Ab-initio quantum chemistry with neural-network wavefunctions

Jan Hermann [1,2,†], James Spencer [3,†], Kenny Choo [4,5,†], Antonio Mezzacapo [6], W. M. C. Foulkes [7], David Pfau [3,7,*], Giuseppe Carleo [8,*], and Frank Noé [1,2,9,10,*]

[1] *Microsoft Research AI4Science, Karl-Liebknecht-Str. 32, 10178 Berlin, Germany*
[2] *FU Berlin, Department of Mathematics and Computer Science, Arnimallee 12, 14195 Berlin, Germany*
[3] *DeepMind, 6 Pancras Square, London N1C 4AG, United Kingdom*
[4] *University of Zurich, Department of Physics, Winterthurerstrasse 190, 8057 Zurich, Switzerland*
[5] *IBM Quantum, IBM Research Zurich, Saumerstrasse 4, 8803 Ruschlikon, Switzerland*
[6] *IBM Quantum, Thomas J. Watson Research Center, 1101 Kitchawan Rd, Yorktown Heights, New York 10598, USA*
[7] *Imperial College London, Department of Physics, South Kensington Campus, London SW7 2AZ, United Kingdom*
[8] *EPFL, Institute of Physics, CH-1015 Lausanne, Switzerland*
[9] *FU Berlin, Department of Physics, Arnimallee 12, 14195 Berlin, Germany*
[10] *Rice University, Department of Chemistry, 6100 Main St, Houston, TX 77005, United States*

Abstract Deep learning methods have outperformed human capabilities in many pattern recognition and data processing problems, in game playing, and now also play an increasingly important role in scientific discovery. A key application of machine learning in the molecular sciences is to learn potential energy surfaces or force fields from ab-initio solutions of the electronic Schrödinger equation using datasets obtained with density functional theory, coupled cluster, or other quantum chemistry methods. Here we review a recent and complementary approach—using machine learning to aid the direct solution of quantum chemistry problems from first principles. Specifically, we focus on quantum Monte Carlo (QMC) methods that use neural network ansatzes in order to solve the electronic Schrödinger equation, both in first and second quantization, computing ground and excited states, and generalizing over multiple nuclear configurations. While still at their infancy and far from maturity and routine usage, these new methods are already able to generate virtually exact solutions of the Schrödinger equation for small systems, and rival advanced conventional quantum chemistry methods for systems with up to a few dozen electrons.

1 Introduction

In the past decade, machine learning (ML) has made inroads into many areas of the physical sciences,[1] 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 structure,[9] either across conformational[10] or chemical space,[11] as well as with unsupervised learning for the generation of novel molecules.[12] These methods all require a pre-existing dataset of molecules and their properties for training, typically obtained with standard methods of QC such as density functional theory[13] or coupled cluster.[14] In these scenarios, ML accurately approximates a given method of QC at vastly increased computational efficiency. This approach has been already reviewed in other works cited above. In contrast, the present review focuses on the complementary use of ML as an ab-initio technique in QC, which requires no external data and instead 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, where 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 may be changing with the advent of principled and accurate ML-based empirical methods (we use "empirical" to refer to the use of any 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 a various degree also semi-empirical methods of QC fit into supervised learning, but using mostly relatively simple and physically motivated functional forms rather than the more general and highly flexible functions typical for ML. Vice versa, the many recent successful supervised ML models that predict energies or other molecular properties based on 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 framework 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

[†] These authors contributed equally
[*] Emails: franknoe@microsoft.com, giuseppe.carleo@epfl.ch, pfau@google.com

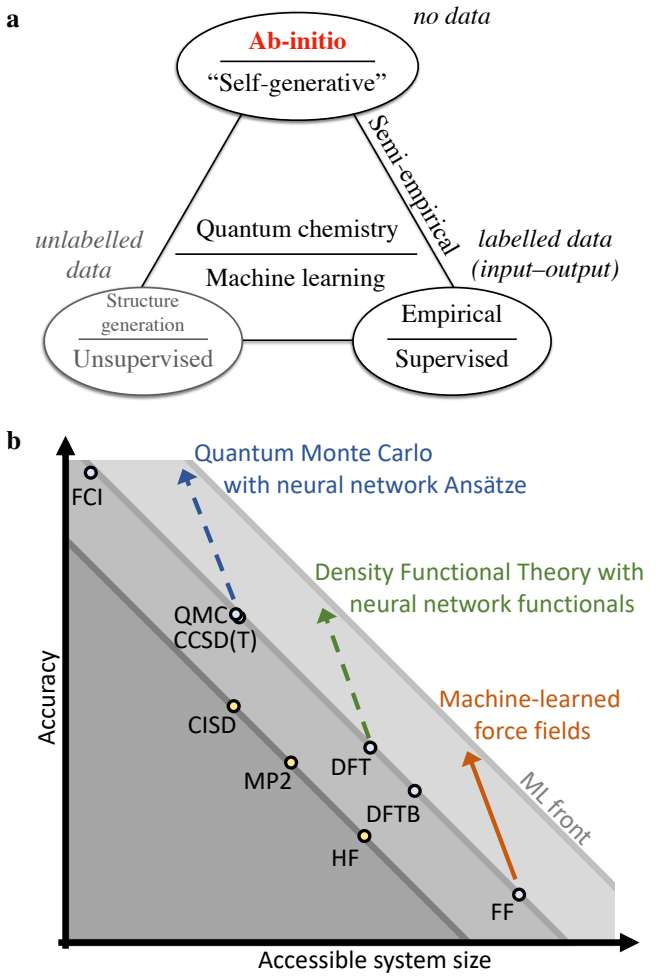


Fig. 1 | Quantum chemistry and machine learning. (**a**) Machine learning (ML) disciplines and their dependence on data can be mapped to disciplines in quantum chemistry (QC). This work reviews the use of machine learning in ab-initio QC, where the only input to ML is the Schrödinger equation itself. 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, though 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 of QC methods. New technologies can lead to better trade-offs, i.e. pushing the front (diagonal lines) to better accuracy and/or system sizes. Currently, 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 (blue, this review). Acronyms—FF: empirical force field; DFT: density functional theory; DFTB: density-functional-based tight binding; CCSD(T): coupled-cluster with singlets, doublets, and perturbative triplets; FCI: full configuration interaction; HF: Hartree–Fock; MP2: Møller–Plesset perturbation theory to second order; CISD: configuration interaction with singlets and doublets. For further discussion see Sec. 6.

to learn the underlying probability distribution that would generate a given dataset. Examples in chemistry include generative models for structural formulas[24] as well as full 3D structures of molecules,[4,25] and in physics the estimation of quantum states from measurements, known as quantum tomography.[26] 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.[27] 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 tokamaks.[31] 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.[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 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 a 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 previously published as a book chapter.[34,35] is a complementary review of neural network approaches for solving the electronic Schrödinger equation.

While one obvious contribution of the methods reviewed below 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 modeled (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 difficult cases, without reducing the accessible system sizes too much. And while none of the methods reviewed here are yet mature enough to routinely model systems with more than a few dozen electrons with guaranteed accuracy, the results obtained so far and reviewed below indicate that with further effort and progress, this aspiration might be achievable.

Section 2 briefly reviews the components of electronic structure theory necessary for the development of the ML methods discussed later on. The electronic structure problem is mapped to ML in Section 3, which is followed by a review of the ab-initio

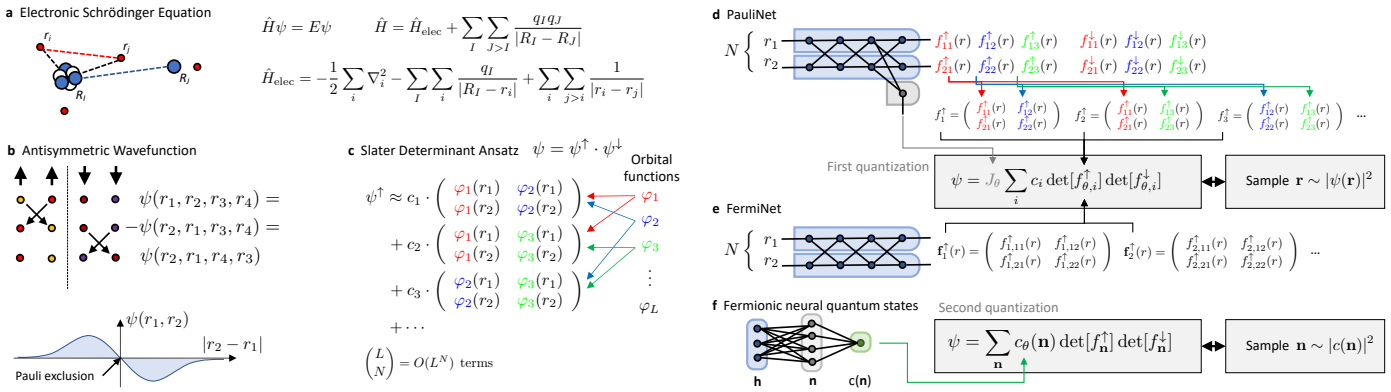


Fig. 2 | Electronic structure problem and its neural-network solutions. (**a**) The problem is fully specified by the geometry of a molecule and the electronic Schrödinger equation. (**b**) Only fully antisymmetric wavefunctions are admissible as solutions due to the Pauli exclusion principle and (**c**) these are often represented with Slater determinants. (**d**,**e**) Solutions formulated in first quantization use antisymmetric neural networks 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.

ML methods for QC formulated in real space and in a discrete basis in Sections 4 and 5, respectively. The review is concluded in Section 6.

2 Electronic structure

2.1 Schrödinger equation

QC aims at finding approximate solutions of the electronic Schrödinger equation that strike a good balance between accuracy and efficiency[36] (Fig. 1b). The nonrelativistic electronic Schrödinger equation within the Born–Oppenheimer approximation for a given molecule specified by the charges and coordinates of the nuclei, Z_I, $\mathbf{R}_I$, 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):

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

$$\hat{H} := \sum_i \left(-\tfrac{1}{2}\nabla^2_{\mathbf{r}_i} - \sum_I \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} \right) + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}. \tag{2}$$

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

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

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

$$E = \min_\psi \langle \hat{H} \rangle_\psi. \tag{4}$$

2.2 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 nonrelativistic Hamiltonian does not operate on spin, so the spin coordinate of each electron can be considered fixed. In this scenario [for full treatment, see 37, Sec. IV.E], the *spatial* wavefunction must be antisymmetric only with respect to the exchange of same-spin electrons, i.e., when $s_i = s_j$ (Fig. 2b),

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

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

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

When interpreting $\phi_j(\mathbf{r}_i)$ as the j-th component of a N-dimensional feature vector for the i-th electron (using ML parlance), $\boldsymbol{\phi}(\mathbf{r}_i)$, a Slater determinant is in fact the only antisymmetric function of N feature vectors that is linear in every one of them, making it a natural choice. Alternative antisymmetric forms exist, such as the Pfaffian[38] or the Vandermonde determinant and its generalizations,[39,40] but these are far less common and we will not discuss them here.

Slater determinants formed from different orbitals can be further mixed in a linear combination without breaking the antisymmetry (Fig. 2c). In fact, 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 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).

2.3 Variational wavefunction methods

An important class of QC methods derives directly from the variational principle (Eq. 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,

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

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

One can distinguish two strategies to construct the ansatzes. First, traditional QC uses relatively simple forms, such that the integral of Eq. (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 stochastically.[42] The latter is a natural framework to incorporate neural networks, and we introduce it in more detail in Section 3.1. Here we introduce three ansatzes for electronic wavefunctions of the first (traditional) kind, since they serve as scaffolding for the neural-network ansatzes of Sections 4 and 5.

Hartree–Fock Perhaps the simplest nontrivial ansatz in QC is the single Slater determinant of Eq. (6), where the orbitals $\phi_j(\mathbf{r})$ 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, $\varphi_k(\mathbf{r})$, $k = 1, \ldots, K$, with $K \sim N$ in most cases:

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

The use of a finite basis set turns the functional optimization problem of Eq. (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.

Density functional theory (DFT) is quite a different approach and relies instead on an in-principle exact mapping of the ab-initio Hamiltonian (Eq. 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),

$$\psi(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \sum_p c_p D_{\boldsymbol{\phi}_p}(\mathbf{r}_1, \ldots, \mathbf{r}_N). \tag{9}$$

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 (see Box 1). The simplest version of CI, called full CI (FCI), considers all $\binom{M}{N}$ 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 theory, [14]; matrix product states, [44]), deterministic pruning (selected CI, [45]), or stochastic sampling (FCI-QMC, [46]). Section 5 explores a novel way of "compressing" the CI expansion through neural networks.

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 $D_{\boldsymbol{\phi}_p}(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ 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 (Eqs. 8 and 9) is harder to optimize but much more expressive. The most common variant is to consider all $\binom{M'}{N'}$ Slater determinants formed by letting $N' < N$ electrons occupy a space of $M' < M$ orbitals, while the remaining $N - N'$ electrons occupy a fixed set of inactive orbitals. This is called the *complete active space self-consistent field* (CASSCF) method.[47] Due 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 (Eq. 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 CASSCF 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 Section 4 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.

3 Machine learning for electronic Schrödinger equation

3.1 Mapping quantum mechanics to machine learning

A 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 (Section 3.2) represents a wavefunction, which accepts electron coordinates (first quantization)

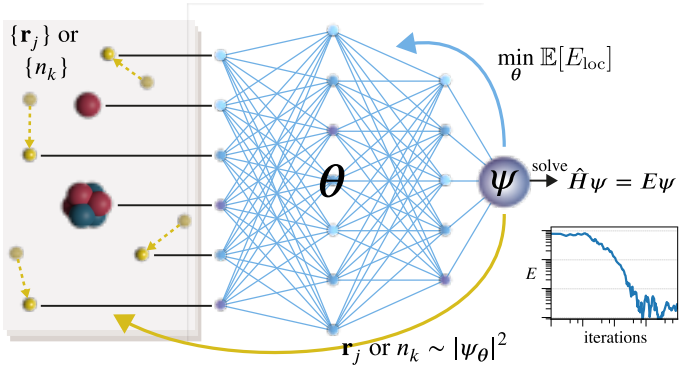


Fig. 3 | Variational Monte Carlo with neural networks. Electron positions, $\mathbf{r}_j$, or orbital occupation numbers, n_k, describe an electron configuration (Box 1) which is an input to the wavefunction, ψ, represented by a neural network parametrized with $\boldsymbol{\theta}$. The wavefunction is used in two ways: first, to sample new electron configurations which provide new input to the neural network (yellow, Box 2), and second, to evaluate the electronic energy, which is minimized by varying the network parameters (blue, Box 3).

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 2). The optimization methods used (Box 3) 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 & Troyer[52] 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.[53–56] In essence, this review is concerned with the extension of this approach to electrons in molecules.

3.2 Deep learning

The standard practice in ab-initio QC today is in some ways analogous to the state of computer vision before the rise of deep learning. Prior to 2012, the best pipelines for large-scale image recognition consisted of a combination of hand-designed features and simple ML models.[61] A single deep convolutional neural network trained end-to-end was able to cut the recognition error in half relative to these systems,[62] 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) 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 significantly 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 Schmidhuber,[63] and for a review of the fundamental concepts in deep learning, see LeCun et al..[64]

Neural networks date back to the very beginning of computer science,[65] and their modern form originates with the single perceptron "unit",[66] which produces as output a non-linear function of the sum of a constant, known as the bias, and a linear combination of its inputs. The non-linear 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 to the next, the resulting multi-layer perceptron (MLP) can, in theory, represent any smooth function to arbitrary accuracy given enough units.[67] 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,[68–70] was found to be effective for training neural networks.[71] 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.

Several factors were instrumental in rehabilitating neural networks under the banner of "deep learning": a combination of algorithmic advances[72] and the use of modern GPU hardware[73] made the computations much faster, and the resulting ability to train larger networks made issues with local minima less severe.[74,75] Furthermore, with the help of stochastic gradient descent, deep neural networks can be applied straightforwardly and efficiently to large datasets, unlike other ML models.[76,77] Finally, empirical successes like winning the ImageNet Large Scale Visual Recognition Challenge[78] 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, thanks to a mature ecosystem of software libraries for numerical computing with automatic differentiation and hardware accelerators.[79–81] 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 outputs.[82] However, linear models also have this issue, and these issues are of limited concern in the applications discussed here, where only average performance matters.

3.3 Neural network architectures

The starting point for most neural networks is the multi-layer perceptron (MLP), formed as a composition of L layers,

$$\begin{aligned} \mathrm{MLP}(\mathbf{x}) &= f^L \circ f^{L-1} \circ \cdots \circ f^1(\mathbf{x}), \\ f^\ell(\mathbf{z}) &= f\left(\mathbf{W}^\ell \mathbf{z} + \mathbf{b}^\ell\right), \end{aligned} \tag{11}$$

where $\circ$ denotes function composition, f is some non-linear *activation* function, and $\mathbf{W}^\ell$ and $\mathbf{b}^\ell$ are the matrices of weights

and vectors of biases to learn. While a vanilla 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 should transform in a corresponding way. For instance, the weights in a layer of a convolutional neural network (ConvNet)[83] are restricted to be a discrete convolution operator, which constrains each layer to be translation-equivariant, a natural constraint for image recognition, and dramatically reduces the number of possible weights in a layer.

Equivariance to permutation is another frequently useful property that is especially important in real-space approaches to representing electronic wavefunctions (see Section 4). A simple permutation-equivariant layer first proposed by Shawe-Taylor[84] can be constructed by applying the same transformation to each input and summing the results. More sophisticated permutation-equivariant layers are used by models like the Transformer[85] or SchNet.[86] 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 to be equivariant to leads naturally to recipes for constructing the appropriate neural network layers.[87]

Another class of neural network architectures, which have been influential as wavefunction ansatzes, are restricted Boltzmann machines (RBMs).[88] These were originally developed for unsupervised learning, but in the VMC setting considered here they lead to a simple 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 autoencoders,[89] generative adversarial networks,[90] normalizing flows,[91] autoregressive models,[92,93] and diffusion models.[94] In fact, some of these newer models have started to have an impact as neural network wavefunction ansatzes for spin systems. Examples are deep autoregressive quantum states,[95] convolutional neural networks,[96] recurrent neural networks,[97] and normalizing flows.[98]

3.4 Neural network architectures for molecules

In the past 15 years, there has been very active research on designing ML architectures for molecular physics and chemistry tasks that are expressive and flexible, yet incorporate physical concepts and constraints. Invariance, e.g. of energy functions with respect to translation and rotation of the molecular coordinates, and with respect to exchange of identical particles, has already been incorporated into early architectures by designing features with the respective invariances.[99–101] Message-passing architectures such as SchNet have leveraged the increased expressiveness of deep learning by defining graph convolutions parametrized by roto-translationally invariant features.[86,102] Many recent efforts went into the design of rotation-equivariant neural networks which have vectorial or tensorial features rotating with the inputs.[103–108] It has been shown that such neural network architectures can reach high accuracy not only in the large-data limit, but also for small datasets which could previously only be tackled by kernel approaches.[107,108] Energy functions based on many-body expansions were considered in the form of using permutationally invariant polynomials,[109] atomic cluster expansion,[110] and permutationally invariant polynomials for molecules.[111]

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

4 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, $\psi(\mathfrak{r};\boldsymbol{\theta})$. Here $\mathfrak{r}$ stands for the N-tuple of electron coordinates, $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N$, and sampling is realized over electronic positions $\mathfrak{r}$ (Box 2). The antisymmetry constraint (Eq. 5) must be imposed in ψ to avoid collapsing onto a lower-energy bosonic state. A commonly adopted form is $\psi(\mathfrak{r};\boldsymbol{\theta}) = S(\mathfrak{r};\boldsymbol{\theta})\times A(\mathfrak{r};\boldsymbol{\theta})$, where 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 (Eq. 6). As discussed in Section 2, single Slater determinants with fixed orbitals have limited expressiveness, and many such determinants need to be combined to achieve high accuracy.[115] A natural generalization of a sum of fixed-orbital Slater determinants is the commonly-used Slater–Jastrow wavefunction

$$\psi(\mathfrak{r};\boldsymbol{\theta}) = \mathrm{e}^{J(\{\mathfrak{r}\};\boldsymbol{\theta})} \sum_k c_k \begin{vmatrix} \phi_1^k(\mathbf{r}_1;\boldsymbol{\theta}) & \cdots & \phi_1^k(\mathbf{r}_N;\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \phi_N^k(\mathbf{r}_1;\boldsymbol{\theta}) & \cdots & \phi_N^k(\mathbf{r}_N;\boldsymbol{\theta}) \end{vmatrix} \quad (12)$$

where the Jastrow factor, $J(\{\mathfrak{r}\};\boldsymbol{\theta})$ constitutes the symmetric ("bosonic") part of the state and typically contains one- and two-body (and in many cases higher-order) parameterized correlations. The set notation, $\{\mathfrak{r}\} \equiv \{\mathbf{r}_1, \ldots, \mathbf{r}_2\}$, indicates that J does not depend on the order of the electron coordinates. The determinants in Eq. (12) are typically replaced with the product of spin-up and spin-down determinants.[37] Separating the up- 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.

4.1 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

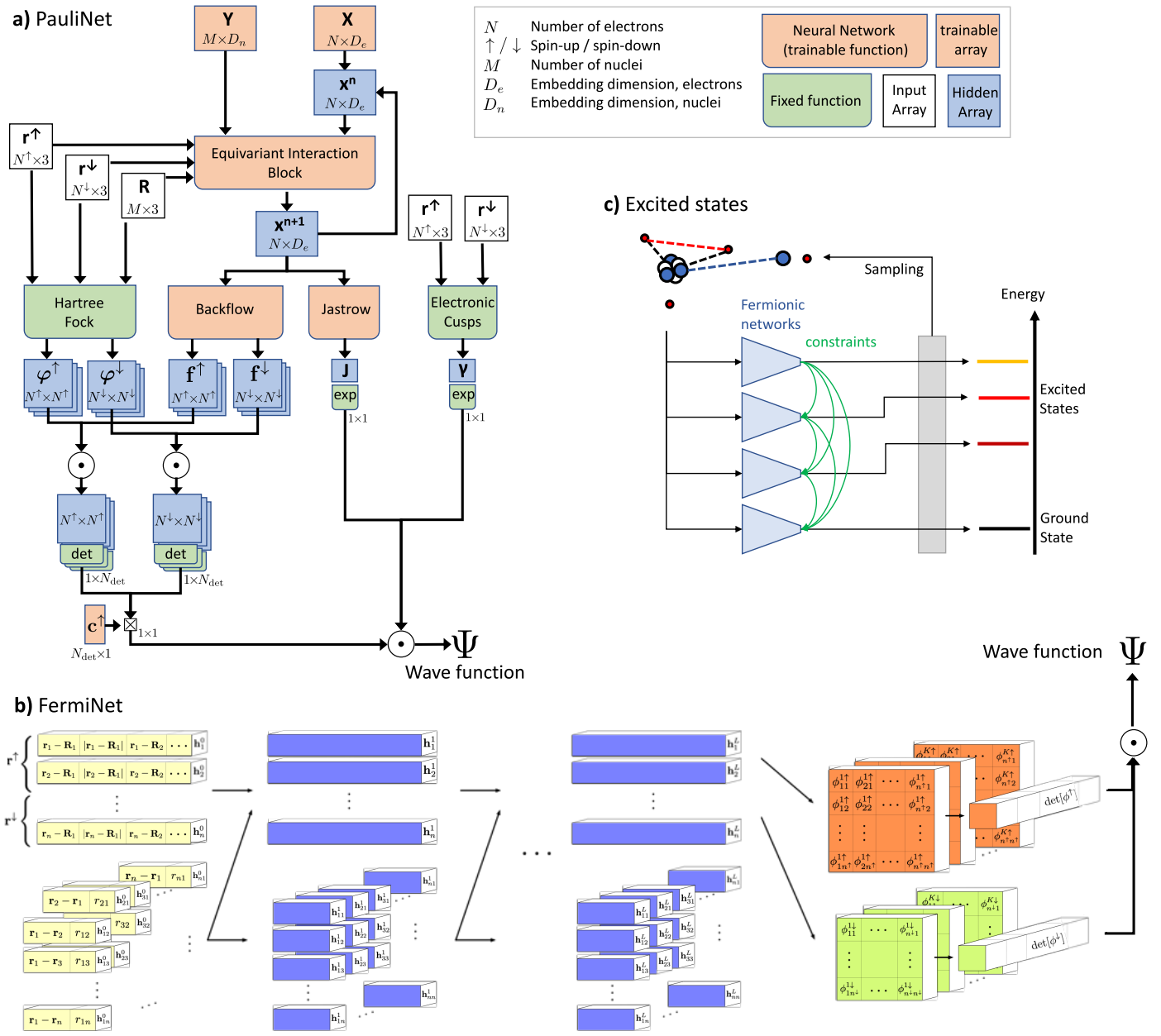


Fig. 4 | Neural-network architectures for selected real-space wavefunctions. (**a**) Original PauliNet architecture from.[112] (**b**) Original FermiNet architecture from.[113] Both architectures have been modified and extended by various contributions mentioned in this review. (**c**) Approach to computing exciting states in.[114]

$S(\mathbf{r};\theta)$ can be readily parameterized with a strategy closely related to NQS for spins:

$$S(\mathbf{r};\theta) = g(n(\mathbf{r});\theta), \tag{13}$$

where $n(\mathbf{r})$ is the unique occupation number representation corresponding to the electronic positions $\mathbf{r}$ and g represent a generic function which could be represented by a neural network. Since 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 NN architectures also adopted for spin systems[52] or lattice bosons[53] can be used to represent the symmetric part S. Early works on the Hubbard model adopted positive-definite RBM-based parameterizations of $S(\mathbf{r};\theta)$,[54] while more recent works have adopted deep-network parameterizations allowing for sign changes.[116]

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

$$A(\mathbf{r};\theta) = \begin{vmatrix} \phi_1(\mathbf{r}_1;\theta) & \cdots & \phi_1(\mathbf{r}_N;\theta) \\ \vdots & \ddots & \vdots \\ \phi_N(\mathbf{r}_1;\theta) & \cdots & \phi_N(\mathbf{r}_N;\theta) \end{vmatrix}, \tag{14}$$

where 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 since it effectively fixes the anti-symmetric part to a mean-field reference solution.

Neural backflow A significant improvement is obtained by considering a many-body backflow transformation of the orbitals.[117,118] 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:

$$\tilde{\Phi}_{ij}(\boldsymbol{\theta}) = \Phi_{ij}(\boldsymbol{\theta}) + \Delta_{ij}(n(\mathfrak{r}); \boldsymbol{\theta}), \tag{15}$$

where Δ is a correction to the single-particle orbitals Φ. In physics-inspired parameterizations, Δ is typically taken to be a simple function of the electronic occupation numbers.[119] The neural backflow method[120] instead introduced a flexible parameterization of the backflow orbitals based on artificial neural networks. In this case, Δ is parameterized with a 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 wave function. The constrained hidden fermion formalism builds on the idea of introducing a set of $\tilde{N}$ auxiliary fermionic particles, with positions $\mathfrak{q}$, and living on $\tilde{M}$ lattice sites. These auxiliary particles are used to effectively mediate correlations among the physical degrees of freedom.[121] Calling $\tilde{A}(\mathfrak{r}, \mathfrak{q}; \boldsymbol{\theta})$ a Slater determinant for the extended (physical+hidden) system, the resulting antisymmetric form for the physical system is given by

$$A(\mathfrak{r}; \boldsymbol{\theta}) = \tilde{A}(\mathfrak{r}, F(\mathfrak{r}; \boldsymbol{\theta})). \tag{16}$$

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 model.[121]

4.2 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 may be improved in a matter suitable for use with neural quantum states by adding a backflow transformation, in which the one-electron orbitals $\phi_i(\mathbf{r}_j; \boldsymbol{\theta})$ are replaced by many-electron functions $\tilde{\phi}_i(\mathbf{r}_j, \{\mathfrak{r}\}; \boldsymbol{\theta})$. The backflow transformation can either modify the orbitals directly via a multiplicative and/or additive term:

$$\tilde{\phi}_i(\mathbf{r}_j, \{\mathfrak{r}\}; \boldsymbol{\theta}) = \phi_i(\mathbf{r}_j) f_i^{\otimes}(\mathbf{r}_j, \{\mathfrak{r}\}; \boldsymbol{\theta}) + f_i^{\oplus}(\mathbf{r}_j, \{\mathfrak{r}\}; \boldsymbol{\theta}), \tag{17}$$

or act as a quasiparticle transformation of the electron coordinates:

$$\tilde{\phi}_i(\mathbf{r}_j, \{\mathfrak{r}\}; \boldsymbol{\theta}) = \phi_i\big(\mathbf{r}_j + \boldsymbol{\xi}(\{\mathfrak{r}\}; \boldsymbol{\theta})\big), \tag{18}$$

where the parameterized functions, $f_i^{\otimes}, f_i^{\oplus}, \boldsymbol{\xi}$, are invariant to permutations of $\{\mathfrak{r}\}$, and $\boldsymbol{\xi}(\{\mathfrak{r}\}; \boldsymbol{\theta})$ is a three-component vector

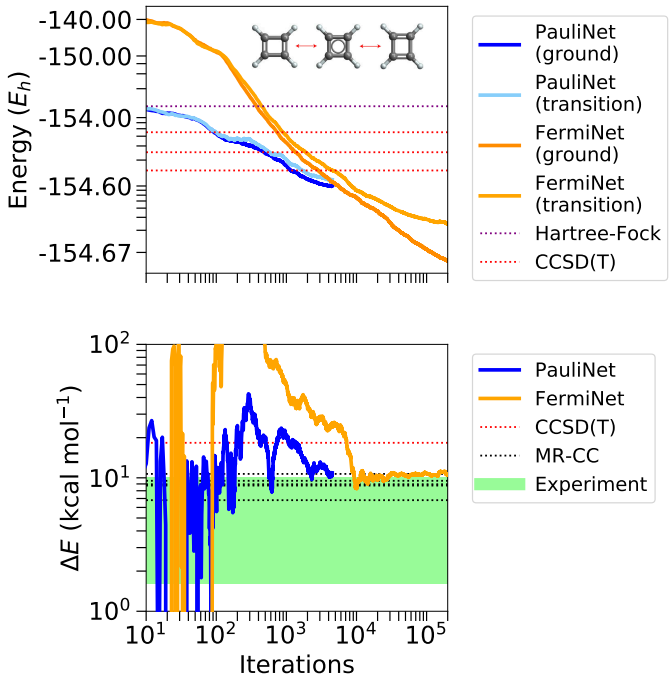


Fig. 5 | Automerization of cyclobutadiene with neural-network ansatzes. Top: PauliNet converges more quickly, while FermiNet reaches lower total energy. Hartree–Fock and CCSD(T) absolute energies (three red lines correspond to cc-pVnZ basis sets, $n = \mathrm{D}, \mathrm{T}, \mathrm{Q}$) for the equilibrium 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 flavours of the CC theory). Figure modified from Spencer et al..[124]

that modifies $\mathbf{r}_j$. If we consider a determinant of orbitals of this form,

$$\begin{vmatrix} \phi_1(\mathbf{r}_1; \{\mathfrak{r}\}) & \cdots & \phi_1(\mathbf{r}_N; \{\mathfrak{r}\}) \\ \vdots & \ddots & \vdots \\ \phi_N(\mathbf{r}_1; \{\mathfrak{r}\}) & \cdots & \phi_N(\mathbf{r}_N; \{\mathfrak{r}\}) \end{vmatrix}, \tag{19}$$

then we see that orbitals with backflow transformations are just one example of a broader class of functions. In order for the determinant to be antisymmetric, the matrix with elements $\Phi_{ij} = \phi_i(\mathbf{r}_j; \{\mathfrak{r}\})$ must be permutation-equivariant; that is, exchanging electrons k and l also exchanges columns k and l. While conventional Slater–Jastrow–backflow (SJB) wavefunctions have had considerable success, they also have limitations due 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 fermion-inc systems, such as those encountered in nuclear physics. Real-space VMC with neural-network ansatzes has been successfully used to model those as well.[122,123]

Iterative backflow Taddei et al.[125] introduced a form of backflow that applied Eq. (18) repeatedly in an interative fashion. Such an ansatz is formally equivalent to expressing the backflow as a deep neural network,[126] albeit with artificial restriction on the dimensionality of the hidden layers. The iterative backflow was used for studying the ^{3}He and ^{4}He liquids.

DeepWF The DeepWF[39] approach uses an ansatz similar to a Slater–Jastrow wavefunction but with a simpler antisymmetric term:

$$\psi(\mathbf{r}) = S(\{\mathbf{r}\}, R)A^{\uparrow}(\mathbf{r}^{\uparrow})A^{\downarrow}(\mathbf{r}^{\downarrow}). \tag{20}$$

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, $A^{\sigma} = \prod_{1\leq i\leq j\leq N}\left(a(\mathbf{r}_i, \mathbf{r}_j, r_{ij}) - a(\mathbf{r}_j, \mathbf{r}_i, r_{ij})\right)$. The two-body antisymmetric function is entirely learned. Such a functional form can be evaluated in $\mathcal{O}(N^2)$ operations, compared to $\mathcal{O}(N^3)$ for 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 below do much better: vanilla PauliNet obtained 99.94% and 97.3% of the correlation energies for the beryllium and boron atoms, and FermiNet 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 exist.[127]

PauliNet PauliNet[112] builds upon HF or CASSCF orbitals as a physically meaningful baseline and takes a neural network approach to the SJB wavefunction in order 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 wavefunction.[128] A graph-convolutional block based on SchNet[86] 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.[112] introduced PauliNet with a purely multiplicative backflow as shown in Fig. 4a; Schätzle et al.[129] generalized this to a multiplicative and additive backflow as shown in Eq. (17). PauliNet is optimized with a fixed number of Slater determinants. Most of the results reported in Hermann et al., Schätzle et al.[112,129] were obtained with around 10 determinants.

FermiNet FermiNet[113] takes a more minimalist (or machine-learning maximalist) approach and attempts to train a neural network to represent the entire wavefunction (Fig. 4b). FermiNet uses two parallel networks, describing one- 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- 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.,[113] 16 determinants were used. FermiNet builds up a rich description of electron-electron interactions from the permutation-equivariant mixing of information describing one- and two-electron features. In particular, the electron-nuclear and electron-electron cusps in the wavefunction are represented accurately, despite not being encoded explicitly. Whereas PauliNet is usually trained with the first-order ADAM optimizer,[130] FermiNet training was found to be substantially improved when employing the KFAC optimizer (Martens & Grosse,[60] Box 3).

While both PauliNet and FermiNet exceed the accuracy of conventional SJB wavefunctions on small systems, there are important tradeoffs 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 While 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.[131] 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 like benzene and the potassium atom, and in fewer iterations. Glehn et al.[132] introduced the Psiformer, which marks a more dramatic departure from prior work, by replacing the neural network part of the FermiNet with a sequence of self-attention layers similar to the Transformer.[85] The Psiformer was found to reach even lower energies than Gerard et al.[131] on benzene, and on larger systems like 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.,[98] though for estimation of the density matrix rather than ground state.

Potential energy surfaces Typically one optimises 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. Scherbela et al.[133] developed a training methodology that allows weight sharing between (simplified) PauliNet architectures targeting different geometries. By switching the geometry being trained at each epoch, 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 initialised from a wavefunction for a smaller molecule and could then be fine-tuned in a relatively short optimization stage. Pretraining neural network wavefunctions from smaller systems has also been shown to dramatically accelerate convergence for Kagome lattice models.[134]

Similarly, Gao & Günnemann, Gao & Günnemann[135,136] demonstrated a hypernetwork approach (a network that predicts weights of another network), where 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 model to fully 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 the wavefunction form is sufficiently flexible.

Potential energy surfaces can also be estimated locally by computing forces rather than energies. While the Hellmann-Feynman theorem[137] 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.[138,139] These techniques have been successfully applied to the computation of forces using the FermiNet as a ground state ansatz.[140]

Periodic systems There has also been progress on using first-quantized neural network architectures in periodic systems, such as interacting quantum gases in low dimension,[141] the electron gas,[142–144] and for small cells of solids such as lithium hydride and graphene.[144] Again, sufficiently expressive networks at the VMC level have been found capable of rivaling or surpassing the accuracy of fixed-node diffusion Monte Carlo calculations using conventional Slater-Jastrow-backflow trial wavefunctions.

Computational cost Any comparison of the computational cost of the reviewed methods against traditional quantum chemistry will depend largely on the target property, target accuracy, hardware, and system size. On one hand, the scaling of the reviewed methods with number of electrons N are favourable, on the other hand the prefactors are large because neural networks are much more expensive to evaluate than traditional ansatz functions in quantum Chemistry. For the system sizes considered in this review, and ordinary properties such as binding energies, traditional methods such as CCSD(T) 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 GPU-days of Paulinet on cyclobutadiene is comparable to the many dozens of CPU-days of multireference CC methods on the same system.[112] 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.

4.3 Extensions

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 variational Monte Carlo methods, which scale roughly as $\mathcal{O}(Z^5)$,[145] where Z is the nuclear charge. Whilst the core electrons contribute heavily to the total energy, energy differences are largely determined by the behavior of the valence electrons. The core electrons can therefore be removed and the effective nuclear charge reduced by using pseudopotentials. The use of pseudopotentials is common in many methods, including density functional theory and conventional variational Monte Carlo. Li et al.[146] demonstrate that effective core potentials can be readily combined with FermiNet and achieve accuracy comparable to CCSDT(Q) extrapolated to the complete basis set limit 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 (DMC) Projector methods such as DMC[147] and auxiliary-field Monte Carlo[148] 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 in principle, exact but, for many-fermion systems, relies in practice on the fixed-node approximation, in which 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 Slater-Jastrow-backflow trial functions optimized using VMC methods are substantial, explaining why DMC is so often used to provide improved estimates of the ground-state wavefunction and energy. Wilson et al.[149] 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.[149] used a simplified FermiNet that gave VMC energies higher than those reported by Pfau et al.,[113] 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 limit,[129] 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 optimisation. If this is the case, diffusion Monte Carlo with the fixed node approximation may not produce substantially lower energies. On the other hand, since neural network wavefunctions routinely capture over 90% of the correlation energy at the VMC level, the need to perform expensive diffusion Monte Carlo calculations is greatly reduced. More recently, Ren et al.[150] showed that DMC can capture roughly half of the remaining correlation energy for the atoms Li-Ar, when using a very small FermiNet-based architecture. Whilst it is possible to achieve energies within chemical accuracy using FermiNet at the VMC level, these calculations model the case for larger systems where converging the energy with respect to network size might not be feasible. Ren et al.[150] 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 behavior of materials. Fortunately, recent 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,[151,152] in which the average energy over multiple states is minimised and individual states are projected out via diagonalization within the basis

of excited states. Similar techniques are used with other quantum chemistry methods. Zhao & Neuscamman[153] instead minimized a different objective function, such that the state with energy closest to a desired energy target is obtained. Pathak et al.[154] suggested a simple alternative, where 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, in particular, penalty function approaches have recently been explored. As with ground-state calculations, the flexibility of the wavefunction ansatz to represent the desired state is critical. Entwistle et al.[114] 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.[155] 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 question.[156]

5 Electrons in second quantization

Instead of working directly with the infinite-dimensional Hilbert space corresponding to the real-space Hamiltonian of Eq. (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-quantised 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 wave function with respect to the exchange of electrons.

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

$$\hat{H} = \sum_{ij} t_{ij}\, \hat{c}_i^\dagger \hat{c}_j + \sum_{ijkl} u_{ijkl}\, \hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_l\, \hat{c}_k, \tag{21}$$

where

$$t_{ij} = \int \varphi_i^*(\mathbf{r}) \left(-\frac{1}{2}\nabla^2 - \sum_I \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|} \right) \varphi_j(\mathbf{r})\, d\mathbf{r}, \tag{22}$$

$$u_{ijkl} = \iint \varphi_i^*(\mathbf{r}) \varphi_j^*(\mathbf{r}') \frac{1}{|\mathbf{r} - \mathbf{r}'|} \varphi_k(\mathbf{r}) \varphi_l(\mathbf{r}')\, d\mathbf{r}\, d\mathbf{r}', \tag{23}$$

are matrix elements of the one- and two-electron terms in the real-space Hamiltonian of Eq. (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 $\psi(n_1, n_2 \cdots)$, thus encoding the amplitudes for different occupations of the orbitals (Box 1). 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 quantum chemistry corresponds to the Jordan–Wigner mapping,[159] which transforms annihilation and creation operators into, respectively, lowering and raising spin operators $\hat{\sigma}_j^\pm = (\hat{\sigma}_j^x \pm i\hat{\sigma}_j^y)/2$. However, this mapping is not unique, and more recent alternatives exist, such as parity or Bravyi–Kitaev encodings,[160] both of which have been developed in the context of quantum simulations.

Overall, this occupation-number formalism and the corresponding Hamiltonian (21) serve as the starting point for the methods described in this section.

5.1 Fermionic neural quantum states

The many-body amplitudes in the occupation number representation can be readily expressed in terms of neural networks taking discrete inputs. Since the antisymmetry constraint is fully encoded in the Hamiltonian (21), 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 based on complex-valued RBMs,[52] 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

$$\psi(n_1, n_2 \dots n_L; \boldsymbol{\theta}) = e^{\sum_i^L a_i n_i} \prod_{j=1}^{M} 2\cosh\left(b_j + \sum_i^L W_{ij} n_i \right), \tag{24}$$

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

This ansatz can be optimized with VMC techniques (Box 3), in which the occupation numbers are sampled according to the probability density proportional to $|\psi(n_1, n_2, \dots)|^2$, and 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 (21) in the occupation basis. Optimization of the wave function ansatz typically relies on the stochastic reconfiguration[57] approach. A number of works have adopted this simple NQS wave function and achieved competitive variational results for relatively small basis sets,[157,161] even in conjunction with quantum computers.[162,163] In Fig. 6 (a), we show the dissociation curve of C_2, in the STO-3G basis, using the RBM as described above.[157]

An alternative ML-based variational ansatz for second-quantized Hamiltonians based on Gaussian process regression has been also proposed,[164] which has reached accuracy comparable to the NQS approach on model Hamiltonians, but has not yet been applied to the ab-initio quantum chemistry Hamiltonian.

Solids The second-quantization framework also allows one to treat solids, using as a basis the Bloch orbitals obtained by solving the crystalline HF equations.[165] Creation and annihilation operators, $\hat{c}_{i\mathbf{k}}^\dagger$ and $\hat{c}_{i\mathbf{k}}$, for electrons in the band i with crystal momentum k are introduced, and the resulting Hamiltonian is similar to Eq. (21), with the noticeable difference that the one- and two-body matrix elements now depend on the crystal momenta: $t_{ij} \to t_{ij}^{\mathbf{k}}$ and $u_{ijkl} \to u_{ijkl}^{\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3\mathbf{k}_4}$, with the four momenta 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 state,[158] applied this approach to study the electronic structure of solids. In Fig. 6 (b), we show the computed ground-state energies for graphene crystals as a function of the lattice constant.

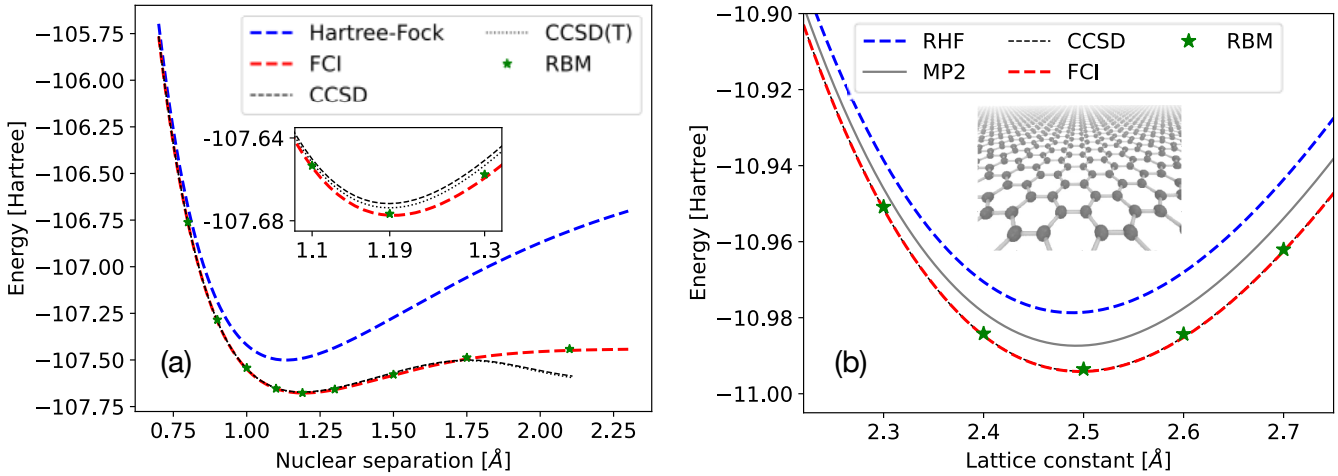


Fig. 6 | Electronic energies for molecules and solids in second quantization (**a**) Dissociation curve for N_2 molecule in the STO-3G basis. The green stars show results for a restricted Boltzmann machine which represents the electrons in discrete space. Figure adapted from.[157] (**b**) Graphene on a honeycomb lattice solved using the cc-pVDZ basis set. Figure adapted from.[158]

Exact Sampling Fermionic NQS are typically sampled using the MCMC approach commonly adopted in VMC (Box 2). However, the mixing rate of the MCMC algorithm is known to be slow in some cases, such as close-to-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 wavefunctions,[95] 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 multi-modal. The exact sampling approach was applied to QC hamiltonians in recent work by Barrett et al..[166] 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 the early applications of Choo et al..[157] Specifically, Zhao et al.[167] 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 (Na_2CO_3 at equilibrium) have been obtained at a relatively modest computational cost.

5.2 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 diagonalising the Hamiltonian on the reduced subspace, one can achieve highly accurate results. This set of approaches is also known as selected CI.[45,168–170] Different flavours of selected CI vary in the way relevant configurations are selected.

One well-known approach is called Monte Carlo CI (MCCI)[171] and can be briefly summarised as follows:

1. Start from a finite set of configurations S_i
2. By considering single or double excitations starting from configurations in S_i, construct an expanded set S_i'.
3. Construct the Hamiltonian $\hat{H}_i$ for the expanded set S_i' and diagonalise to obtain the wavefunction coefficients for the configurations in the set.
4. 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}.
5. Repeat until convergence.

ML techniques can be used to improve selection of the configuration set. One such approach is to perform supervised learning,[172] where a neural network is trained to predict the wavefunction coefficients using the data from the MCCI method, i.e., 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 Coe[172] 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 where 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 minimising the variational energy. This approach was applied in Goings et al.[173] 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 or not, which is the setup used by Chembot [174].

6 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 may become the methodology with the best trade-off between efficiency and accuracy for systems with up to one to

two hundred electrons and a nontrivial electronic structure, which are already too large for exact methods, and for which conventional high-accuracy methods may 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 so far 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 extensivity [175], and these have not yet been established for any of the methods reviewed above. 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, and it is yet to be seen to what degree (perhaps fully?) can the much increased expressiveness of the individual determinants parameterized with neural networks mitigate this issue. One could also imagine scaling of the network size with system size, but how precisely that would need to be done is likewise an open question. Certain degree of size-inconsistency does not necessarily render a method unusable, but it certainly makes its reliable application more difficult. If the current ansatzes suffer from size-inconsistency, any modifications such that the network size or number of determinants would need to grow with system size would introduce additional scaling of computational cost, or, if left untreated, may result in loss of accuracy with growing system size. Whether this is the case or not 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. Since 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, while 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 orbital,[176] 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. While the formal quartic scaling of backflow-based variational QMC with system size is favourable compared to traditional QC methods, the prefactor due to the neural networks is large. Until very recently, 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 reported.[144] The prefactor can be reduced by integrating traditional QC techniques such as pseudopotentials,[146] developing more efficient neural-network architectures, or using ML techniques such as pre-training and transfer learning. Specific to the discrete-basis second-quantized approaches is the issue of basis-set convergence, where sufficiently large basis sets may increase the prefactor by up to three orders of magnitude compared to 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 be often 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 and can be addressed quickly, thanks to the relative simplicity of the framework based on variational QMC compared to 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 that can be reached with variational QMC with neural-network wavefunctions on small systems makes us optimistic that these technologies can make progress on quantum systems that were as yet 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 above have already reached thermochemical (sub-mH) accuracy in ionization potentials and electron affinities, and rival or in some cases even exceed the accuracy of advanced conventional quantum chemistry methods. Yet these networks are just a small subset of possible architectures for representing antisymmetric wavefunctions, and it is unlikely that the optimal ones were found on the first attempt, so we expect that significant innovation in the pursuit for better efficiency/accuracy trade-off lies ahead.

An important aspect of variational QMC is that be solutions can systematically ranked 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 may 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 are available. 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 enables to generalize across different nuclear geometries, i.e. the variational learning over entire potential energy surfaces with one model, as well as to generalize across chemical space. While 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 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.

References

[1] Carleo, Cirac, Cranmer, Daudet, Schuld, Tishby, Vogt-Maranto & Zdeborová. "Machine Learning and the Physical Sciences." *Rev. Mod. Phys.* **91**, 045002 (2019). DOI: 10/ggd5qv.

[2] Jumper, Evans, Pritzel, Green, Figurnov, Ronneberger, Tunyasuvunakool, Bates, Žídek, Potapenko, et al. "Highly Accurate Protein Structure Prediction with AlphaFold." *Nature* **596**, 583 (2021). DOI: 10/gk7nfp.

[3] Deringer, Bernstein, Csányi, Ben Mahmoud, Ceriotti, Wilson, Drabold & Elliott. "Origins of Structural and Electronic Transitions in Disordered Silicon." *Nature* **589**, 59 (2021). DOI: 10/ghsb84.

[4] Noé, Olsson, Köhler & Wu. "Boltzmann Generators: Sampling Equilibrium States of Many-Body Systems with Deep Learning." *Science* **365**, eaaw1147 (2019). DOI: 10/gf7qsx.

[5] Huang & von Lilienfeld. "Quantum Machine Learning Using Atom-in-Molecule-Based Fragments Selected on the Fly." *Nat. Chem.* **12**, 945 (2020). DOI: 10/ghqj49.

[6] Tkatchenko. "Machine Learning for Chemical Discovery." *Nat. Commun.* **11**, 4125 (2020). DOI: 10/gg89ph.

[7] Von Lilienfeld & Burke. "Retrospective on a Decade of Machine Learning for Chemical Discovery." *Nat. Commun.* **11**, 4895 (2020). DOI: 10/ghqj5d.

[8] Noé, Tkatchenko, Müller & Clementi. "Machine Learning for Molecular Simulation." *Annu. Rev. Phys. Chem.* **71**, 361 (2020). DOI: 10/ggz88h.

[9] Dral. "Quantum Chemistry in the Age of Machine Learning." *J. Phys. Chem. Lett.* **11**, 2336 (2020). DOI: 10/gg2bcq.

[10] Unke, Chmiela, Sauceda, Gastegger, Poltavsky, Schütt, Tkatchenko & Müller. "Machine Learning Force Fields." *Chem. Rev.* **121**, 10142 (2021). DOI: 10/gj7n5p.

[11] Von Lilienfeld, Müller & Tkatchenko. "Exploring Chemical Compound Space with Quantum-Based Machine Learning." *Nat. Rev. Chem.* **4**, 347 (2020). DOI: 10/ghdztf.

[12] Bian & Xie. "Generative Chemistry: Drug Discovery with Deep Learning Generative Models." *J. Mol. Model.* **27**, 71 (2021). DOI: 10/gkfx6d.

[13] Jones. "Density Functional Theory: Its Origins, Rise to Prominence, and Future." *Rev. Mod. Phys.* **87**, 897 (2015). DOI: 10/f7phd9.

[14] Bartlett & Musiał. "Coupled-Cluster Theory in Quantum Chemistry." *Rev. Mod. Phys.* **79**, 291 (2007). DOI: 10/fpfwfb.

[15] Deringer, Bartók, Bernstein, Wilkins, Ceriotti & Csányi. "Gaussian Process Regression for Materials and Molecules." *Chem. Rev.* **121**, 10073 (2021). DOI: 10/gmkffv.

[16] Behler. "Four Generations of High-Dimensional Neural Network Potentials." *Chem. Rev.* **121**, 10037 (2021). DOI: 10/gjxh9z.

[17] Musil, Grisafi, Bartók, Ortner, Csányi & Ceriotti. "Physics-Inspired Structural Representations for Molecules and Materials." *Chem. Rev.* **121**, 9759 (2021). DOI: 10/gmmn9f.

[18] Li, Collins, Tanha, Gordon & Yaron. "A Density Functional Tight Binding Layer for Deep Learning of Chemical Hamiltonians." *J. Chem. Theory Comput.* **14**, 5764 (2018). DOI: 10/gfqk9f.

[19] Schütt, Gastegger, Tkatchenko, Müller & Maurer. "Unifying Machine Learning and Quantum Chemistry with a Deep Neural Network for Molecular Wavefunctions." *Nat. Commun.* **10**, 5024 (2019). DOI: 10/ggdh49.

[20] Kirkpatrick et al. "Pushing the Frontiers of Density Functionals by Solving the Fractional Electron Problem." *Science* **374**, 1385 (2021). DOI: 10/gnqwfr.

[21] Chandrasekaran, Kamal, Batra, Kim, Chen & Ramprasad. "Solving the electronic structure problem with machine learning." *npj Computational Materials* **5**, 1 (2019).

[22] Welborn, Cheng & Miller III. "Transferability in machine learning for electronic structure via the molecular orbital basis." *Journal of chemical theory and computation* **14**, 4772 (2018).

[23] Nagai, Akashi & Sugino. "Completing density functional theory by machine learning hidden messages from molecules." *npj Computational Materials* **6**, 1 (2020).

[24] Gómez-Bombarelli, Wei, Duvenaud, Hernández-Lobato, Sánchez-Lengeling, Sheberla, Aguilera-Iparraguirre, Hirzel, Adams & Aspuru-Guzik. "Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules." *ACS Cent. Sci.* **4**, 268 (2018). DOI: 10/ch82.

[25] Hoogeboom, Satorras, Vignac & Welling. "Equivariant Diffusion for Molecule Generation in 3D." 2022. arXiv: 2203.17003.

[26] Torlai, Mazzola, Carrasquilla, Troyer, Melko & Carleo. "Neural-Network Quantum State Tomography." *Nat. Phys.* **14**, 447 (2018). DOI: 10/gcz7jj.

[27] Sutton & Barto. *Reinforcement Learning: An Introduction* (MIT press, 2018).

[28] Tesauro. "TD-Gammon, a Self-Teaching Backgammon Program, Achieves Master-Level Play." *Neural Comput.* **6**, 215 (1994). DOI: 10/bqsszt.

[29] Mnih, Kavukcuoglu, Silver, Rusu, Veness, Bellemare, Graves, Riedmiller, Fidjeland, Ostrovski, et al. "Human-Level Control through Deep Reinforcement Learning." *Nature* **518**, 529 (2015). DOI: 10/gc3h75.

[30] Silver, Huang, Maddison, Guez, Sifre, Van Den Driessche, Schrittwieser, Antonoglou, Panneershelvam, Lanctot, et al. "Mastering the Game of Go with Deep Neural Networks and Tree Search." *Nature* **529**, 484 (2016). DOI: 10/f77tw6.

[31] Degrave et al. "Magnetic Control of Tokamak Plasmas through Deep Reinforcement Learning." *Nature* **602**, 414 (2022). DOI: 10/gpg9dn.

[32] Heinrich, Lanctot & Silver. "Fictitious Self-Play in Extensive-Form Games." In *Proceedings of the 32nd International Conference on Machine Learning* (PMLR, 2015), 805–813. URL: https://proceedings.mlr.press/v37/heinrich15.html.

[33] Silver et al. "A General Reinforcement Learning Algorithm That Masters Chess, Shogi, and Go through Self-Play." *Science* **362**, 1140 (2018). DOI: 10/cxq3.

[34] Battaglia. "Machine learning wavefunction." In *Quantum Chemistry in the Age of Machine Learning* (Elsevier, 2023), 577–616.

[35] Manzhos. "Machine learning for the solution of the Schrödinger equation." *Machine Learning: Science and Technology* **1**, 013002 (2020). DOI: 10.1088/2632-2153/ab7d30.

[36] Piela. *Ideas of Quantum Chemistry* 2nd ed. (Elsevier, 2014). DOI: 10.1016/C2017-0-03971-X.

[37] Foulkes, Mitas, Needs & Rajagopal. "Quantum Monte Carlo Simulations of Solids." *Rev. Mod. Phys.* **73**, 33 (2001). DOI: 10/dmgh76.

[38] Bajdich, Mitas, Drobný, Wagner & Schmidt. "Pfaffian Pairing Wave Functions in Electronic-Structure Quantum Monte Carlo Simulations." *Phys. Rev. Lett.* **96**, 130201 (2006). DOI: 10/dms82t.

[39] Han, Zhang & E. "Solving Many-Electron Schrödinger Equation Using Deep Neural Networks." *J. Comput. Phys.* **399**, 108929 (2019). DOI: 10/gmg56w.

[40] Acevedo, Curry, Joshi, Leroux & Malaya. "Vandermonde Wave Function Ansatz for Improved Variational Monte Carlo." In *2020 IEEE/ACM Fourth Workshop on Deep Learning on Supercomputers (DLS)* (IEEE, 2020), 40–47. DOI: 10/gm7822.

[41] Szabo & Ostlund. *Modern Quantum Chemistry* (Dover Publications, 1996).

[42] Becca & Sorella. *Quantum Monte Carlo Approaches for Correlated Systems* 1st ed. (Cambridge Univ. Press, 2017). DOI: 10/h83n.

[43] Teale et al. "DFT Exchange: Sharing Perspectives on the Workhorse of Quantum Chemistry and Materials Science." *Phys. Chem. Chem. Phys.* (2022). DOI: 10/gqphzw.

[44] Chan & Sharma. "The Density Matrix Renormalization Group in Quantum Chemistry." *Annu. Rev. Phys. Chem.* **62**, 465 (2011). DOI: 10/bkvrzh.

[45] Huron, Malrieu & Rancurel. "Iterative Perturbation Calculations of Ground and Excited State Energies from Multiconfigurational Zeroth-order Wavefunctions." *J. Chem. Phys.* **58**, 5745 (1973). DOI: 10/czvzff.

[46] Booth, Thom & Alavi. "Fermion Monte Carlo without Fixed Nodes: A Game of Life, Death, and Annihilation in Slater Determinant Space." *J. Chem. Phys.* **131**, 054106 (2009). DOI: 10/d3828h.

[47] Olsen. "The CASSCF Method: A Perspective and Commentary: CASSCF Method." *Int. J. Quantum Chem.* **111**, 3267 (2011). DOI: 10/cw4pgx.

[48] Neuscamman. "The Jastrow Antisymmetric Geminal Power in Hilbert Space: Theory, Benchmarking, and Application to a Novel Transition State." *J. Chem. Phys.* **139**, 194105 (2013). DOI: 10/gkr59c.

[49] Sabzevari & Sharma. "Improved Speed and Scaling in Orbital Space Variational Monte Carlo." *J. Chem. Theory Comput.* **14**, 6276 (2018). DOI: 10/gqpjjf.

[50] Rubenstein. "Introduction to the variational monte carlo method in quantum chemistry and physics." In *Variational Methods in Molecular Modeling* (Springer, 2017), 285–313.

[51] Toulouse, Assaraf & Umrigar. "Introduction to the variational and diffusion Monte Carlo methods." In *Advances in Quantum Chemistry* (Elsevier, 2016). Vol. 73, 285–314.

[52] Carleo & Troyer. "Solving the Quantum Many-Body Problem with Artificial Neural Networks." *Science* **355**, 602 (2017). DOI: 10/f9qd53.

[53] Saito. "Solving the Bose–Hubbard Model with Machine Learning." *J. Phys. Soc. Jpn.* **86**, 093001 (2017). DOI: 10/gbv224.

[54] Nomura, Darmawan, Yamaji & Imada. "Restricted Boltzmann Machine Learning for Solving Strongly Correlated Quantum Systems." *Phys. Rev. B* **96**, 205152 (2017). DOI: 10/gfxfv6.

[55] Adams, Carleo, Lovato & Rocco. "Variational Monte Carlo Calculations of A≤4 Nuclei with an Artificial Neural-Network Correlator Ansatz." *Phys. Rev. Lett.* **127**, 022502 (2021). DOI: 10/gpz4r3.

[56] Astrakhantsev, Westerhout, Tiwari, Choo, Chen, Fischer, Carleo & Neupert. "Broken-Symmetry Ground States of the Heisenberg Model on the Pyrochlore Lattice." *Phys. Rev. X* **11**, 041021 (2021). DOI: 10/gpz4r2.

[57] Sorella. "Green Function Monte Carlo with Stochastic Reconfiguration." *Phys. Rev. Lett.* **80**, 4558 (1998). DOI: 10/dh43x9.

[58] Amari. "Natural Gradient Works Efficiently in Learning." *Neural Comput.* **10**, 251 (1998). DOI: 10/dtsdwt.

[59] Ay, Jost, Lê & Schwachhöfer. *Information Geometry* (Springer, 2017). DOI: 10/h83m.

[60] Martens & Grosse. "Optimizing Neural Networks with Kronecker-Factored Approximate Curvature." In *Proceedings of the 32nd International Conference on Machine Learning* (PMLR, 2015), 2408–2417. URL: https://proceedings.mlr.press/v37/martens15.html.

[61] Perronnin, Liu, Sánchez & Poirier. "Large-Scale Image Retrieval with Compressed Fisher Vectors." In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE, 2010), 3384–3391. DOI: 10/bzns35.

[62] Krizhevsky, Sutskever & Hinton. "Imagenet Classification with Deep Convolutional Neural Networks." In *Advances in Neural Information Processing Systems 25* (Curran Associates, 2012), 1097–1105. URL: https://papers.nips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html.

[63] Schmidhuber. "Deep Learning in Neural Networks: An Overview." *Neural Netw.* **61**, 85 (2015). DOI: 10/f6v78n.

[64] LeCun, Bengio & Hinton. "Deep Learning." *Nature* **521**, 436 (2015). DOI: 10/bmqp.

[65] McCulloch & Pitts. "A Logical Calculus of the Ideas Immanent in Nervous Activity." *Bull. Math. Biophys.* **5**, 115 (1943). DOI: 10/djsbj6.

[66] Rosenblatt. "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain." *Psychol. Rev.* **65**, 386 (1958). DOI: 10/fg6wr5.

[67] Hornik, Stinchcombe & White. "Multilayer Feedforward Networks Are Universal Approximators." *Neural Netw.* **2**, 359 (1989). DOI: 10/c4wch5.

[68] Werbos. "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences." PhD thesis. Harvard University, 1974.

[69] Linnainmaa. "The Representation of the Cumulative Rounding Error of an Algorithm as a Taylor Expansion of the Local Rounding Errors." MA thesis. University of Helsinki, 1970.

[70] Linnainmaa. "Taylor Expansion of the Accumulated Rounding Error." *BIT Numer. Math.* **16**, 146 (1976). DOI: 10/cv4d63.

[71] Rumelhart, Hinton & Williams. "Learning Representations by Back-Propagating Errors." *Nature* **323**, 533 (1986). DOI: 10/cvjdpk.

[72] Glorot & Bengio. "Understanding the Difficulty of Training Deep Feedforward Neural Networks." In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics* (PMLR, 2010), 249–256. URL: https://proceedings.mlr.press/v9/glorot10a.html.

[73] Hooker. "The Hardware Lottery." 2020. arXiv: 2009.06489.

[74] Dauphin, Pascanu, Gulcehre, Cho, Ganguli & Bengio. "Identifying and Attacking the Saddle Point Problem in High-Dimensional Non-Convex Optimization." In *Advances in Neural Information Processing Systems 27* (Curran Associates, 2014), 2933–2941. URL: https://papers.nips.cc/paper/2014/hash/17e23e50bedc63b4095e3d8204ce063b-Abstract.html.

[75] Choromanska, Henaff, Mathieu, Arous & LeCun. "The Loss Surfaces of Multilayer Networks." In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics* (PMLR, 2015), 192–204. URL: https://proceedings.mlr.press/v38/choromanska15.html.

[76] Bottou & Bousquet. "Learning Using Large Datasets." In *Mining Massive Data Sets for Security* (IOS Press, 2008), 15–26. DOI: 10/hr97.

[77] Bottou & Bousquet. "The Tradeoffs of Large-Scale Learning." In *Optimization for Machine Learning* (MIT Press, 2011), 351–368. URL: https://mitpress.mit.edu/9780262016469/optimization-for-machine-learning/.

[78] Russakovsky, Deng, Su, Krause, Satheesh, Ma, Huang, Karpathy, Khosla, Bernstein, et al. "Imagenet Large Scale Visual Recognition Challenge." *Int. J. Comput. Vis.* **115**, 211 (2015). DOI: 10/gcgk7w.

[79] Abadi et al. "TensorFlow: A System for Large-Scale Machine Learning." In *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation* (USENIX Association, 2016), 265–283. URL: https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi.

[80] Paszke, Gross, Chintala, Chanan, Yang, DeVito, Lin, Desmaison, Antiga & Lerer. "Automatic Differentiation in PyTorch." 2017. URL: https://openreview.net/forum?id=BJJsrmfCZ.

[81] Bradbury, Frostig, Hawkins, Johnson, Leary, Maclaurin & Wanderman-Milne. *JAX: Composable Transformations of Python+ NumPy Programs*. 2018. URL: http://github.com/google/jax.

[82] Goodfellow, Shlens & Szegedy. "Explaining and Harnessing Adversarial Examples." *Third International Conference on Learning Representations (ICLR)* (2015).

[83] LeCun, Bottou, Bengio & Haffner. "Gradient-Based Learning Applied to Document Recognition." *Proc. IEEE* **86**, 2278 (1998). DOI: 10/d89c25.

[84] Shawe-Taylor. "Building Symmetries into Feedforward Networks." In *1989 First IEE International Conference on Artificial Neural Networks* (IET, 1989), 158–162. URL: https://ieeexplore.ieee.org/document/51951.

[85] Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser & Polosukhin. "Attention Is All You Need." In *Advances in Neural Information Processing Systems 30* (Curran Associates, 2017), 5998–6008. URL: https://papers.nips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

[86] Schütt, Sauceda, Kindermans, Tkatchenko & Müller. "Schnet – a Deep Learning Architecture for Molecules and Materials." *J. Chem. Phys.* **148**, 241722 (2018). DOI: 10/gdtbj6.

[87] Bronstein, Bruna, Cohen & Veličković. "Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges." 2021. arXiv: 2104.13478.

[88] Hinton & Salakhutdinov. "Reducing the Dimensionality of Data with Neural Networks." *Science* **313**, 504 (2006). DOI: 10/b7vgmj.

[89] Kingma & Welling. "Auto-Encoding Variational Bayes." 2013. arXiv: 1312.6114.

[90] Goodfellow, Pouget-Abadie, Mirza, Xu, Warde-Farley, Ozair, Courville & Bengio. "Generative Adversarial Nets." In *Advances in Neural Information Processing Systems 27* (Curran Associates, 2014), 2672–2680. URL: https://papers.nips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html.

[91] Rezende & Mohamed. "Variational Inference with Normalizing Flows." In *Proceedings of the 32nd International Conference on Machine Learning* (PMLR, 2015), 1530–1538. URL: https://proceedings.mlr.press/v37/rezende15.html.

[92] Van den Oord, Dieleman, Zen, Simonyan, Vinyals, Graves, Kalchbrenner, Senior & Kavukcuoglu. "WaveNet: A Generative Model for Raw Audio." 2016. arXiv: 1609.03499.

[93] Van den Oord, Kalchbrenner, Vinyals, Espeholt, Graves & Kavukcuoglu. "Conditional Image Generation with PixelCNN Decoders." In *Advances in Neural Information Processing Systems 29* (Curran Associates, 2016), 4797–4805. URL: https://papers.

nips.cc/paper/2016/hash/b1301141feffabac455e1f90a7de2054-Abstract.html.

[94] Sohl-Dickstein, Weiss, Maheswaranathan & Ganguli. "Deep Unsupervised Learning Using Nonequilibrium Thermodynamics." In *Proceedings of the 32nd International Conference on Machine Learning* (PMLR, 2015), 2256–2265. URL: https://proceedings.mlr.press/v37/sohl-dickstein15.html.

[95] Sharir, Levine, Wies, Carleo & Shashua. "Deep Autoregressive Models for the Efficient Variational Simulation of Many-Body Quantum Systems." *Phys. Rev. Lett.* **124**, 020503 (2020). DOI: 10/ggw9km.

[96] Choo, Neupert & Carleo. "Two-Dimensional Frustrated J_1-J_2 Model Studied with Neural Network Quantum States." *Phys. Rev. B* **100**, 125124 (2019). DOI: 10/ghmzrr.

[97] Hibat-Allah, Ganahl, Hayward, Melko & Carrasquilla. "Recurrent Neural Network Wave Functions." **2**, 023358 (2020). DOI: 10/ghmzrq.

[98] Xie, null & Wang. "Ab-Initio Study of Interacting Fermions at Finite Temperature with Neural Canonical Transformation." *J. Mach. Learn.* **1**, 38 (2022). DOI: 10/gqpv2j.

[99] Behler & Parrinello. "Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces." *Phys. Rev. Lett.* **98**, 146401 (2007). DOI: 10/c7kbsq.

[100] Rupp, Tkatchenko, Müller & Lilienfeld. "Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning." *Phys. Rev. Lett.* **108**, 058301 (2012).

[101] Bartók, Kondor & Csányi. "On representing chemical environments." en. *Phys. Rev. B* **87**, 184115 (2013). DOI: 10.1103/PhysRevB.87.184115.

[102] Schütt, Arbabzadah, Chmiela, Müller & Tkatchenko. "Quantum-Chemical Insights from Deep Tensor Neural Networks." *Nat. Commun.* **8**, 13890 (2017). DOI: 10/f9kzdv.

[103] Thomas, Smidt, Kearnes, Yang, Li, Kohlhoff & Riley. "Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds." *arXiv preprint arXiv:1802.08219* (2018). DOI: 10.48550/ARXIV.1802.08219.

[104] Schütt, Unke & Gastegger. *Equivariant message passing for the prediction of tensorial properties and molecular spectra*. en. Tech. rep. arXiv:2102.03150 [physics] type: article. arXiv, 2021. URL: http://arxiv.org/abs/2102.03150.

[105] Miller, Geiger, Smidt & Noé. *Relevance of Rotationally Equivariant Convolutions for Predicting Molecular Properties*. Tech. rep. arXiv, 2020. URL: http://arxiv.org/abs/2008.08461.

[106] Geiger & Smidt. "e3nn: Euclidean Neural Networks." *arXiv preprint arXiv:2207.09453* (2022).

[107] Batzner, Musaelian, Sun, Geiger, Mailoa, Kornbluth, Molinari, Smidt & Kozinsky. "E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials." *Nat. Commun.* **13**, 2453 (2022). DOI: 10.1038/s41467-022-29939-5.

[108] Batatia, Kovács, Simm, Ortner & Csányi. "MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields." *arXiv preprint arXiv:2206.07697* (2022).

[109] Bastiaan J. Braams and Joel M. Bowman. "Ab initio potential energy and dipole moment surfaces for H5O2+." *J. Chem. Phys* **122**, 44308 (2005).

[110] Drautz. "Atomic Cluster Expansion for Accurate and Transferable Interatomic Potentials." *Phys. Rev. B* **99**, 014104 (2019). DOI: 10/ghdhfv.

[111] Allen, Dusson, Ortner & Csányi. "Atomic permutationally invariant polynomials for fitting molecular force fields." *Mach. Learn.: Sci. Tech* **2**, 025017 (2021).

[112] Hermann, Schätzle & Noé. "Deep-Neural-Network Solution of the Electronic Schrödinger Equation." *Nat. Chem.* **12**, 891 (2020). DOI: 10/ghcm5p.

[113] Pfau, Spencer, Matthews & Foulkes. "Ab Initio Solution of the Many-Electron Schrödinger Equation with Deep Neural Networks." *Phys. Rev. Research* **2**, 033429 (2020). DOI: 10/ghmf5c.

[114] Entwistle, Schätzle, Erdman, Hermann & Noé. "Electronic Excited States in Deep Variational Monte Carlo." 2022. arXiv: 2203.09472.

[115] Benali, Gasperich, Jordan, Applencourt, Luo, Bennett, Krogel, Shulenburger, Kent, Loos, et al. "Toward a systematic improvement of the fixed-node approximation in diffusion Monte Carlo for solids—A case study in diamond." *The Journal of Chemical Physics* **153**, 184111 (2020).

[116] Stokes, Izaac, Killoran & Carleo. "Quantum Natural Gradient." *Quantum* **4**, 269 (2020). DOI: 10/ghpb9k.

[117] Feynman & Cohen. "Energy Spectrum of the Excitations in Liquid Helium." *Phys. Rev.* **102**, 1189 (1956). DOI: 10/c77rcn.

[118] Kwon, Ceperley & Martin. "Effects of Three-Body and Backflow Correlations in the Two-Dimensional Electron Gas." *Phys. Rev. B* **48**, 12037 (1993). DOI: 10/bnjx23.

[119] Tocchio, Becca, Parola & Sorella. "Role of Backflow Correlations for the Nonmagnetic Phase of the *t*–*t*' Hubbard Model." *Phys. Rev. B* **78**, 041101 (2008). DOI: 10/fqzg2f.

[120] Luo & Clark. "Backflow Transformations via Neural Networks for Quantum Many-Body Wave Functions." *Phys. Rev. Lett.* **122**, 226401 (2019). DOI: 10/ghp332.

[121] Robledo Moreno, Carleo, Georges & Stokes. "Fermionic Wave Functions from Neural-Network Constrained Hidden States." *Proc. Natl. Acad. Sci.* **119**, e2122059119 (2022). DOI: 10/gqpvnx.

[122] Lovato, Adams, Carleo & Rocco. "Hidden-nucleons neural-network quantum states for the nuclear many-body problem." *Physical Review Research* **4**, 043178 (2022).

[123] Yang & Zhao. "Deep-neural-network solution of the ab initio nuclear structure." *arXiv preprint arXiv:2211.13998* (2022).

[124] Spencer, Pfau, Botev & Foulkes. "Better, Faster Fermionic Neural Networks." 2020. arXiv: 2011.07125.

[125] Taddei, Ruggeri, Moroni & Holzmann. "Iterative Backflow Renormalization Procedure for Many-Body Ground-State Wave Functions of Strongly Interacting Normal Fermi Liquids." *Phys. Rev. B* **91**, 115106 (2015). DOI: 10/gm782w.

[126] Ruggeri, Moroni & Holzmann. "Nonlinear Network Description for Many-Body Quantum Systems in Continuous Space." *Phys. Rev. Lett.* **120**, 205302 (2018). DOI: 10/gdg64n.

[127] Chakravorty, Gwaltney & Davidson'. "Ground-state correlation energies for atomic ions with to 18 electrons." *Phys. Rev. A* **44**, 7071 (1991).

[128] Ma, Towler, Drummond & Needs. "Scheme for Adding Electron–Nucleus Cusps to Gaussian Orbitals." *J. Chem. Phys.* **122**, 224322 (2005). DOI: 10/ffqp2b.

[129] Schätzle, Hermann & Noé. "Convergence to the Fixed-Node Limit in Deep Variational Monte Carlo." *J. Chem. Phys.* **154**, 124108 (2021). DOI: 10/gm7xvm.

[130] Kingma & Ba. "Adam: A Method for Stochastic Optimization" (2015). URL: https://arxiv.org/abs/1412.6980.

[131] Gerard, Scherbela, Marquetand & Grohs. "Gold-Standard Solutions to the Schrödinger Equation Using Deep Learning: How Much Physics Do We Need?" 2022. arXiv: 2205.09438.

[132] Glehn, Spencer & Pfau. "A Self-Attention Ansatz for Ab-initio Quantum Chemistry." *arXiv preprint arXiv:2211.13672* (2022).

[133] Scherbela, Reisenhofer, Gerard, Marquetand & Grohs. "Solving the Electronic Schrödinger Equation for Multiple Nuclear Geometries with Weight-Sharing Deep Neural Networks." *Nat. Comput. Sci.* **2**, 331 (2022). DOI: 10/gqpvzj.

[134] Yang, Hu & Li. "Scalable Variational Monte Carlo with Graph Neural Ansatz." 2020. arXiv: 2011.12453.

[135] Gao & Günnemann. "Ab-Initio Potential Energy Surfaces by Pairing GNNs with Neural Wave Functions." In *International Conference on Learning Representations (ICLR 2022)* (OpenReview, 2022). URL: https://openreview.net/forum?id=apv504XsysP.

[136] Gao & Günnemann. "Sampling-Free Inference for Ab-Initio Potential Energy Surface Networks." 2022. arXiv: 2205.14962.

[137] Feynman. "Forces in Molecules." *Physical Review* **56**, 340 (1939).

[138] Assaraf & Caffarel. "Zero-variance zero-bias principle for observables in quantum Monte Carlo: Application to forces." *The Journal of Chemical Physics* **119**, 10536 (2003).

[139] Umrigar. "Two aspects of quantum Monte Carlo: determination of accurate wavefunctions and determination of potential energy surfaces of molecules." *International Journal of Quantum Chemistry* **36**, 217 (1989).

[140] Qian, Fu, Ren & Chen. "Interatomic force from neural network based variational quantum Monte Carlo." *The Journal of Chemical Physics* **157**, 164104 (2022).

[141] Pescia, Han, Lovato, Lu & Carleo. "Neural-Network Quantum States for Periodic Systems in Continuous Space." *Phys. Rev. Res.* **4**, 023138 (2022). DOI: 10/gqpvnw.

[142] Wilson, Moroni, Holzmann, Gao, Wudarski, Vegge & Bhowmik. "Wave Function Ansatz (but Periodic) Networks and the Homogeneous Electron Gas." 2022. arXiv: 2202.04622.

[143] Cassella, Sutterud, Azadi, Drummond, Pfau, Spencer & Foulkes. "Discovering Quantum Phase Transitions with Fermionic Neural Networks." 2022. arXiv: 2202.05183.

[144] Li, Li & Chen. "Ab Initio Calculation of Real Solids via Neural Network Ansatz." 2022. arXiv: 2203.15472v2.

[145] Hammond, Reynolds & Lester. "Valence Quantum Monte Carlo with Ab Initio Effective Core Potentials." *J. Chem. Phys.* **87**, 1130 (1987). DOI: 10/b59sdb.

[146] Li, Fan, Ren & Chen. "Fermionic Neural Network with Effective Core Potential." *Phys. Rev. Research* **4**, 013021 (2022). DOI: 10/gpz4rx.

[147] Needs, Towler, Drummond, López Ríos & Trail. "Variational and Diffusion Quantum Monte Carlo Calculations with the CASINO Code." *J. Chem. Phys.* **152**, 154106 (2020). DOI: 10/gm783w.

[148] Shi & Zhang. "Some Recent Developments in Auxiliary-Field Quantum Monte Carlo for Real Materials." *J. Chem. Phys.* **154**, 024107 (2021). DOI: 10/gptz6w.

[149] Wilson, Gao, Wudarski, Rieffel & Tubman. "Simulations of State-of-the-Art Fermionic Neural Network Wave Functions with Diffusion Monte Carlo." 2021. arXiv: 2103.12570.

[150] Ren, Fu & Chen. "Towards the Ground State of Molecules via Diffusion Monte Carlo on Neural Networks." 2022. arXiv: 2204.13903.

[151] Schautz & Filippi. "Optimized Jastrow–Slater Wave Functions for Ground and Excited States: Application to the Lowest States of Ethene." *J. Chem. Phys.* **120**, 10931 (2004). DOI: 10/cn3tqr.

[152] Dash, Feldt, Moroni, Scemama & Filippi. "Excited States with Selected Configuration Interaction-Quantum Monte Carlo: Chemically Accurate Excitation Energies and Geometries." *J. Chem. Theory Comput.* **15**, 4896 (2019). DOI: 10/gqpvn4.

[153] Zhao & Neuscamman. "An Efficient Variational Principle for the Direct Optimization of Excited States." *J. Chem. Theory Comput.* **12**, 3436 (2016). DOI: 10/gqpvn3.

[154] Pathak, Busemeyer, Rodrigues & Wagner. "Excited States in Variational Monte Carlo Using a Penalty Method." *J. Chem. Phys.* **154**, 034101 (2021). DOI: 10/gm9xw2.

[155] Choo, Carleo, Regnault & Neupert. "Symmetries and Many-Body Excitations with Neural-Network Quantum States." *Phys. Rev. Lett.* **121**, 167204 (2018). DOI: 10/gfgtrg.

[156] Cuzzocrea, Scemama, Briels, Moroni & Filippi. "Variational Principles in Quantum Monte Carlo: The Troubled Story of Variance Minimization." *J. Chem. Theory Comput.* **16**, 4203 (2020). DOI: 10/gqpvn2.

[157] Choo, Mezzacapo & Carleo. "Fermionic Neural-Network States for Ab-Initio Electronic Structure." *Nat. Commun.* **11**, 2368 (2020). DOI: 10/ghmzr5.

[158] Yoshioka, Mizukami & Nori. "Solving Quasiparticle Band Spectra of Real Solids Using Neural-Network Quantum States." *Commun. Phys.* **4**, 1 (2021). DOI: 10/gj6tnr.

[159] Jordan & Wigner. "Über Das Paulische Äquivalenzverbot." *Z. Phys.* **47**, 631 (1928). DOI: 10/d4dxmd.

[160] Bravyi & Kitaev. "Fermionic Quantum Computation." *Ann. Phys.* **298**, 210 (2002). DOI: 10/cx53kk.

[161] Yang, Sugiyama, Tsuda & Yanai. "Artificial Neural Networks Applied as Molecular Wave Function Solvers." *J. Chem. Theory Comput.* **16**, 3513 (2020). DOI: 10/gm782z.

[162] Torlai, Mazzola, Carleo & Mezzacapo. "Precise Measurement of Quantum Observables with Neural-Network Estimators." *Phys. Rev. Res.* **2**, 022060 (2020). DOI: 10/ghmzsh.

[163] Iouchtchenko, Gonthier, Perdomo-Ortiz & Melko. "Neural Network Enhanced Measurement Efficiency for Molecular Groundstates." 2022. arXiv: 2206.15449.

[164] Glielmo, Rath, Csányi, De Vita & Booth. "Gaussian Process States: A Data-Driven Representation of Quantum Many-Body Physics." *Phys. Rev. X* **10**, 041026 (2020). DOI: 10/gmjhdq.

[165] Del Re, Ladik & Biczó. "Self-Consistent-Field Tight-Binding Treatment of Polymers. I. Infinite Three-Dimensional Case." *Phys. Rev.* **155**, 997 (1967). DOI: 10/czgzkm.

[166] Barrett, Malyshev & Lvovsky. "Autoregressive Neural-Network Wavefunctions for Ab Initio Quantum Chemistry." *Nat. Mach. Intell.* **4**, 351 (2022). DOI: 10/gp6r33.

[167] Zhao, Stokes & Veerapaneni. "Scalable Neural Quantum States Architecture for Quantum Chemistry." 2022. arXiv: 2204.13903.

[168] Giner, Scemama & Caffarel. "Using Perturbatively Selected Configuration Interaction in Quantum Monte Carlo Calculations." *Can. J. Chem.* **91**, 879 (2013). DOI: 10/gffmj3.

[169] Holmes, Tubman & Umrigar. "Heat-Bath Configuration Interaction: An Efficient Selected Configuration Interaction Algorithm Inspired by Heat-Bath Sampling." *J. Chem. Theory Comput.* **12**, 3674 (2016). DOI: 10/f82q4p.

[170] Sharma, Holmes, Jeanmairet, Alavi & Umrigar. "Semistochastic Heat-Bath Configuration Interaction Method: Selected Configuration Interaction with Semistochastic Perturbation Theory." *J. Chem. Theory Comput.* **13**, 1595 (2017). DOI: 10/f9wgq3.

[171] Greer. "Monte Carlo Configuration Interaction." *J. Comput. Phys.* **146**, 181 (1998). DOI: 10/bckb4c.

[172] Coe. "Machine Learning Configuration Interaction." *J. Chem. Theory Comput.* **14**, 5739 (2018). DOI: 10/gfpc2m.

[173] Goings, Hu, Yang & Li. "Reinforcement Learning Configuration Interaction." *J. Chem. Theory Comput.* **17**, 5482 (2021). DOI: 10/gpz4rz.

[174] Pineda Flores. "Chembot: A Machine Learning Approach to Selective Configuration Interaction." *Journal of Chemical Theory and Computation* **17**, 4028 (2021). DOI: 10.1021/acs.jctc.1c00196.

[175] Nooijen*, Shamasundar & Mukherjee. "Reflections on size-extensivity, size-consistency and generalized extensivity in many-body theory." *Molecular Physics* **103**, 2277 (2005).

[176] Hutter. "On Representing (Anti)Symmetric Functions." 2020. arXiv: 2007.15298.

Acknowledgements

We 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, 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 the NCCR MARVEL, a National Centre of Competence in Research, under Grant No. 205602. We are grateful to Nobuyuki Yoshioka for providing us with the raw data of Ref..[158] WMCF and his collaborators gratefully acknowledge PRACE for awarding them access to the JUWELS Booster supercomputer (https://apps.fz-juelich.de/jsc/hps/juwels/booster-overview.html); the HPC RIVR consortium (https://www.hpc-rivr.si) and EuroHPC JU (https://eurohpc-ju.europa.eu) 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/T022221/1) and the Digital Research Infrastructure programme (EP/W032244/1) and is operated by Advanced Research Computing at the University of Birmingham.

Competing interests

There are no competing interests.

Glossary of quantum and machine-learning 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, due 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 is antisymmetric with respect to electron exchange.

Variational principle Enables to solve the Schrödinger equation by minimizing the energy E over all antisymmetric wavefunctions ψ

Basis set and orbitals 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 to solve 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.

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/Equivariance Functions whose output does not change/changes in the same way in response to group transformation of the inputs, e.g. rotations, translations, permutations.

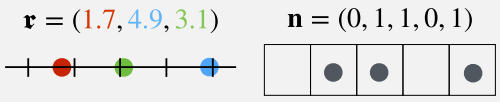
Markov chain Monte Carlo Numerical method to sample a probability distribution by generating a sequence of random variables (RV) in which each RV is generated by perturbing the previous RV.

Box 1 | 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 [36, Vol. 2, Appx. U]. In first quantization, one works with the individual electrons and their coordinates directly in real space ($\mathbf{r}_i \in \mathbb{R}^3$, $i = 1, \ldots, N$) as in Eq. (1). Here, the wavefunction $\psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ 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), labeled 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 $\psi_{n_1 n_2 \cdots} \equiv \psi(n_1, n_2 \cdots)$ 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 1D and in a finite basis of size 5. $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$.

Box 2 | Variational Monte Carlo

Optimization of wavefunctions with neural networks naturally leads to the variational Monte Carlo (VMC) framework. First, Monte Carlo integration of Eq. (3) can handle ansatzes with arbitrary analytical forms for which analytical integrals are not available. Second, VMC samples these integrals stochastically which 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 approaches[37] and more recently in the discrete-basis second-quantized setting.[48,49]

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

$$\langle \hat{H} \rangle_\psi = \mathbb{E}_{\mathbf{x} \sim |\psi(\mathbf{x})|^2} \left[E_{\mathrm{loc}}(\mathbf{x}; \psi) \right]. \tag{10}$$

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

A straightforward and generally applicable way to obtain the samples is 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 [50, Sec. 2.2]. 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 [51, Sec. 1.4].

The VMC formula for the expectation value is exact in the limit of infinite sample size, $\mathcal{N} \to \infty$. Still, in practice, it incurs a statistical error proportional to $\sqrt{\mathrm{Var}[E_{\mathrm{loc}}]/\mathcal{N}}$. While $1/\sqrt{\mathcal{N}}$ converges slowly with 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.

Box 3 | Optimizing neural-network ansatzes

Up to the statistical error, the VMC expectation value for the energy (Box 2) obeys the variational principle (Eq. 4). VMC exploits 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 Eq. (1) and the ansatz will approximate the ground state wavefunction Ψ.

The most straightforward optimization method is gradient descent, where 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 2).

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 account.[57] Stochastic reconfiguration approximates an imaginary-time evolution where each iteration tries to approximate best the state $\mathrm{e}^{-\eta \hat{H}} |\psi\rangle$, where η is a step size. It is similar to the natural gradient descent algorithm[58] that is well-known in the ML community, and where the correlations between parameters are encoded in the Fisher information matrix.[59] Kronecker-factored approximate curvature (KFAC) approach is an approximate version of the natural gradient descent that has been designed to be efficient specifically for neural networks.[60]