer vision research.

In ab initio QC, ground-state solutions to the Schrödinger equation are usually represented by a wavefunction ansatz with a relatively simple functional form, and parameters are usually fit through a mix of procedures (fixed-point iteration, variational optimization and others) rather than a unified end-to-end estimation of all parameters simultaneously. The development of deep QMC methods is driven by the hope that the use of neural networks will greatly increase the expressiveness of wavefunction ansatzes, enabling large leaps in accuracy as in image recognition. To appreciate how and why deep neural networks can be usefully applied in QC, a brief review of their application in artificial intelligence is necessary. For a thorough review of the history of deep learning, see Schmidhuber⁵⁵, and for a review of the fundamental concepts in deep learning, see LeCun et al.⁵⁶.

Neural networks date back to the very beginning of computer science⁵⁷, and their modern form originates with the single perceptron ‘unit’⁵⁸, which produces a nonlinear function of the sum of a constant as output – known as the bias – and a linear combination of its inputs. The nonlinear function rises from zero to one as its input increases, mimicking the activation function of a biological neuron. When many such units are assembled in parallel to form a ‘layer’, and several layers are computed serially, taking the output from one layer as the input for the next, the resulting multilayer perceptron (MLP) can, in theory, represent any smooth function to arbitrary accuracy given enough units⁵⁹. However, actually fitting or learning a set of parameters that matches any given function is a different matter. A form of gradient descent utilizing derivatives computed using backpropagation, or reverse-mode automatic differentiation^{60–62}, was found to be effective for training neural networks⁶³. This led to a wave of enthusiasm for neural networks, which eventually faded as several issues were discovered, such as the infamous ‘vanishing gradients’ and getting stuck in local minima.

Different factors were instrumental in rehabilitating neural networks under the banner of ‘deep learning’: a combination of algorithmic advances⁶⁴ and the use of modern graphics processing units⁶⁵ made computations faster, and the resulting ability to train larger networks made issues with local minima less severe^{66,67}. Furthermore, with the help of stochastic gradient descent, deep neural networks can be applied straightforwardly and efficiently to large data sets, unlike other ML models^{68,69}. Finally, empirical successes such as winning the ImageNet Large Scale Visual Recognition Challenge⁷⁰ helped legitimize deep learning research and generate excitement among researchers.

Today, the barrier to entry for developing and training deep neural networks is quite low, owing to a mature ecosystem of software libraries for numerical computing with automatic differentiation and hardware accelerators^{71–73}. However, achieving good performance from a deep learning model still requires finesse and the application of various heuristics. Deep neural networks also sometimes suffer from odd failure modes, such as imperceptible perturbations to the inputs causing enormous changes to the outputs⁷⁴. However, linear models also have this issue, and these issues are of limited concern in the applications discussed here, where only average performance matters.

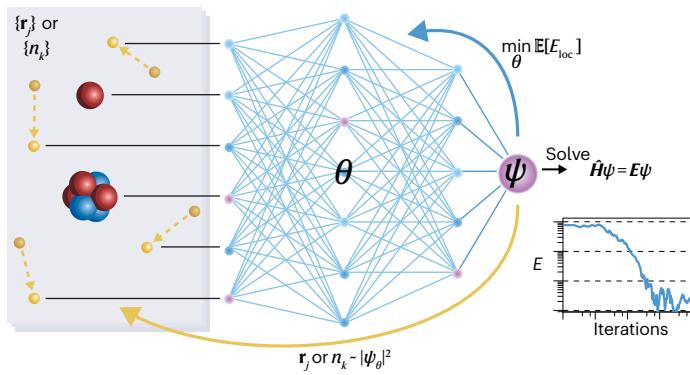


Fig. 3 | Variational Monte Carlo with neural networks. Electron positions, r_j , or orbital occupation numbers, n_k , describe an electron configuration (Box 2), which is an input to the wavefunction, ψ , represented by a neural network parameterized with θ . The wavefunction is used in two ways: first, to sample new electron configurations that provide new input to the neural network (yellow; Box 3), and second, to evaluate the electronic energy, which is minimized by varying the network parameters (blue; Box 4). Adapted with permission from ref. 116, Springer Nature.

Neural-network architectures

The starting point for most neural networks is the MLP, formed as a composition of L layers,

$$\text{MLP}(\mathbf{x}) = f^L \circ f^{L-1} \circ \dots \circ f^1(\mathbf{x}), \\ f^\ell(\mathbf{z}) = f(\mathbf{W}^\ell \mathbf{z} + \mathbf{b}^\ell), \quad (10)$$

where \circ denotes function composition, f is some nonlinear activation function and \mathbf{W}^ℓ and \mathbf{b}^ℓ are the matrices of weights and vectors of biases to learn. Although a basic MLP is capable of representing arbitrary functions, the real power of neural networks comes from more sophisticated architectures. Many of these architectures are designed to encode some particular ‘invariance’ or ‘equivariance’ – that is, when the input to the network is transformed in a particular way, the output should either be unchanged or transform in a corresponding way. For instance, the weights in a layer of a convolutional neural network – known as ConvNet⁷⁵ – are restricted to be a discrete convolution operator, which constrains each layer to be translation-equivariant and markedly reduces the number of possible weights in a layer. This is a natural constraint in, for example, image recognition.

Equivariance to permutation is another frequently useful property that is especially important in real-space approaches to representing electronic wavefunctions. A simple permutation-equivariant layer first proposed by Shawe-Taylor⁷⁶ can be constructed by applying the same transformation to each input and summing the results. More sophisticated permutation-equivariant layers are used by models such as the Transformer⁷⁷ or SchNet⁷⁸. Many of these equivariant layers can be unified in a conceptual framework based around the language of geometry and group theory, wherein the choice of transformation with respect to which to be equivariant naturally leads to recipes for constructing the appropriate neural-network layers⁷⁹.

Another class of neural-network architectures, which have been influential as wavefunction ansatzes, are restricted Boltzmann machines (RBMs)⁸⁰. These were originally developed for unsupervised learning, but in the VMC setting considered here they lead to a simple

Box 3

Variational Monte Carlo

Optimization of wavefunctions with neural networks naturally leads to the variational Monte Carlo (VMC) framework. First, Monte Carlo integration of equation (3) can handle ansatzes with arbitrary analytical forms for which analytical integrals are not available. Second, VMC samples these integrals stochastically that naturally combines with the stochastic gradient descent used for optimizing neural networks. In conventional QC, VMC has been used extensively with real-space first-quantized approaches³⁷ and also in the discrete-basis second-quantized setting^{169,170}.

The expectation value of any operator, such as the Hamiltonian (equation (3)), can be written as a Monte Carlo integral over a continuous or discrete basis:

$$\langle \hat{H} \rangle_{\psi} = \mathbb{E}_{\mathbf{x} \sim |\psi(\mathbf{x})|^2} [E_{\text{loc}}(\mathbf{x}; \psi)]. \quad (24)$$

Here, the expectation value is obtained as an expected value of local energy, $E_{\text{loc}}(\mathbf{x}) = \int_{\mathbf{x}'} H_{\mathbf{xx}'} \psi(\mathbf{x}')/\psi(\mathbf{x})$, defined for every basis element \mathbf{x} , over the square of the wavefunction. The local energy is calculated from the matrix elements $H_{\mathbf{xx}'}$ of the Hamiltonian in a given basis.

A straightforward and generally applicable way to obtain the samples is through Markov-chain Monte Carlo (MCMC). MCMC is an iterative procedure, in which a new sample point, \mathbf{x}' , is produced from a current one, \mathbf{x} , by making a proposal step with probability $g(\mathbf{x}'|\mathbf{x})$ and then accepting or rejecting the proposal with a certain probability (see Sec. 2.2 of ref. 171). The resulting Markov chain then samples $|\psi(\mathbf{x})|^2$. Variants of MCMC differ in the construction of the proposal steps and g and include the simplest Metropolis algorithm as well as more sophisticated flavours such as Langevin Monte Carlo (see Sec. 1.4 of ref. 172).

The VMC formula for the expectation value is exact in the limit of infinite sample size, $N \rightarrow \infty$. Still, in practice, it incurs a statistical error proportional to $\sqrt{\text{Var}[E_{\text{loc}}]/N}$. Although $1/\sqrt{N}$ converges slowly with increasing sample size, VMC has the great benefit that as the ansatz converges to the exact eigenstates, the local energy converges to a constant (the exact energy), and as such its variance vanishes and so does the statistical sampling error.

deterministic expression for the log probability closely resembling a one-layer MLP. Despite their early popularity, RBMs have been largely eclipsed in the AI community by other methods for unsupervised learning, such as variational autoencoders⁸¹, generative adversarial networks⁸², normalizing flows⁸³, autoregressive models^{84,85} and diffusion models⁸⁶. Some of these newer models have started to have an impact as neural-network wavefunction ansatzes for spin systems. Examples are deep autoregressive quantum states⁸⁷, convolutional neural networks⁸⁸, recurrent neural networks⁸⁹ and normalizing flows⁹⁰.

Neural-network architectures for molecules

Since the mid-2000s, designing ML architectures for molecular physics and chemistry tasks that are expressive and flexible, yet incorporate physical concepts and constraints, has been a very active field.

Invariance, for example, of energy functions with respect to translation and rotation of the molecular coordinates, and with respect to exchange of identical particles, has already been designed and incorporated into early architectures^{91–93}. Message-passing architectures such as SchNet have leveraged the increased expressiveness of deep learning by defining graph convolutions parametrized by rototranslationally invariant features^{78,94}. Many efforts went into the design of rotation-equivariant neural networks, which have vectorial or tensorial features rotating with the inputs^{95–100}. It has been shown that such neural-network architectures can reach high accuracy not only in the large-data limit but also for small data sets which could previously only be tackled by kernel approaches^{99,100}. Energy functions based on many-body expansions were considered in the form of using permutationally invariant polynomials¹⁰¹, atomic cluster expansion¹⁰² and permutationally invariant polynomials for molecules¹⁰³.

Although these neural-network architectures have been predominantly developed for supervised learning of potential energy surfaces from QC data, they have provided tools and concepts that are applicable to QMC with neural-network wavefunctions.

Electrons in first quantization

One approach to studying the electronic problem with deep learning is to work with parameterized many-body wavefunctions in first quantization, $\psi(\tau; \theta)$. Here τ stands for the N -tuple of electron coordinates, $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$, and sampling is realized over electronic positions τ (Box 3). The antisymmetry constraint (equation (5)) must be imposed in ψ to avoid collapsing onto a lower-energy bosonic state. A commonly adopted form is $\psi(\tau; \theta) = S(\tau; \theta) \times A(\tau; \theta)$, in which the first factor is symmetric (or ‘bosonic’) under exchange of electron coordinates and the second factor carries the necessary antisymmetry. The simplest and most common approach is to build the antisymmetric part of the wavefunctions using Slater determinants. As discussed earlier, single Slater determinants with fixed orbitals have limited expressiveness, and many such determinants need to be combined to achieve high accuracy¹⁰⁴. A natural generalization of a sum of fixed-orbital Slater determinants is the commonly used Slater–Jastrow wavefunction:

$$\psi(\tau; \theta) = e^{-J(\{\tau\}; \theta)} \sum_k c_k \begin{vmatrix} \phi_1^k(\mathbf{r}_1; \theta) & \cdots & \phi_1^k(\mathbf{r}_N; \theta) \\ \vdots & \ddots & \vdots \\ \phi_N^k(\mathbf{r}_1; \theta) & \cdots & \phi_N^k(\mathbf{r}_N; \theta) \end{vmatrix}, \quad (11)$$

in which the Jastrow factor, $J(\{\tau\}; \theta)$, constitutes the symmetric (bosonic) part of the state and typically contains one-body and two-body (and in many cases higher order) parameterized correlations. The set notation, $\{\tau\} \equiv \{\mathbf{r}_1, \dots, \mathbf{r}_N\}$, indicates that τ does not depend on the order of the electron coordinates. The determinants in equation (11) are typically replaced with the product of spin-up and spin-down determinants³⁷. Separating the up-spin and down-spin determinants improves computational efficiency, simplifies the implementation and makes it easier to handle the electron–electron cusps while leaving expectation values of spin-independent operators unchanged.

More flexible parametric forms can be obtained by leveraging the approximation power of artificial neural networks. In the following, we discuss neural-network-based strategies to parameterize these forms.

Discrete space

The first applications of neural networks to electronic systems were for electrons moving in discretized space, as realized, for example,

in the 2D Hubbard model of strongly interacting electrons. In the following, for simplicity, we discuss the case of N spinless electrons in M lattice sites and denote with $l(\mathbf{r}) \in [1, M]$, the discrete lattice index corresponding to electron position \mathbf{r} . The extension to the spinful case will be considered more in detail when discussing continuous space later on. The symmetric part $S(\mathbf{r}; \boldsymbol{\theta})$ can be readily parameterized with a strategy closely related to NQS for spins:

$$S(\mathbf{r}; \boldsymbol{\theta}) = g(n(\mathbf{r}); \boldsymbol{\theta}), \quad (12)$$

where $n(\mathbf{r})$ is the unique occupation-number representation corresponding to the electronic positions \mathbf{r} , and g represents a generic function that could be represented by a neural network. As the occupation numbers $n(\mathbf{r})$ are invariant under permutation of the electron positions, $g(n(\mathbf{r}))$ is also symmetric under exchange. Any of the neural-network architectures also adopted for spin systems⁴⁸ or lattice bosons⁴⁹ can be used to represent the symmetric part (S). Early works on the Hubbard model adopted positive-definite RBM-based parameterizations of $S(\mathbf{r}; \boldsymbol{\theta})$ ⁵⁰, whereas works now adopt deep network parameterizations allowing for sign changes¹⁰⁵.

The simplest parameterization for the antisymmetric part, $A(\mathbf{r}; \boldsymbol{\theta})$, is again a Slater determinant:

$$A(\mathbf{r}; \boldsymbol{\theta}) = \begin{vmatrix} \phi_1(\mathbf{r}_1; \boldsymbol{\theta}) & \cdots & \phi_1(\mathbf{r}_N; \boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \phi_N(\mathbf{r}_1; \boldsymbol{\theta}) & \cdots & \phi_N(\mathbf{r}_N; \boldsymbol{\theta}) \end{vmatrix}, \quad (13)$$

in which the matrix $\Phi \in \mathbb{C}^{N \times M}$ of discrete orbitals $\phi_i(\mathbf{r}_j) = \Phi_{i,l(\mathbf{r}_j)}$ holds the variational parameters to be optimized. This approach, however, has the important drawback of not providing enough variational flexibility as it effectively fixes the antisymmetric part to a mean-field reference solution.

Neural backflow. A considerable improvement is obtained by considering a many-body backflow transformation of the orbitals^{106,107}. In this variational form, the matrix of one-electron orbitals Φ is promoted to a parameterized many-electron function depending on all the occupation numbers:

$$\tilde{\Phi}_j(\boldsymbol{\theta}) = \Phi_j(\boldsymbol{\theta}) + \Delta_j(n(\mathbf{r}); \boldsymbol{\theta}), \quad (14)$$

where Δ is a correction to the single-electron orbitals Φ . In physics-inspired parameterizations, Δ is typically taken to be a simple function of the electronic occupation numbers¹⁰⁸. The neural backflow method¹⁰⁹ instead introduced a flexible parameterization of the backflow orbitals on the basis of artificial neural networks. In this case, Δ is parameterized with an MLP taking as inputs the electronic occupation numbers and outputting a many-body correction to the matrix Φ . This approach allows the orbitals to change dynamically depending on the positions of the electrons, thus allowing one to include genuinely many-body correlations in the antisymmetric part of the wavefunction.

Constrained hidden fermions. Neural backflow transformations are not the only way to introduce flexible parameterizations of the antisymmetric part of the wavefunction. The constrained hidden fermion formalism builds on the idea of introducing a set of \tilde{N} auxiliary fermionic particles, with positions \mathbf{q} , distributed over \tilde{M} lattice sites. These auxiliary particles are used to effectively mediate correlations among

the physical degrees of freedom¹¹⁰. Calling $\tilde{A}(\mathbf{r}, \mathbf{q}; \boldsymbol{\theta})$ a Slater determinant for the extended (physical + hidden) system, the resulting antisymmetric form for the physical system is given by

$$A(\mathbf{r}; \boldsymbol{\theta}) = \tilde{A}(\mathbf{r}, F(\mathbf{r}; \boldsymbol{\theta}); \boldsymbol{\theta}). \quad (15)$$

In this expression, F is a function parameterized by a neural network, mapping the physical positions to the hidden ones. This approach has been shown to improve systematically over the neural backflow form for the 2D Hubbard model¹¹⁰.

Continuous space

We now focus on describing the important case of first-quantized electrons in continuous space, directly corresponding to the electronic Schrödinger equation. As in the discrete-space case, the Slater–Jastrow form can be improved in a manner suitable for use with NQS by adding a backflow transformation, in which the one-electron orbitals $\phi_i(\mathbf{r}_j; \boldsymbol{\theta})$ are replaced by many-electron functions $\tilde{\phi}_i(\mathbf{r}_j, \{\mathbf{r}\}; \boldsymbol{\theta})$. The backflow transformation can either modify the orbitals directly via a multiplicative and/or additive term:

$$\tilde{\phi}_i(\mathbf{r}_j, \{\mathbf{r}\}; \boldsymbol{\theta}) = \phi_i(\mathbf{r}_j) f_i^{\otimes}(\mathbf{r}_j, \{\mathbf{r}\}; \boldsymbol{\theta}) + f_i^{\oplus}(\mathbf{r}_j, \{\mathbf{r}\}; \boldsymbol{\theta}), \quad (16)$$

Box 4

Optimizing neural-network ansatzes

Up to the statistical error, the VMC expectation value for the energy (Box 3) obeys the variational principle (equation (4)). VMC uses this by varying a parametric wavefunction ansatz ($\psi_{\boldsymbol{\theta}}$) so as to minimize the energy. For a sufficiently expressive ansatz, the variational energy will eventually approximate the ground-state energy of equation (1) and the ansatz will approximate the ground-state wavefunction, Ψ .

The most straightforward optimization method is gradient descent, in which the parameters are iteratively updated in the direction of the negative gradient of the loss function with respect to the parameters. This gradient can be expressed as certain expectation values and can be efficiently estimated using Monte Carlo integration (Box 3).

In some cases, the optimization can be sped up and made more stable with higher-order methods, such as the stochastic reconfiguration scheme, which takes the correlation between individual variational parameters into account¹⁵⁰. Stochastic reconfiguration approximates an imaginary-time evolution in which each iteration tries to approximate best the state $e^{-\eta H} |\psi\rangle$, where η is a step size. It is similar to the natural gradient descent algorithm¹⁷³ that is well known in the ML community and where the correlations between parameters are encoded in the Fisher information matrix¹⁷⁴. Kronecker-factored approximate curvature approach is an approximate version of the natural gradient descent that has been designed to be efficient specifically for neural networks¹²¹.

or act as a quasiparticle transformation of the electron coordinates:

$$\tilde{\phi}_i(\mathbf{r}_j; \{\mathbf{r}\}; \boldsymbol{\theta}) = \phi_i(\mathbf{r}_j + \boldsymbol{\xi}(\{\mathbf{r}\}; \boldsymbol{\theta})), \quad (17)$$

in which the parameterized functions, $f_i^\otimes, f_i^\oplus, \boldsymbol{\xi}$, are invariant to permutations of $\{\mathbf{r}\}$, and $\boldsymbol{\xi}(\{\mathbf{r}\}; \boldsymbol{\theta})$ is a three-component vector that modifies \mathbf{r}_j . If we consider a determinant of orbitals of this form,

$$\begin{vmatrix} \phi_1(\mathbf{r}_1; \{\mathbf{r}\}) & \dots & \phi_1(\mathbf{r}_N; \{\mathbf{r}\}) \\ \vdots & \ddots & \vdots \\ \phi_N(\mathbf{r}_1; \{\mathbf{r}\}) & \dots & \phi_N(\mathbf{r}_N; \{\mathbf{r}\}) \end{vmatrix}, \quad (18)$$

then we see that orbitals with backflow transformations are just one example of a broader class of functions. For the determinant to be antisymmetric, the matrix with elements $\Phi_{ij} = \phi_i(\mathbf{r}_j; \{\mathbf{r}\})$ must be permutation-equivariant; that is, exchanging electrons k and l also exchanges columns k and l . Although conventional Slater–Jastrow–backflow (SJB) wavefunctions have had considerable success, they also have limitations owing to the choice of fixed functional forms. Therefore, the goal is to develop more flexible permutation-equivariant functions. Here we highlight several approaches that share this common theme. We also note that similar model architectures can be used for other fermionic systems, such as those encountered in nuclear physics. Real-space VMC with neural-network ansatzes has been successfully used to model those as well^{111,112}.

Iterative backflow. Taddei et al.¹¹³ introduced a form of backflow that applied equation (17) repeatedly in an iterative fashion. Such an ansatz is formally equivalent to expressing the backflow as a deep neural network¹¹⁴, albeit with artificial restriction on the dimensionality of the hidden layers. The iterative backflow was used for studying the ^3He and ^4He liquids.

DeepWF. The Deep WaveFunction (DeepWF)³⁹ approach uses an ansatz similar to a Slater–Jastrow wavefunction but with a simpler antisymmetric term:

$$\psi(\mathbf{r}) = S(\{\mathbf{r}\}, R) A^\dagger(\mathbf{r}^\uparrow) A^\downarrow(\mathbf{r}^\downarrow). \quad (19)$$

The learned symmetric function S is similar to a Jastrow factor and ensures that the wavefunction captures the electron–nuclear and electron–electron cusp conditions. The antisymmetric factors, A^σ , are constructed from the Vandermonde-like determinant of an explicitly antisymmetric two-body function, $A^\sigma = \prod_{1 \leq i \leq j \leq N} (a(\mathbf{r}_i, \mathbf{r}_j, r_{ij}) - a(\mathbf{r}_j, \mathbf{r}_i, r_{ij}))$. The two-body antisymmetric function is entirely learned. Such a functional form can be evaluated more efficiently than a determinant. However, using a simplified antisymmetric function is also likely to limit the accuracy achieved: DeepWF obtains only 43.6% of the correlation energy for the beryllium atom and does not even reach HF accuracy for the boron atom. The PauliNet and FermiNet approaches described subsequently do much better: PauliNet obtained 99.94% and 97.3% of the correlation energies for the beryllium and boron atoms, and FermiNet obtained 99.97% and 99.83%, respectively. Furthermore, FermiNet and PauliNet substantially surpass conventional SJB wavefunctions on first-row atoms, for which nearly exact benchmark values exist¹¹⁵.

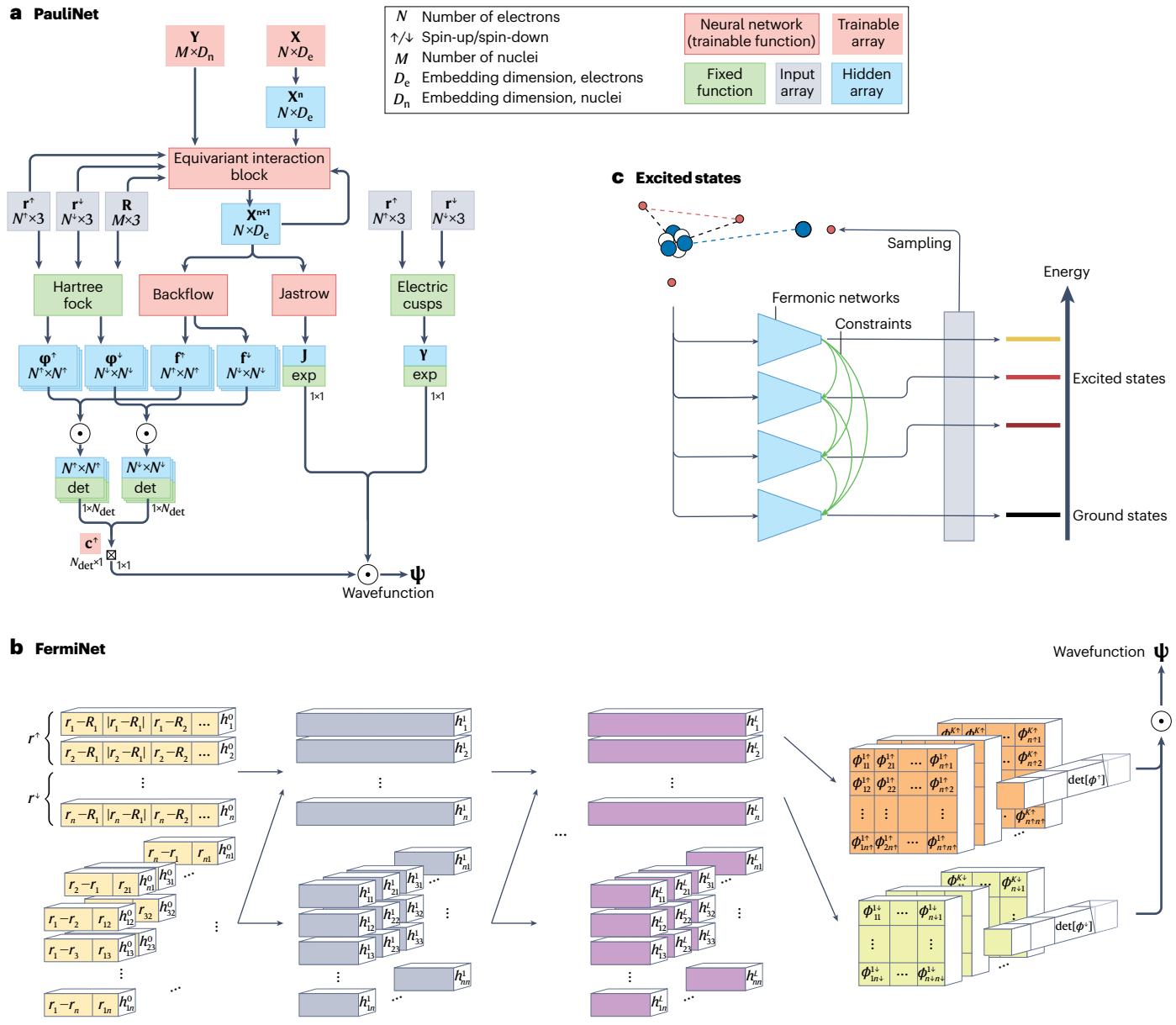
PauliNet. PauliNet¹¹⁶ builds on HF or CASSCF orbitals as a physically meaningful baseline and takes a neural-network approach to the SJB

wavefunction to correct this baseline towards a high-accuracy solution (Fig. 4a). Cusp conditions are explicitly met via the inclusion of cusp correction terms in the wavefunction¹¹⁷. A graph-convolutional block based on SchNet⁷⁸ is used to create a permutation-equivariant latent space representation depending on the many-electron configuration. This embedding is then passed into separate deep neural networks that learn the Jastrow factor and a (cuspless) backflow transformation. Hermann et al.¹¹⁶ introduced PauliNet with a purely multiplicative backflow as shown in Fig. 4a, and Schätzle et al.¹¹⁸ generalized this to a multiplicative and additive backflow as shown in equation (16). PauliNet is optimized with a fixed number of Slater determinants. Most of the results reported in Hermann et al.¹¹⁶ and Schätzle et al.¹¹⁸ were obtained with around 10 determinants.

FermiNet. FermiNet¹¹⁹ takes a more minimalist (or ML maximalist) approach and attempts to train a neural network to represent the entire wavefunction (Fig. 4b). FermiNet uses two parallel networks, describing one-electron and two-electron features, respectively. The inputs to each layer in the one-electron stream are permutation-equivariant functions of the activations from the previous layers of the one-electron and two-electron streams. The final layer projects the latent space into the required number of orbitals, from which determinants can be formed and evaluated. As with PauliNet, the final wavefunction is a sum over a number of determinants. For most of the results reported in Pfau et al.¹¹⁹, 16 determinants were used. FermiNet builds a rich description of electron–electron interactions from the permutation-equivariant mixing of information describing one-electron and two-electron features. In particular, the electron–nuclear and electron–electron cusps in the wavefunction are represented accurately, despite not being encoded explicitly. Although PauliNet is usually trained with the first-order Adam optimizer¹²⁰, FermiNet training was found to be substantially improved when using the Kronecker-factored approximate-curvature optimizer (ref. 121; Box 4).

Although both PauliNet and FermiNet exceed the accuracy of conventional SJB wavefunctions on small systems, there are important trade-offs between the two models. Results from both on the automerization of cyclobutadiene can be seen in Fig. 5. The FermiNet is typically trained with a larger number of parameters than the PauliNet, requiring more iterations and more computation per iteration to converge, but it typically converges to lower absolute energy.

Novel ansatzes. Although the FermiNet and PauliNet were the first neural-network ansatzes in first quantization to outperform other methods, newer neural-network ansatzes have emerged which are even more accurate. Gerard et al.¹²² proposed a hybrid ansatz that uses neural-network layers similar to the SchNet and PauliNet in a FermiNet-like architecture. This hybrid ansatz was found to reach even lower absolute energies than the FermiNet on systems such as benzene and the potassium atom and in fewer iterations. von Glehn et al.¹²³ introduced the Psiformer, which marks a more dramatic departure from the previous work, by replacing the neural-network part of the FermiNet with a sequence of self-attention layers similar to the transformer⁷⁷. The Psiformer was found to reach even lower energies than that obtained by Gerard et al.¹²² on benzene, and on larger systems such as carbon tetrachloride and the benzene dimer, the improvement in performance relative to the FermiNet was as large as 0.1 Ha. Transformers have also been applied to electronic-structure calculations in Xie et al.¹²⁰, although for estimation of the density matrix rather than ground state.



Potential energy surfaces. Typically, one optimizes a wavefunction at a specific geometry, but this quickly becomes prohibitively expensive for exploring the high-dimensional potential energy surface of even relatively small molecules. Scherbera et al.¹²⁴ developed a training methodology that allows weight sharing between (simplified) PauliNet architectures targeting different geometries. By periodically switching the geometry, they showed that the computational cost for training across a set of geometries can be improved by an order of magnitude without affecting the accuracy of the final energies, with 95% of network parameters shared across all geometries. This implies that the network is learning features of electron correlation in general, rather than fitting

to a specific geometry. They also demonstrated that a wavefunction for a larger molecule could be initialized from a wavefunction for a smaller molecule and could then be fine-tuned in a relatively short optimization step. Pretraining neural-network wavefunctions from smaller systems has also been shown to dramatically accelerate convergence for Kagome lattice models¹²⁵.

Similarly, Gao and Günnemann^{126,127} demonstrated a hypernetwork approach (a network that predicts weights of another network), in which a graph neural network is used to parameterize a wavefunction model as a function of nuclear geometries, which can accurately represent the wavefunctions for multiple geometries. This enables a single

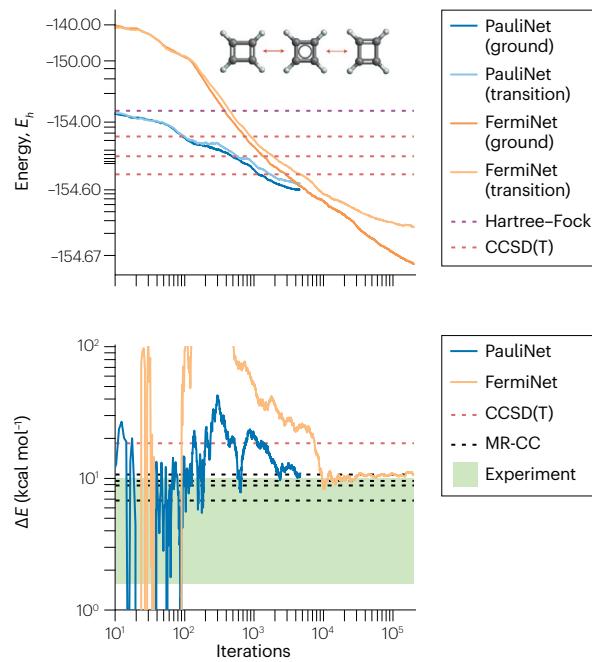


Fig. 5 | Automerization of cyclobutadiene with neural-network ansatzes. Top: PauliNet converges more quickly, whereas FermiNet reaches lower total energy. Hartree–Fock and coupled cluster with singlets, doublets and perturbative triplets (CCSD(T)) absolute energies (three red lines correspond to cc-pVnZ basis sets, $n = D, T, Q$) for the equilibrium geometry are also shown. Bottom: Both PauliNet and FermiNet predict relative energies within the range of experimental values and agree with multireference coupled cluster (MR-CC) methods (different lines denote different MR types of the CC theory). Adapted with permission from ref. 175, Spencer, J. S. et al.

model to fully capture the quantum-mechanical potential energy surface, including generalization to previously unseen geometries. Their approach used a FermiNet-like wavefunction model, but the hypernetwork concept directly applies to other wavefunction representations, assuming that the wavefunction form is sufficiently flexible.

Potential energy surfaces can also be estimated locally by computing forces rather than energies. Although the Hellmann–Feynman theorem¹²⁸ provides an elegant method for computing forces and other gradients of the energy at the true ground state, correction terms are usually necessary when performing calculations from imperfect ansatzes^{129,130}. These techniques have been successfully applied to the computation of forces using the FermiNet as a ground-state ansatz¹³¹.

Periodic systems. There has also been progress on using first-quantized neural-network architectures in periodic systems, such as interacting quantum gases in low dimension¹³², the electron gas^{133–135} and for small cells of solids such as lithium hydride and graphene¹³⁵. Again, sufficiently expressive networks at the VMC level have been found capable of rivalling or surpassing the accuracy of fixed-node diffusion Monte Carlo (DMC) calculations using conventional SJB trial wavefunctions.

Computational cost. Any comparison of the computational cost of the reviewed methods against traditional QC will depend largely on the target property, target accuracy, hardware and system size. On the one

hand, the scaling of the reviewed methods with number of electrons N is favourable; on the other hand, the prefactors are large because neural networks are much more expensive to evaluate than traditional ansatz functions in QC. For the system sizes considered in this Review, and ordinary properties such as binding energies, traditional methods such as coupled cluster are much faster, often by orders of magnitude. This, however, changes when one considers traditional methods able to treat multireference systems – for instance, the computational cost of a few graphics processing unit days of PauliNet on cyclobutadiene is comparable to the many dozens of CPU days of multireference coupled-cluster methods on the same system¹¹⁶. In any case, it is too early to make any conclusions about the efficiency of neural-network QMC, and future work will no doubt focus both on reducing the computational cost and on properly quantifying it.

Extensions

This section describes techniques based on VMC with neural-network ansatzes that in one way or another go beyond ab initio variational ground-state calculations.

Pseudopotentials. The electronic structure of heavy atoms, especially transition metals, is complicated and challenging for all QC methods. The difficulty is compounded by the high computational cost of VMC methods, which scales steeply with the nuclear charge, Z . Although the core electrons contribute heavily to the total energy, energy differences are largely determined by the behaviour of the valence electrons. The core electrons can therefore be removed and the effective nuclear charge is reduced by using pseudopotentials. The use of pseudopotentials is common in many methods, including DFT and conventional VMC. Li et al.¹³⁶ demonstrate that effective core potentials can be readily combined with FermiNet and achieve accuracy comparable to virtually exact methods for first-row transition-metal atoms. The computational time per iteration was reduced by 43% (17%) for the scandium (zinc) atom using an argon core. Again, this approach is not restricted to FermiNet. Pseudopotentials can be used with any first-quantized neural-network wavefunction.

Diffusion Monte Carlo. Projector methods such as DMC¹³⁷ and auxiliary-field Monte Carlo¹³⁸ go beyond VMC by using stochastic algorithms to sample the ground state without requiring its wavefunction to be represented as a known function or network. DMC is exact in principle, but for many-fermion systems relies on the fixed-node approximation in practice. Herein, collapse to the bosonic ground state is avoided by imposing the sign structure of the trial wavefunction on the DMC wavefunction. A DMC simulation, therefore, samples (stochastically) the lowest-energy state with the same sign structure as the trial wavefunction. The improvements that result from applying DMC to conventional SJB trial functions optimized using VMC methods are substantial, explaining why DMC is often used to provide improved estimates of the ground-state wavefunction and energy. Wilson et al.¹³⁹ combined DMC with a FermiNet trial wavefunction. For first-row atoms, DMC captured much of the remaining correlation energy (94% of the difference between the VMC energy and the exact energy in the case of the nitrogen atom). However, Wilson et al.¹³⁹ used a simplified FermiNet that gave VMC energies higher than those reported by Pfau et al.¹¹⁹, which were already within 1 mH of exact results for all first-row atoms. Given the evidence that the mean-field equivalent of PauliNet can essentially match HF in the complete basis set limit¹¹⁸, it is possible that the remaining error in PauliNet and FermiNet wavefunctions is

dominated by errors in the nodal surface, which are rarely sampled regions during optimization. If this is the case, diffusion Monte Carlo with the fixed node approximation might not produce substantially lower energies. By contrast, as neural-network wavefunctions routinely capture more than 90% of the correlation energy at the VMC level, the need to perform expensive diffusion Monte Carlo calculations is greatly reduced. It has been shown that DMC can capture roughly half of the remaining correlation energy for the Li–Ar atoms, when using a very small FermiNet-based architecture¹⁴⁰. Although it is possible to achieve energies within chemical accuracy using FermiNet at the VMC level, these calculations model the case for larger systems in which converging the energy with respect to network size might not be feasible. Ren et al.¹⁴⁰ went on to demonstrate that DMC using FermiNet trial wavefunctions noticeably reduces the energy for larger systems. In the case of the benzene dimer, the reduction was 50 mH.

Excited states. Our discussion so far, and most VMC calculations, have focused on ground-state properties. However, excited states are of critical importance to understanding the behaviour of materials. Fortunately, algorithmic developments by multiple groups have demonstrated that the calculation of excited states using VMC methods is feasible and can achieve an acceptable trade-off in accuracy and cost. Here we highlight three such approaches utilizing conventional VMC wavefunctions. One approach is the state-averaged VMC method^{141,142}, in which the average energy over multiple states is minimized and individual states are projected out via diagonalization within the basis of excited states. Similar techniques are used with other QC methods. Zhao and Neuscammman¹⁴³ instead minimized a different objective function, such that a state with energy closest to the desired energy target is obtained. Pathak et al.¹⁴⁴ suggested a simple alternative, in which a state is forced to be (approximately) orthogonal to all lower energy states via a penalty term. These techniques can be readily applied to VMC using neural-network wavefunctions and as with ground-state calculations, the flexibility of the wavefunction ansatz to represent the desired state is critical. Entwistle et al.¹⁴⁵ demonstrated that the PauliNet architecture combined with a penalty function can represent the lowest few excited states of molecules up to the size of benzene (Fig. 4c). Relatedly, Choo et al.¹⁴⁶ demonstrated that NQS on lattice models can obtain the lowest-energy state of any given Abelian symmetry by performing – what is essentially – a ground-state simulation in that symmetry sector, and multiple states of the same symmetry using a penalty function. However, the most accurate and efficient way to obtain excited states within VMC, irrespective of wavefunction ansatz, remains an open question¹⁴⁷.

Electrons in second quantization

Instead of working directly with the infinite-dimensional Hilbert space – corresponding to the real-space Hamiltonian of equation (2) – it is common practice in QC to use a finite basis set. By choosing a set of electronic basis functions $\{\varphi_1(\mathbf{r}), \varphi_2(\mathbf{r}), \dots\}$, we can define the corresponding second-quantized operators $\hat{c}_i^\dagger (\hat{c}_i)$, which create (annihilate) an electron in the i th basis function and which satisfy the canonical anticommutation relations $\{\hat{c}_i^\dagger, \hat{c}_j\} = \delta_{ij}$. The anti-commutation relations are a direct consequence of the Pauli exclusion principle and of the antisymmetry of the wavefunction with respect to the exchange of electrons.

Projecting the real-space Hamiltonian onto the given set of spin orbitals yields the corresponding discretized Hamiltonian:

$$\hat{H} = \sum_{ij} t_{ij} \hat{c}_i^\dagger \hat{c}_j + \sum_{ijkl} u_{ijkl} \hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_l \hat{c}_k, \quad (20)$$

where

$$t_{ij} = \int \varphi_i^*(\mathbf{r}) \left(-\frac{1}{2} \nabla^2 - \sum_l \frac{Z_l}{|\mathbf{r} - \mathbf{R}_l|} \right) \varphi_j(\mathbf{r}) \, d\mathbf{r}, \quad (21)$$

$$u_{ijkl} = \iint \varphi_i^*(\mathbf{r}) \varphi_j^*(\mathbf{r}') \frac{1}{|\mathbf{r} - \mathbf{r}'|} \varphi_k(\mathbf{r}) \varphi_l(\mathbf{r}') \, d\mathbf{r} d\mathbf{r}', \quad (22)$$

are matrix elements of the one-electron and two-electron terms in the real-space Hamiltonian of equation (2). The matrix elements can be evaluated analytically for simple basis functions such as Gaussians or plane waves. In this framework, the many-electron wavefunction is expressed as $\psi(n_1, n_2, \dots)$, thus encoding the amplitudes for different occupations of the orbitals (Box 2). It should be remarked that there are several non-equivalent ways of defining the occupation-number basis states, $|n_1, n_2, \dots\rangle$. The canonical ordering often adopted in QC corresponds to the Jordan–Wigner mapping¹⁴⁸, which transforms annihilation and creation operators into, respectively, lowering and raising spin operators. However, this mapping is not unique, and alternatives exist, such as parity or Bravyi–Kitaev encodings¹⁴⁹, both of which have been developed in the context of quantum simulations.

Overall, this occupation-number formalism and the corresponding Hamiltonian (equation (20)) serve as the starting point for the methods described in this section.

Fermionic neural quantum states

The many-body amplitudes (values of the wavefunction) in the occupation-number representation can be readily expressed in terms of neural networks taking discrete inputs. As the antisymmetry constraint is fully encoded in the Hamiltonian (equation (20)), the amplitudes $\psi(n_1, n_2, \dots)$ do not carry any specific symmetry and any network architecture can be used to represent them. This observation allows, for example, to directly use NQS representations on the basis of complex-valued RBMs⁴⁸, originally introduced to study spin systems. In this case, for a system with L spin orbitals, the many-body amplitude corresponding to a given occupation number takes the compact form

$$\psi(n_1, n_2, \dots, n_L; \boldsymbol{\theta}) = e^{\sum_i a_i n_i} \prod_{j=1}^M 2 \cosh \left(b_j + \sum_i W_{ij} n_i \right), \quad (23)$$

with parameters $\boldsymbol{\theta} = (a_i, b_j, W_{ij})$.

This ansatz can be optimized with VMC techniques (Box 4), in which the occupation numbers are sampled according to the probability density that is proportional to $|\psi(n_1, n_2, \dots)|^2$. The corresponding local energy estimator $E_{\text{loc}}(n_1, n_2, \dots)$ can be computed taking into account the matrix elements of the Hamiltonian (equation (20)) in the occupation basis. Optimization of the wavefunction ansatz typically relies on the stochastic reconfiguration¹⁵⁰ approach. A number of works have adopted this simple NQS wavefunction and achieved competitive variational results for relatively small basis sets^{151,152}, even in conjunction with quantum computers^{153,154}. In Fig. 6a, we show the dissociation curve of C₂, in the STO-3G basis, using the RBM as described earlier¹⁵¹.

An alternative ML-based variational ansatz for second-quantized Hamiltonians based on Gaussian process regression has also been proposed¹⁵⁵, which has reached accuracy comparable to the NQS approach on model Hamiltonians, but has not yet been applied to the ab initio QC Hamiltonian.

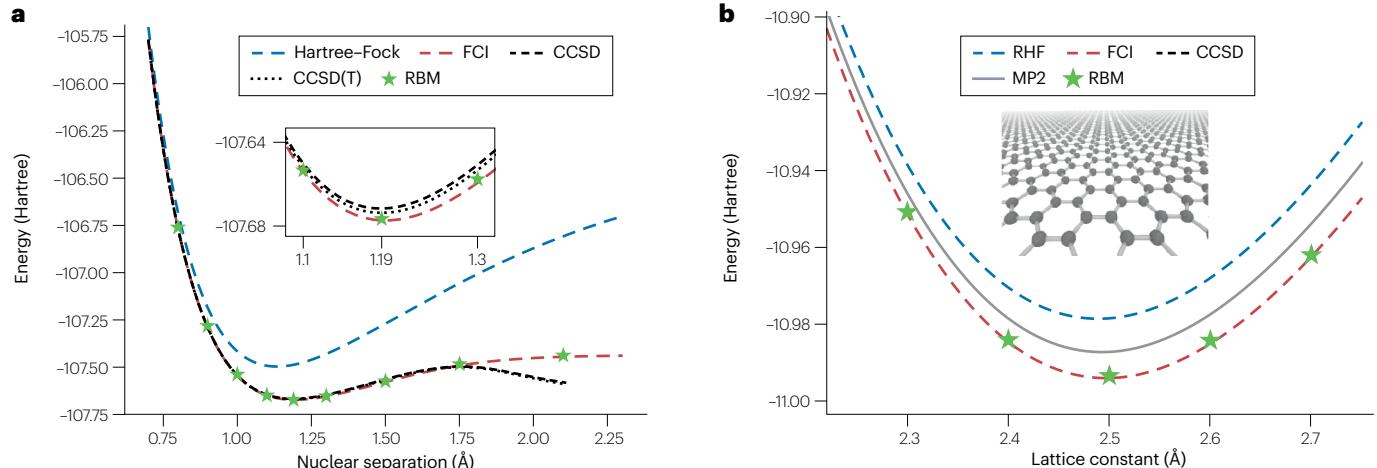


Fig. 6 | Electronic energies for molecules and solids in second quantization.

a, Dissociation curve for N₂ molecule in the STO-3G basis. The green stars show results for a restricted Boltzmann machine (RBM) that represents the electrons in discrete space. **b**, Graphene on a honeycomb lattice solved using the cc-pVDZ

basis set. CCSD(T), coupled cluster with singlets, doublets and perturbative triplets; FCI, full configuration interaction; MP2, Møller–Plesset perturbation theory to second order; RHF, restricted Hartree–Fock. Part **a** is adapted from ref. 151, CC BY 4.0. Part **b** is reprinted from ref. 157, CC BY 4.0.

Solids. The second-quantization framework also allows one to treat solids, using as a basis the Bloch orbitals¹⁵⁶. Creation and annihilation operators, \hat{c}_{ik}^\dagger and \hat{c}_{ik} , respectively, for electrons in the band i with crystal momentum k are introduced into the model, and the resulting Hamiltonian is similar to equation (20), with the noticeable difference that the one-body and two-body matrix elements now depend on the crystal momenta: $t_{ij} \rightarrow t_{ij}^k$ and $u_{ijkl} \rightarrow u_{ijkl}^{k_1 k_2 k_3 k_4}$, with the four momenta (\mathbf{k}) appearing in the two-body integrals satisfying the conservation of the total crystal momentum. Using Gaussian-based atomic functions as the single-particle basis and RBM wavefunctions to represent the many-body state¹⁵⁷, this approach can be applied to studying the electronic structure of solids. In Fig. 6b, we show the computed ground-state energies for graphene crystals as a function of the lattice constant.

Exact sampling. Fermionic NQS are typically sampled using the Markov-chain Monte Carlo (MCMC) approach commonly adopted in VMC (Box 3). However, the mixing rate of the MCMC algorithm is known to be slow in some cases – such as near-phase transitions – and MCMC simulations can suffer from critical slowing down. A way to circumvent this limitation is to introduce model wavefunctions explicitly designed to allow exact sampling of their square modulus, thus avoiding the need to use MCMC. One such family is autoregressive neural-network wavefunctions⁸⁷, a complex-valued generalization of the autoregressive models commonly adopted in deep learning. Such networks represent normalized wavefunctions and allow one to obtain perfectly uncorrelated samples directly; this is useful as the wavefunction distribution for many QC problems can be highly multimodal. The exact sampling approach has been applied to QC Hamiltonians¹⁵⁸. Optimizations in how Hamiltonian matrix elements and the corresponding Monte Carlo estimators are computed have made it possible to treat much larger systems than were accessible in early applications¹⁵¹. Specifically, Zhao et al.¹⁵⁹ obtain competitive variational energies, improving on the CCSD energies of molecules in minimal basis sets. Results for up to around 50 electrons in 80 orbitals (Na₂CO₃ at equilibrium) have been obtained at a relatively modest computational cost.

ML-assisted selected CI

For many QC problems, although the dimension of the Hilbert space grows exponentially with system size, the number of relevant configurations in the ground state typically remains sparse. This suggests that by efficiently selecting the relevant configurations and then diagonalizing the Hamiltonian on the reduced subspace, one can achieve highly accurate results. This set of approaches is also known as selected CI^{45,160–162}. Different types of selected CI vary in the way relevant configurations are selected.

One well-known approach is called Monte Carlo CI (MCCI)¹⁶³ and can be briefly summarized in simple steps, starting from a finite set of configurations (S_i). Following this, and by considering single or double excitations starting from configurations in S_i , an expanded set S'_i can be constructed. Next, construct the Hamiltonian \hat{H} for the expanded set S'_i and diagonalize to obtain the wavefunction coefficients for the configurations in the set. Finally, discard the configurations whose coefficient is less than a given threshold c_{\min} . The remaining configurations then form a new set of configurations S_{i+1} , after which the procedure is repeated until convergence.

ML techniques can be used to improve selection of the configuration set. One such approach is to perform supervised learning¹⁶⁴, in which a neural network is trained to predict the wavefunction coefficients using the data from the MCCI method, for example, the wavefunction coefficients of the configurations in the set S'_i . After training, the network can be queried or sampled to select the configurations with the largest coefficients. In other words, the network is used to bootstrap and predict the coefficients of configurations not yet seen in the data set. It was shown in Coe¹⁶⁴ that such an approach converges faster than the vanilla MCCI method.

The task of selecting configurations for selected CI can also be cast as a reinforcement-learning task, in which the state is the current set of configurations and an agent is trained to perform actions on the set to iteratively modify the configurations with the aim of minimizing the variational energy. This approach was applied in Goings et al.¹⁶⁵ to achieve near-FCI accuracy for small molecules in a small basis set. Another option is to formulate the problem as a classification task to

decide whether to include a configuration, which is the setup used by Chembot¹⁶⁶.

Challenges and outlook

Ab initio QC with neural-network wavefunctions has only just emerged as a viable path to highly accurate electronic-structure methods, yet it already competes on small systems with established approaches that have been developed for decades. We imagine that it can become the methodology with the best trade-off between efficiency and accuracy for systems with up to 100–200 electrons and a non-trivial electronic structure, which are already too large for exact methods, and for which conventional high-accuracy methods can fail. Before that can happen, however, several challenges must be addressed.

All reviewed methods are currently in a development stage, and only limited benchmarking is available. As such, it is not yet clear whether the excellent accuracy seen on small systems will be maintained across a broader range of chemical systems and, perhaps more importantly, across system sizes. The latter is referred to as size consistency and size extensivity¹⁶⁷, and these have not yet been established for any of the methods reviewed earlier. Size consistency and extensivity are crucial for a method to satisfy, for it to predict interaction energies (energy differences) accurately, and most conventional QC methods are size-consistent either exactly by construction or approximately to a good degree. It is reasonable to imagine that the fixed network size or the fixed number of Slater determinants in the reviewed methods could introduce size inconsistency. Truncated determinant expansions are certainly the source of severe size inconsistency in conventional methods. The degree to which the increased expressiveness of the individual determinants parameterized with neural networks can mitigate this is yet to be seen. One could also imagine scaling the network size with system size, but precisely how that would need to be done is likewise an open question. A certain degree of size inconsistency does not necessarily render a method unusable, but it certainly makes its reliable application more difficult. Alternatively, a modification that restores size consistency would have to be found, but this could potentially introduce additional scaling with system size. Whether this is the case is left for future work to establish.

A related issue is our incomplete understanding of what limits the accuracy of neural-network ansatzes, and how their success or failure is related to physical phenomena such as strong correlation. As the underlying electronic problem is exponentially hard but the algorithms are polynomial, they must be limited in accuracy in some ways. It is not currently clear, however, whether the limitations seen to date are caused by the restricted expressiveness of the neural networks or by difficulties in optimization or both. For instance, although it has been proven that a single generalized Slater determinant is in principle sufficient to represent any antisymmetric function and that an equivariant neural network can represent any equivariant generalized orbital¹⁶⁸, it might not be possible to train it within a polynomially scaling time. We note that in practice, several generalized determinants are usually necessary to reach high accuracy, suggesting that these proofs do not translate to practically meaningful results.

Apart from these fundamental issues, there are many practical challenges. Although the formal quartic scaling of backflow-based variational QMC with system size is favourable compared with traditional QC methods, the prefactor owing to the neural networks is large. This limited applications to systems no larger than the benzene molecule (42 electrons), which is three to four times below our envisaged applicability range, although results for a 108-electron simulation cell of

solid LiH have now been reported¹³⁵. The prefactor can be reduced by integrating traditional QC techniques such as pseudopotentials¹³⁶, developing more efficient neural-network architectures or using ML techniques such as pretraining and transfer learning. Specific to the discrete-basis second-quantized approaches is the issue of basis-set convergence, in which sufficiently large basis sets can increase the prefactor by up to three orders of magnitude compared with minimal basis sets. Another challenge is related to stochastic optimization, which produces noise in the converged energies that is especially amplified when calculating small energy differences and dependent on a particular system and ansatz in a way that is currently poorly understood. This issue can often be resolved in practice by averaging individual total energies over multiple optimization runs, at the cost of increased computational effort.

We are, however, optimistic that many of these challenges can be addressed quickly, owing to the relative simplicity of the framework on the basis of variational QMC compared with conventional QC approaches and to the rate of innovation in deep learning architectures. Indeed, the simplicity of the approach has already enabled rapid development of multiple extensions to the first single-point ground-state calculations on molecules, including transferable wavefunctions, excited states and formulations for periodic systems, all originating from multiple independent research groups.

The high accuracy reached by variational QMC with neural-network wavefunctions on small systems makes us optimistic that such technologies can make progress on quantum systems that were challenging with existing QC approaches, including strongly correlated electronic states such as certain transition states, transition-metal complexes and many excited states. On a number of select, small test systems, the novel neural-network architectures reviewed earlier have already reached thermochemical (sub-millihartree) accuracy in ionization potentials and electron affinities, and rival or in some cases even exceed the accuracy of advanced conventional QC methods. Yet these networks are just a small subset of possible architectures for representing antisymmetric wavefunctions, and it is unlikely that the optimal ones would be found on the first attempt, so we expect that considerable innovation in the pursuit of better efficiency or accuracy trade-off lies ahead.

An important aspect of variational QMC is that solutions can be ranked systematically from better to worse together with other variational methods. This enables the construction of ‘leaderboards’ of high-accuracy solutions of quantum states of small molecules, which could serve to produce high-quality training data for methods that are computationally more efficient but a priori less accurate, such as machine-learned density functionals and neural-network potentials.

Another promising aspect is the flexibility of neural-network wavefunctions, which not only leads to high accuracy, but perhaps more importantly might enable us to get accurate results for non-standard Hamiltonians, for which no efficient off-the-shelf methods exist. Examples include Hamiltonians with electron–photon and electron–phonon interactions, which are important for the computation of certain material properties, superconductivity and superfluidity.

Finally, the flexibility of the wavefunction ansatz in principle also allows researcher to generalize across different nuclear geometries, such as the variational learning over entire potential energy surfaces with one model, as well as to generalize across chemical space. Although the first steps towards the former goal have already been made and reviewed in this article, progress on the latter will likely require development of new neural-network architectures. This paradigm of training large ‘foundation’ models or ‘emulators’, which can

later be used for inference at a much reduced cost, has been successfully used in nature language processing and protein structure prediction, and we believe that it is also possible to build similar systems for electronic-structure prediction.

Overall, we are confident that *ab initio* methods based on neural-network wavefunctions will become an integral part of the QC toolbox that will enable straightforward electronic-structure calculations of complex molecular systems.

Published online: 9 August 2023

References

- Carleo, G. et al. Machine learning and the physical sciences. *Rev. Mod. Phys.* **91**, 045002 (2019).
- Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
- Deringer, V. L. et al. Origins of structural and electronic transitions in disordered silicon. *Nature* **589**, 59–64 (2021).
- Noé, F., Olsson, S., Köhler, J. & Wu, H. Boltzmann generators: sampling equilibrium states of many-body systems with deep learning. *Science* **365**, eaaw1147 (2019).
- Huang, B. & von Lilienfeld, O. A. Quantum machine learning using atom-in-molecule-based fragments selected on the fly. *Nat. Chem.* **12**, 945–951 (2020).
- Tkatchenko, A. Machine learning for chemical discovery. *Nat. Commun.* **11**, 4125 (2020).
- von Lilienfeld, O. A. & Burke, K. Retrospective on a decade of machine learning for chemical discovery. *Nat. Commun.* **11**, 4895 (2020).
- Noé, F., Tkatchenko, A., Müller, K.-R. & Clementi, C. Machine learning for molecular simulation. *Annu. Rev. Phys. Chem.* **71**, 361–390 (2020).
- Dral, P. O. Quantum chemistry in the age of machine learning. *J. Phys. Chem. Lett.* **11**, 2336–2347 (2020).
- Unke, O. T. et al. Machine learning force fields. *Chem. Rev.* **121**, 10142–10186 (2021).
- von Lilienfeld, O. A., Müller, K.-R. & Tkatchenko, A. Exploring chemical compound space with quantum-based machine learning. *Nat. Rev. Chem.* **4**, 347–358 (2020).
- Bian, Y. & Xie, X.-Q. Generative chemistry: drug discovery with deep learning generative models. *J. Mol. Model.* **27**, 71 (2021).
- Jones, R. O. Density functional theory: its origins, rise to prominence, and future. *Rev. Mod. Phys.* **87**, 897–923 (2015).
- Bartlett, R. J. & Musiāk, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.* **79**, 291–352 (2007).
- Deringer, V. L. et al. Gaussian process regression for materials and molecules. *Chem. Rev.* **121**, 10073–10141 (2021).
- Behler, J. Four generations of high-dimensional neural network potentials. *Chem. Rev.* **121**, 10037–10072 (2021).
- Musil, F. et al. Physics-inspired structural representations for molecules and materials. *Chem. Rev.* **121**, 9759–9815 (2021).
- Li, H., Collins, C., Tanha, M., Gordon, G. J. & Yaron, D. J. A density functional tight binding layer for deep learning of chemical Hamiltonians. *J. Chem. Theory Comput.* **14**, 5764–5776 (2018).
- Schütt, K. T., Gastegger, M., Tkatchenko, A., Müller, K.-R. & Maurer, R. J. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nat. Commun.* **10**, 5024 (2019).
- Kirkpatrick, J. et al. Pushing the frontiers of density functionals by solving the fractional electron problem. *Science* **374**, 1385–1389 (2021).
- Chandrasekaran, A. et al. Solving the electronic structure problem with machine learning. *Comput. Mater.* **5**, 1–7 (2019).
- Welborn, M., Cheng, L. & Miller III, T. F. Transferability in machine learning for electronic structure via the molecular orbital basis. *J. Chem. Theory Comput.* **14**, 4772–4779 (2018).
- Nagai, R., Akashi, R. & Sugino, O. Completing density functional theory by machine learning hidden messages from molecules. *Comput. Mater.* **6**, 1–8 (2020).
- Gómez-Bombarelli, R. et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **4**, 268–276 (2018).
- Hoogeboom, E., Satorras, V. G., Vignac, C. & Welling, M. Equivariant diffusion for molecule generation in 3D. Preprint at <http://arxiv.org/abs/2203.17003> (2022).
- Torlai, G. et al. Neural-network quantum state tomography. *Nat. Phys.* **14**, 447–450 (2018).
- Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT Press, 2018).
- Tesauro, G. TD-Gammon, a self-teaching backgammon program, achieves master-level play. *Neural Comput.* **6**, 215–219 (1994).
- Mnih, V. et al. Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015).
- Silver, D. et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).
- Degrave, J. et al. Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature* **602**, 414–419 (2022).
- Heinrich, J., Lanctot, M. & Silver, D. Fictitious self-play in extensive-form games. In *Proceedings of the 32nd International Conference on Machine Learning* 805–813 (PMLR, 2015).
- Silver, D. et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science* **362**, 1140–1144 (2018).
- Battaglia, S. Machine learning wavefunction. In *Quantum Chemistry in the Age of Machine Learning* 577–616 (Elsevier, 2023).
- Manzhos, S. Machine learning for the solution of the Schrödinger equation. *Mach. Learn. Sci. Techn.* **1**, 013002 (2020).
- Pielas, L. *Ideas of Quantum Chemistry* 2nd edn (Elsevier, 2014).
- Foulkes, W. M. C., Mitas, L., Needs, R. J. & Rajagopal, G. Quantum Monte Carlo simulations of solids. *Rev. Mod. Phys.* **73**, 33–83 (2001).
- Bajdich, M., Mitas, L., Dröbný, G., Wagner, L. K. & Schmidt, K. E. Pfaffian pairing wave functions in electronic-structure quantum Monte Carlo Simulations. *Phys. Rev. Lett.* **96**, 130201 (2006).
- Han, J., Zhang, L. & Weinan, E. Solving many-electron Schrödinger equation using deep neural networks. *J. Comput. Phys.* **399**, 108929 (2019).
- Acevedo, A., Curry, M., Joshi, S. H., Leroux, B. & Malaya, N. Vandermonde wave function ansatz for improved variational Monte Carlo. In *2020 IEEE/ACM Fourth Workshop on Deep Learning on Supercomputers (DLS)* 40–47 (IEEE, 2020).
- Szabo, A. & Ostlund, N. S. *Modern Quantum Chemistry* (Dover Publications, 1996).
- Becca, F. & Sorella, S. *Quantum Monte Carlo Approaches for Correlated Systems* 1st edn (Cambridge Univ. Press, 2017).
- Teale, A. M. et al. DFT exchange: sharing perspectives on the workhorse of quantum chemistry and materials science. *Phys. Chem. Chem. Phys.* **24**, 28700–28781 (2022).
- Chan, G. K.-L. & Sharma, S. The density matrix renormalization group in quantum chemistry. *Annu. Rev. Phys. Chem.* **62**, 465 (2011).
- Huron, B., Malrieu, J. P. & Rancurel, P. Iterative perturbation calculations of ground and excited state energies from multiconfigurational zeroth-order wavefunctions. *J. Chem. Phys.* **58**, 5745–5759 (1973).
- Booth, G. H., Thom, A. J. W. & Alavi, A. Fermion Monte Carlo without fixed nodes: a game of life, death, and annihilation in Slater determinant space. *J. Chem. Phys.* **131**, 054106 (2009).
- Olsen, J. The CASSCF method: a perspective and commentary: CASSCF Method. *Int. J. Quantum Chem.* **111**, 3267–3272 (2011).
- Carleo, G. & Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **355**, 602–606 (2017).
- Saito, H. Solving the Bose-Hubbard model with machine learning. *J. Phys. Soc. Jpn.* **86**, 093001 (2017).
- Nomura, Y., Darmawan, A. S., Yamaji, Y. & Imada, M. Restricted Boltzmann machine learning for solving strongly correlated quantum systems. *Phys. Rev. B* **96**, 205152 (2017).
- Adams, C., Carleo, G., Lovato, A. & Rocco, N. Variational Monte Carlo calculations of A≤4 nuclei with an artificial neural-network correlator ansatz. *Phys. Rev. Lett.* **127**, 022502 (2021).
- Astrakhantsev, N. et al. Broken-symmetry ground states of the Heisenberg model on the pyrochlore lattice. *Phys. Rev. X* **11**, 041021 (2021).
- Perronnin, F., Liu, Y., Sánchez, J. & Poirier, H. Large-scale image retrieval with compressed fisher vectors. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 3384–3391 (IEEE, 2010).
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* Vol. 25, 1097–1105 (Curran Associates, 2012).
- Schmidhuber, J. Deep learning in neural networks: an overview. *Neural Netw.* **61**, 85–117 (2015).
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- McCulloch, W. S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **5**, 115–133 (1943).
- Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **65**, 386 (1958).
- Hornik, K., Stinchcombe, M. & White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **2**, 359–366 (1989).
- Werbos, P. Beyond regression: new tools for prediction and analysis in the behavioral sciences. PhD thesis. Harvard Univ. (1974).
- Linnainmaa, S. The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors. Master's thesis (in Finnish), Univ. Helsinki (1970).
- Linnainmaa, S. Taylor expansion of the accumulated rounding error. *BIT Numer. Math.* **16**, 146–160 (1976).
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
- Glorot, X. & Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics* 249–256 (PMLR, 2010).
- Hooker, S. The hardware lottery. *Commun. ACM* **64**, 58–65 (2021).
- Dauphin, Y. et al. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. In *Advances in Neural Information Processing Systems* Vol. 27, 2933–2941 (Curran Associates, 2014).
- Choromanska, A., Henaff, M., Mathieu, M., Arous, G. B. & LeCun, Y. The loss surfaces of multilayer networks. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics* 192–204 (PMLR, 2015).
- Bottou, L. & Bousquet, O. Learning using large datasets. In *Mining Massive Data Sets for Security* 15–26 (IOS Press, 2008).

69. Bottou, L. & Bousquet, O. The tradeoffs of large-scale learning. In *Optimization for Machine Learning*, 351–368 (MIT Press, 2011).
70. Russakovsky, O. et al. ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**, 211–252 (2015).
71. Abadi, M. et al. TensorFlow: a system for large-scale machine learning. In *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation* 265–283 (USENIX Association, 2016).
72. Paszke, A. et al. Automatic differentiation in PyTorch. In *NIPS Workshop on Automatic Differentiation* (2017).
73. Bradbury, J. et al. JAX: Composable Transformations of Python+ NumPy Programs (GitHub, 2018); <https://github.com/google/jax>.
74. Goodfellow, I. J., Shlens, J. & Szegedy, C. Explaining and harnessing adversarial examples. *Third International Conference on Learning Representations (ICLR)* (ICLR, 2015).
75. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
76. Shawe-Taylor, J. Building symmetries into feedforward networks. In *1989 First IEEE International Conference on Artificial Neural Networks* 158–162 (IET, 1989).
77. Vaswani, A. et al. Attention is all you need. In *Advances in Neural Information Processing Systems* Vol. 30, 5998–6008 (Curran Associates, 2017).
78. Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. SchNet — a deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
79. Bronstein, M. M., Bruna, J., Cohen, T. & Veličković, P. Geometric deep learning: grids, groups, graphs, geodesics, and gauges. Preprint at <http://arxiv.org/abs/2104.13478> (2021).
80. Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science* **313**, 504–507 (2006).
81. Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. Preprint at <http://arxiv.org/abs/1312.6114> (2013).
82. Goodfellow, I. et al. Generative adversarial nets. In *Advances in Neural Information Processing Systems* Vol. 27, 2672–2680 (Curran Associates, 2014).
83. Rezende, D. & Mohamed, S. Variational inference with normalizing flows. In *Proceedings of the 32nd International Conference on Machine Learning* 1530–1538 (PMLR, 2015).
84. van den Oord, A. et al. WaveNet: a generative model for raw audio. Preprint at <http://arxiv.org/abs/1609.3499> (2016).
85. van den Oord, A. et al. Conditional image generation with PixelCNN decoders. In *Advances in Neural Information Processing Systems* Vol. 29, 4797–4805 (Curran Associates, 2016).
86. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N. & Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning* 2256–2265 (PMLR, 2015).
87. Sharir, O., Levine, Y., Wies, N., Carleo, G. & Shashua, A. Deep autoregressive models for the efficient variational simulation of many-body quantum systems. *Phys. Rev. Lett.* **124**, 020503 (2020).
88. Choo, K., Neupert, T. & Carleo, G. Two-dimensional frustrated J_1 - J_2 model studied with neural network quantum states. *Phys. Rev. B* **100**, 125124 (2019).
89. Hibat-Allah, M., Ganahl, M., Hayward, L. E., Melko, R. G. & Carrasquilla, J. Recurrent neural network wave functions. *Phys. Rev. Res.* **2**, 023358 (2020).
90. Xie, H., null, L. Z. & Wang, L. Ab-initio study of interacting fermions at finite temperature with neural canonical transformation. *J. Mach. Learn.* **1**, 38 (2022).
91. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
92. Rupp, M., Tkatchenko, A., Müller, K.-R. & von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **108**, 058301 (2012).
93. Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).
94. Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K. R. & Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **8**, 13890 (2017).
95. Thomas, N. et al. Tensor field networks: rotation- and translation-equivariant neural networks for 3D point clouds. Preprint at <https://arxiv.org/abs/1802.08219> (2018).
96. Schütt, K. T., Unke, O. T. & Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *Proceedings of the 38th International Conference on Machine Learning* 9377–9388 (PMLR, 2021).
97. Miller, B. K., Geiger, M., Smidt, T. E. & Noé, F. Relevance of rotationally equivariant convolutions for predicting molecular properties. Preprint at <https://arxiv.org/abs/2008.08461> (2020).
98. Geiger, M. & Smidt, T. e3nn: Euclidean neural networks. Preprint at <http://arxiv.org/abs/2207.09453> (2022).
99. Battatia, I., Kovács, D. P., Simm, G. N., Ortner, C. & Csányi, G. MACE: higher order equivariant message passing neural networks for fast and accurate force fields. In *Advances in Neural Information Processing Systems* Vol. 35, 11423–11436 (Curran Associates, 2022).
100. Huang, X., Braams, B. J. & Bowman, J. M. Ab initio potential energy and dipole moment surfaces for H_2O^{2+} . *J. Chem. Phys.* **122**, 44308 (2005).
101. Drautz, R. Atomic cluster expansion for accurate and transferable interatomic potentials. *Phys. Rev. B* **99**, 014104 (2019).
102. Allen, A. E. A., Dušson, G., Ortner, C. & Csányi, G. Atomic permutationally invariant polynomials for fitting molecular force fields. *Mach. Learn.: Sci. Tech.* **2**, 025017 (2021).
103. Benali, A. et al. Toward a systematic improvement of the fixed-node approximation in diffusion Monte Carlo for solids — A case study in diamond. *J. Chem. Phys.* **153**, 184111 (2020).
104. Stokes, J., Isaac, J., Killoran, N. & Carleo, G. Quantum natural gradient. *Quantum* **4**, 269 (2020).
105. Feynman, R. P. & Cohen, M. Energy spectrum of the excitations in liquid helium. *Phys. Rev.* **102**, 1189–1204 (1956).
106. Kwon, Y., Ceperley, D. M. & Martin, R. M. Effects of three-body and backflow correlations in the two-dimensional electron gas. *Phys. Rev. B* **48**, 12037–12046 (1993).
107. Tocchio, L. F., Becca, F., Parola, A. & Sorella, S. Role of backflow correlations for the nonmagnetic phase of the t - t' Hubbard model. *Phys. Rev. B* **78**, 041101 (2008).
108. Luo, D. & Clark, B. K. Backflow transformations via neural networks for quantum many-body wave functions. *Phys. Rev. Lett.* **122**, 226401 (2019).
109. Robledo Moreno, J., Carleo, G., Georges, A. & Stokes, J. Fermionic wave functions from neural-network constrained hidden states. *Proc. Natl Acad. Sci. USA* **119**, e2122059119 (2022).
110. Lovato, A., Adams, C., Carleo, G. & Rocco, N. Hidden-nucleons neural-network quantum states for the nuclear many-body problem. *Phys. Rev. Res.* **4**, 043178 (2022).
111. Yang, Y. & Zhao, P. Deep-neural-network approach to solving the ab initio nuclear structure problem. *Phys. Rev. C* **107**, 034320 (2023).
112. Taddei, M., Ruggeri, M., Moroni, S. & Holzmann, M. Iterative backflow renormalization procedure for many-body ground-state wave functions of strongly interacting normal Fermi liquids. *Phys. Rev. B* **91**, 115106 (2015).
113. Ruggeri, M., Moroni, S. & Holzmann, M. Nonlinear network description for many-body quantum systems in continuous space. *Phys. Rev. Lett.* **120**, 205302 (2018).
114. Chakravorty, S. J., Gwaltney, S. R. & Davidson, E. R. Ground-state correlation energies for atomic ions with 18 electrons. *Phys. Rev. A* **44**, 7071 (1991).
115. Hermann, J., Schätzle, Z. & Noé, F. Deep-neural-network solution of the electronic Schrödinger equation. *Nat. Chem.* **12**, 891–897 (2020).
116. Ma, A., Towler, M. D., Drummond, N. D. & Needs, R. J. Scheme for adding electron-nucleus cusps to Gaussian orbitals. *J. Chem. Phys.* **122**, 224322 (2005).
117. Schätzle, Z., Hermann, J. & Noé, F. Convergence to the fixed-node limit in deep variational Monte Carlo. *J. Chem. Phys.* **154**, 124108 (2021).
118. Pfau, D., Spencer, J. S., Matthews, A. G. D. G. & Foulkes, W. M. C. Ab initio solution of the many-electron Schrödinger equation with deep neural networks. *Phys. Rev. Res.* **2**, 033429 (2020).
119. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at <http://arxiv.org/abs/1412.6980> (2015).
120. Martens, J. & Grosse, R. Optimizing neural networks with kronecker-factored approximate curvature. In *Proceedings of the 32nd International Conference on Machine Learning* 2408–2417 (PMLR, 2015).
121. Gerard, L., Scherbel, M., Marquetand, P. & Grohs, P. Gold-standard solutions to the Schrödinger equation using deep learning: how much physics do we need? In *Advances in Neural Information Processing Systems*, Vol. 35, 10282–10294 (Curran Associates, 2022).
122. von Glehn, I., Spencer, J. S. & Pfau, D. A self-attention ansatz for ab-initio quantum chemistry. In *International Conference on Learning Representations (ICLR 2023)* (OpenReview, 2023).
123. Scherbel, M., Reisenhofer, R., Gerard, L., Marquetand, P. & Grohs, P. Solving the electronic Schrödinger equation for multiple nuclear geometries with weight-sharing deep neural networks. *Nat. Comput. Sci.* **2**, 331 (2022).
124. Yang, L., Hu, W. & Li, L. Scalable variational Monte Carlo with graph neural ansatz. Preprint at <http://arxiv.org/abs/2011.12453> (2020).
125. Gao, N. & Günemann, S. Ab-initio potential energy surfaces by pairing GNNs with neural wave functions. In *International Conference on Learning Representations (ICLR 2022)* (OpenReview, 2022).
126. Gao, N. & Günemann, S. Sampling-free inference for ab-initio potential energy surface networks. In *International Conference on Learning Representations (ICLR 2023)* (OpenReview, 2023).
127. Feynman, R. P. Forces in molecules. *Phys. Rev.* **56**, 340 (1939).
128. Assaraf, R. & Caffarel, M. Zero-variance zero-bias principle for observables in quantum Monte Carlo: application to forces. *J. Chem. Phys.* **119**, 10536–10552 (2003).
129. Umrigar, C. Two aspects of quantum Monte Carlo: determination of accurate wavefunctions and determination of potential energy surfaces of molecules. *Int. J. Quantum Chem.* **36**, 217–230 (1989).
130. Qian, Y., Fu, W., Ren, W. & Chen, J. Interatomic force from neural network based variational quantum Monte Carlo. *J. Chem. Phys.* **157**, 164104 (2022).
131. Pescia, G., Han, J., Lovato, A., Lu, J. & Carleo, G. Neural-network quantum states for periodic systems in continuous space. *Phys. Rev. Res.* **4**, 023138 (2022).
132. Wilson, M. et al. Neural network ansatz for periodic wave functions and the homogeneous electron gas. *Phys. Rev. B* **107**, 235139 (2023).
133. Cassella, G. et al. Discovering quantum phase transitions with fermionic neural networks. *Phys. Rev. Lett.* **130**, 036401 (2023).
134. Li, X., Li, Z. & Chen, J. Ab initio calculation of real solids via neural network ansatz. *Nat. Commun.* **13**, 7895 (2022).
135. Li, X., Fan, C., Ren, W. & Chen, J. Fermionic neural network with effective core potential. *Phys. Rev. Res.* **4**, 013021 (2022).

137. Needs, R. J., Towler, M. D., Drummond, N. D., López Ríos, P. & Trail, J. R. Variational and diffusion quantum Monte Carlo calculations with the CASINO code. *J. Chem. Phys.* **152**, 154106 (2020).
138. Shi, H. & Zhang, S. Some recent developments in auxiliary-field quantum Monte Carlo for real materials. *J. Chem. Phys.* **154**, 024107 (2021).
139. Wilson, M., Gao, N., Wudarski, F., Rieffel, E. & Tubman, N. M. Simulations of state-of-the-art fermionic neural network wave functions with diffusion Monte Carlo. Preprint at <http://arxiv.org/abs/2103.12570> (2021).
140. Ren, W., Fu, W. & Chen, J. Towards the ground state of molecules via diffusion Monte Carlo on neural networks. *Nat. Commun.* **14**, 1860 (2023).
141. Schautz, F. & Filippi, C. Optimized Jastrow–Slater wave functions for ground and excited states: application to the lowest states of ethene. *J. Chem. Phys.* **120**, 10931 (2004).
142. Dash, M., Feldt, J., Moroni, S., Scemama, A. & Filippi, C. Excited states with selected configuration interaction–quantum Monte Carlo: chemically accurate excitation energies and geometries. *J. Chem. Theory Comput.* **15**, 4896 (2019).
143. Zhao, L. & Neuscamman, E. An efficient variational principle for the direct optimization of excited states. *J. Chem. Theory Comput.* **12**, 3436 (2016).
144. Pathak, S., Busemeyer, B., Rodrigues, J. N. B. & Wagner, L. K. Excited states in variational Monte Carlo using a penalty method. *J. Chem. Phys.* **154**, 034101 (2021).
145. Entwistle, M., Schätzle, Z., Erdman, P. A., Hermann, J. & Noé, F. Electronic excited states in deep variational Monte Carlo. *Nat. Commun.* **14**, 274 (2023).
146. Choo, K., Carleo, G., Regnault, N. & Neupert, T. Symmetries and many-body excitations with neural-network quantum states. *Phys. Rev. Lett.* **121**, 167204 (2018).
147. Cuzzocrea, A., Scemama, A., Briels, W. J., Moroni, S. & Filippi, C. Variational principles in quantum Monte Carlo: the troubled story of variance minimization. *J. Chem. Theory Comput.* **16**, 4203 (2020).
148. Jordan, P. & Wigner, E. über das Paulische Äquivalenzverbot. *Z. Phys.* **47**, 631 (1928).
149. Bravyi, S. & Kitaev, A. Fermionic quantum computation. *Ann. Phys.* **298**, 210–226 (2002).
150. Sorella, S. Green function Monte Carlo with stochastic reconfiguration. *Phys. Rev. Lett.* **80**, 4558–4561 (1998).
151. Choo, K., Mezzacapo, A. & Carleo, G. Fermionic neural-network states for ab-initio electronic structure. *Nat. Commun.* **11**, 2368 (2020).
152. Yang, P.-J., Sugiyama, M., Tsuda, K. & Yanai, T. Artificial neural networks applied as molecular wave function solvers. *J. Chem. Theory Comput.* **16**, 3513–3529 (2020).
153. Torlai, G., Mazzola, G., Carleo, G. & Mezzacapo, A. Precise measurement of quantum observables with neural-network estimators. *Phys. Rev. Res.* **2**, 022060 (2020).
154. Iouchtchenko, D., Gonthier, J. F., Perdomo-Ortiz, A. & Melko, R. G. Neural network enhanced measurement efficiency for molecular groundstates. *Mach. Learn. Sci. Technol.* **4**, 015016 (2023).
155. Glielmo, A., Rath, Y., Csányi, G., De Vita, A. & Booth, G. H. Gaussian process states: a data-driven representation of quantum many-body physics. *Phys. Rev. X* **10**, 041026 (2020).
156. Del Re, G., Ladik, J. & Biczó, G. Self-consistent-field tight-binding treatment of polymers. I. Infinite three-dimensional case. *Phys. Rev.* **155**, 997–1003 (1967).
157. Yoshioka, N., Mizukami, W. & Nori, F. Solving quasiparticle band spectra of real solids using neural-network quantum states. *Commun. Phys.* **4**, 1–8 (2021).
158. Barrett, T. D., Malyshev, A. & Lvovsky, A. I. Autoregressive neural-network wavefunctions for ab initio quantum chemistry. *Nat. Mach. Intell.* **4**, 351 (2022).
159. Zhao, T., Stokes, J. & Veerapaneni, S. Scalable neural quantum states architecture for quantum chemistry. *Mach. Learn. Sci. Technol.* **4**, 025034 (2023).
160. Giner, E., Scemama, A. & Caffarel, M. Using perturbatively selected configuration interaction in quantum Monte Carlo calculations. *Can. J. Chem.* **91**, 879–885 (2013).
161. Holmes, A. A., Tubman, N. M. & Umrigar, C. Heat-bath configuration interaction: an efficient selected configuration interaction algorithm inspired by heat-bath sampling. *J. Chem. Theory Comput.* **12**, 3674–3680 (2016).
162. Sharma, S., Holmes, A. A., Jeanmairet, G., Alavi, A. & Umrigar, C. J. Semistochastic heat-bath configuration interaction method: selected configuration interaction with semistochastic perturbation theory. *J. Chem. Theory Comput.* **13**, 1595–1604 (2017).
163. Greer, J. Monte Carlo configuration interaction. *J. Comput. Phys.* **146**, 181–202 (1998).
164. Coe, J. P. Machine learning configuration interaction. *J. Chem. Theory Comput.* **14**, 5739–5749 (2018).
165. Goings, J. J., Hu, H., Yang, C. & Li, X. Reinforcement learning configuration interaction. *J. Chem. Theory Comput.* **17**, 5482–5491 (2021).
166. Pineda Flores, S. D. Chembot: a machine learning approach to selective configuration interaction. *J. Chem. Theory Comput.* **17**, 4028 (2021).
167. Noojien, M., Shamasundar, K. & Mukherjee, D. Reflections on size-extensivity, size-consistency and generalized extensivity in many-body theory. *Mol. Phys.* **103**, 2277–2298 (2005).
168. Hutter, M. On representing (anti)symmetric functions. Preprint at <http://arxiv.org/abs/2007.15298> (2020).
169. Neuscamman, E. The Jastrow antisymmetric geminal power in Hilbert space: theory, benchmarking, and application to a novel transition state. *J. Chem. Phys.* **139**, 194105 (2013).
170. Sabzevari, I. & Sharma, S. Improved speed and scaling in orbital space variational Monte Carlo. *J. Chem. Theory Comput.* **14**, 6276–6286 (2018).
171. Rubenstein, B. Introduction to the variational monte carlo method in quantum chemistry and physics. In *Variational Methods in Molecular Modeling* 285–313 (Springer, 2017).
172. Toulouse, J., Assaraf, R. & Umrigar, C. J. Introduction to the variational and diffusion Monte Carlo methods. In *Advances in Quantum Chemistry* Vol. 73, 285–314 (Elsevier, 2016).
173. Amari, S. Natural gradient works efficiently in learning. *Neural Comput.* **10**, 251–276 (1998).
174. Ay, N., Jost, J., Lé, H. V. & Schwachhöfer, L. *Information Geometry*. No. 64 in a Series of Modern Surveys in Mathematics (Springer, 2017).
175. Spencer, J. S., Pfau, D., Botev, A. & Foukkes, W. M. C. Better, faster fermionic neural networks. Preprint at <http://arxiv.org/abs/2011.07125> (2020).

Acknowledgements

The authors acknowledge funding from the German Ministry for Education and Research (Berlin Institute for the Foundations of Learning and Data, BIFOLD), the Berlin Mathematics Research Center MATH+ (AA1-6 and AA2-8) and European Commission (ERC CoG 772230 ScaleCell). G.C. is supported by the Swiss National Science Foundation under Grant No. 200021_200336, and by the NCCR MARVEL, a National Centre of Competence in Research, under Grant No. 205602. The authors are grateful to N. Yoshioka for providing us with the raw data of ref. ¹⁵⁷. W.M.C.F. and his co-workers gratefully acknowledge PRACE for awarding them access to the JUWELS Booster supercomputer (<https://apps.fz-juelich.de/sc/hps/jewels/booster-overview.html>); the HPC RIVR Consortium (<https://www.hpc-rivr.si>) and EuroHPC JU for providing computing resources on the Vega HPC system at the Institute of Information Science (<https://www.izum.si>) and the UK Engineering and Physical Sciences Research Council for providing computing resources at the Baskerville Tier 2 HPC service (<https://www.baskerville.ac.uk>). Baskerville was funded by the EPSRC and UKRI through the World Class Labs scheme (EP/T02221/1) and the Digital Research Infrastructure programme (EP/W032244/1) and is operated by Advanced Research Computing at the University of Birmingham.

Author contributions

J.H., J.S. and K.C. contributed equally and all authors were part of the discussion, reviewing and editing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature Reviews Chemistry* thanks Pavlo Dral and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© Springer Nature Limited 2023

¹Microsoft Research AI4Science, Berlin, Germany. ²FU Berlin, Department of Mathematics and Computer Science, Berlin, Germany. ³DeepMind, London, UK. ⁴Department of Physics, University of Zurich, Zurich, Switzerland. ⁵IBM Quantum, IBM Research Zurich, Ruschlikon, Switzerland. ⁶IBM Quantum, Thomas J. Watson Research Center, New York, NY, USA. ⁷Imperial College London, Department of Physics, London, UK. ⁸EPFL, Institute of Physics, Lausanne, Switzerland. ⁹FU Berlin, Department of Physics, Berlin, Germany. ¹⁰Department of Chemistry, Rice University, Houston, TX, USA. ¹¹These authors contributed equally: Jan Hermann, James Spencer, Kenny Choo.