

2 Variational Wave Functions for Molecules and Solids

W.M.C. Foulkes

Department of Physics

Imperial College London

Contents

1	Introduction	2
2	Slater determinants	4
3	The Hartree-Fock approximation	11
4	Configuration expansions	14
5	Slater-Jastrow wave functions	21
6	Beyond Slater determinants	27

1 Introduction

Chemists and condensed matter physicists are lucky to have a reliable “grand unified theory” — the many-electron Schrödinger equation — capable of describing almost every phenomenon we encounter. If only we were able to solve it! Finding the exact solution is believed to be “NP hard” in general [1], implying that the computational cost almost certainly scales exponentially with N . Until we have access to a working quantum computer, the best we can do is seek good approximate solutions computable at a cost that rises less than exponentially with system size. Another problem is that our approximate solutions have to be surprisingly accurate to be useful. The energy scale of room-temperature phenomena is $k_B T \approx 0.025$ eV per electron, and the energy differences between competing solid phases can be as small as 0.01 eV per atom [2]. Quantum chemists say that 1 kcal mol⁻¹ (≈ 0.043 eV) is “chemical accuracy” and that methods with errors much larger than this are not good enough to provide quantitative predictions of room temperature chemistry. Yet the natural energy scale built in to the many-electron Schrödinger equation is 1 Hartree (≈ 27.2 eV) per electron,¹ which is about 630 times larger. The total energy of a medium-sized atom can be hundreds of Hartrees, so we need to be able to calculate energies to at least five significant figures. Describing low-temperature many-body phenomena such as magnetism, superconductivity, heavy fermions, and spin liquids requires another couple of orders of magnitude. Single-precision arithmetic (accurate to about seven significant figures) is not good enough.

Thinking about a simple non-interacting electron gas shows that calculating the properties of solids to very high precision also requires very large simulations. Suppose that you want to reach an accuracy of 0.1 eV per electron in an electron gas with Fermi energy $E_F = 10$ eV ≈ 0.37 Hartrees. Since

$$E = \frac{1}{2}k^2, \quad \left[E = \frac{\hbar^2 k^2}{2m} \text{ in MKS units} \right] \quad (1)$$

the energy ratio $E/E_F = 0.01$ implies a wave vector ratio $k/k_F = 0.1$. The Fermi wave vector $k_F = \sqrt{2E_F} \approx 0.86 a_0^{-1}$, so the wave vector k associated with an accuracy of 0.1 eV is $0.086 a_0^{-1}$. The corresponding length scale is $2\pi/k \approx 73 a_0$. Given that the electron density $n = k_F^3/(3\pi^2)$ is around $0.021 a_0^{-3}$, you need to solve the Schrödinger equation for a simulation cell containing of order $n\lambda^3 \approx 8,000$ electrons, which is rarely possible in interacting systems. A great deal of effort has gone into understanding and correcting the finite-size errors that arise when smaller simulation cells are used [3].

The obvious conclusion is that attempting to study chemical reactions by starting from the many-electron Schrödinger equation is a fool’s errand; it would be much better to work with

¹This chapter uses dimensionless equations involving only the numerical values of physical quantities. The numerical values are as measured in Hartree atomic units, where Planck’s constant $\hbar = 1$, the permittivity of free space $\epsilon_0 = 1/4\pi$ (so $4\pi\epsilon_0 = 1$), the electron mass $m = 1$, and the elementary charge $e = 1$. Distances are made dimensionless by dividing by the Hartree atomic unit of length, $a_0 = 4\pi\epsilon_0\hbar^2/(me^2) \approx 0.529 \cdot 10^{-10}$ m, which is also known as the Bohr radius. Energies are made dimensionless by dividing by the Hartree atomic unit of energy, $\hbar^2/(ma_0^2) = e^2/(4\pi\epsilon_0 a_0) \approx 27.2$ eV.

a low-energy effective theory. Most of the strongly correlated phenomena of interest in many-body physics take place at energy scales smaller than 0.025 eV and are indeed treated using low-energy theories (even the Hubbard model, which ignores all but a few bands near the Fermi level, is a low-energy theory by electronic structure standards), but we have at present no reliable low-energy theory of chemical bond breaking and formation. The pseudo-potential approximation allows the core electrons to be eliminated from the Schrödinger equation with only a small loss in accuracy, but that is as far as we can go. If we want to use quantum theory to understand the mechanical properties of solids or follow chemical reactions in real time, the only option is to solve the Schrödinger equation to extraordinarily high precision.

How might we accomplish this? The most widely used electronic structure method is density functional theory (DFT) [4–6], which is reasonably accurate and can, with enough effort, be scaled to thousands of electrons. Although the Hohenberg-Kohn theorem shows that DFT is in principle capable of producing exact ground-state energies and electron densities, this guarantee is of little value in practice because we do not know the exact exchange-correlation functional. DFT calculations for weakly correlated Fermi liquids give excellent qualitative results and reasonably good quantitative results, but current exchange-correlation functionals are far from capable of delivering chemical accuracy consistently. DFT's main contribution to the study of strong-correlation effects has been as a useful framework on which to build more sophisticated approaches. Dynamical mean-field theory, the *GW* approximation, and the Bethe-Salpeter equation [7] are examples of these.

Guessing the form of the many-electron wave function has proved to be a surprisingly successful approach to complicated many-electron problems. The Bethe Ansatz [8] for one-dimensional systems, the BCS theory of superconductivity [9], and Laughlin's treatment of the fractional quantum Hall effect [10] are all good examples. When seeking ground states, a common approach is to guess a trial wave function Ψ with a number of adjustable parameters and vary the parameters until the energy expectation value,

$$E[\Psi] = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle}, \quad (2)$$

is minimized. According to the variational principle, this is the best you can do given the constraints imposed by the assumed functional form.

The greatest successes of guessing the wave function have been in many-body theory, but this article is about approximate wave functions used in electronic structure theory and quantum chemistry. The aim here is to guess the ground-state wave function accurately enough to identify the most stable molecular and crystal structures, study chemical and biochemical reactions, and follow atomic rearrangements in solids, such as those associated with fracture processes or the motion of dislocations. It is rare to achieve chemical accuracy in systems larger than small molecules, but it is possible to outperform DFT in most cases. For the most part we will work in the Schrödinger picture, using trial wave functions of the form $\Psi(x_1, x_2, \dots, x_N)$, where $x_i = (\mathbf{r}_i, \sigma_i)$ is shorthand for the combined spatial and spin coordinates of electron i .

Before we can use the variational principle to optimize the parameters of a trial wave function, we need to be able to work out the energy expectation value $E[\Psi]$. This is not an easy task when

the number of electrons N is 10, never mind when it is 100 or 1000. In the variational quantum Monte Carlo (VMC) method [11, 7], the energy expectation value is rewritten as

$$E[\Psi] = \int \left(\frac{\hat{H}\Psi(x_1, \dots, x_N)}{\Psi(x_1, \dots, x_N)} \right) \left(\frac{|\Psi(x_1, \dots, x_N)|^2}{\int |\Psi|^2 dx_1 \dots dx_N} \right) dx_1 \dots dx_N, \quad (3)$$

where integrals over x are understood to include a spin sum:

$$\int dx' = \sum_{\sigma'} \iiint d^3r'. \quad (4)$$

The $\hat{H}\Psi/\Psi$ term is called the *local energy*, and the $|\Psi|^2/\int|\Psi|^2$ term, which is positive and integrates to one, is interpreted as a probability density in coordinate space. Points in this space are specified by giving $3N$ position variables and N binary spin variables.

As long as it is possible to evaluate the local energy, one can obtain statistical estimates of the value of $E[\Psi]$ using Monte Carlo integration. The Metropolis algorithm [7, 11] is used to sample random coordinate-space points from the probability density $|\Psi|^2/\int|\Psi|^2$, and the values of the local energy at the sampled points are averaged. Neither the Metropolis algorithm (which uses only ratios of the probability density at different points) nor the evaluation of the local energy require knowledge of the normalization of the wave function, so wave functions used in VMC simulations do not need to be normalized. Other more sophisticated and accurate quantum Monte Carlo (QMC) methods, including diffusion quantum Monte Carlo (DMC) [11, 7] and auxiliary-field QMC [12], are also used to simulate molecules and solids and produce much more accurate results, but all require trial wave functions as a starting point.

To whittle down the amount of material, I have had to leave out several important types of trial wave function: the subject is larger than is apparent from this article. I have omitted all discussion of pairing wave functions such as the BCS wave function [9], geminals [13] and Pfaffians [14]. Until recently I would have said that attempts to use pairing wave functions to describe non-superconducting electrons had produced disappointing results, but a new preprint [15] has changed my mind. I have also omitted the family of trial wave functions that developed from the density matrix renormalization group [16] and includes matrix product states and tensor network states [17]. These are very important in low-dimensional model systems and becoming more important in quantum chemistry.

2 Slater determinants

Non-interacting electrons

Let us start by thinking about a molecule or periodically-repeated simulation cell containing N *non-interacting* electrons. The many-electron Hamiltonian is

$$\hat{H} = \sum_{i=1}^N \left(-\frac{1}{2} \nabla_{\mathbf{r}_i}^2 + V(x_i) \right) = \sum_{i=1}^N \hat{h}(x_i), \quad (5)$$

where $\hat{h}(x_i) = -\frac{1}{2}\nabla_{\mathbf{r}_i}^2 + V(x_i)$ acts on the coordinates of electron i only. In the simplest plausible model of a molecule or solid, the effective potential might look like this:

$$V(x) = V_{\text{nuc}}(\mathbf{r}) + V_{\text{Hartree}}(\mathbf{r}) = -\sum_I \frac{Z_I}{|\mathbf{r} - \mathbf{d}_I|} + \int \frac{n(x')}{|\mathbf{r} - \mathbf{r}'|} dx', \quad (6)$$

where Z_I is the atomic number of the (fixed, classical) nucleus at position \mathbf{d}_I , and $n(x) = n(\mathbf{r}, \sigma)$ is the number density of spin σ electrons at point \mathbf{r} . In Hartree-Fock theory [18, 5, 7], $V(x)$ also contains a non-local spin-dependent exchange potential. In the Kohn-Sham equations of density functional theory (DFT) [4–6], it contains a local exchange-correlation potential.

The many-electron Schrödinger equation for the N non-interacting electrons,

$$\hat{H}\Psi(x_1, x_2, \dots, x_N) = E\Psi(x_1, x_2, \dots, x_N), \quad (7)$$

is a separable partial differential equation with solutions of the form

$$\Psi(x_1, x_2, \dots, x_N) = \varphi_1(x_1) \varphi_2(x_2) \dots \varphi_N(x_N). \quad (8)$$

(Such solutions are not totally antisymmetric, but let us ignore this problem for the time being.) Substituting the trial solution into the Schrödinger equation gives

$$(\hat{h}\varphi_1)\varphi_2 \dots \varphi_N + \varphi_1(\hat{h}\varphi_2) \dots \varphi_N + \dots + \varphi_1 \varphi_2 \dots (\hat{h}\varphi_N) = E\varphi_1 \varphi_2 \dots \varphi_N. \quad (9)$$

If we now divide by $\varphi_1 \varphi_2 \dots \varphi_N$ we get,

$$\frac{\hat{h}(x_1)\varphi_1(x_1)}{\varphi_1(x_1)} + \frac{\hat{h}(x_2)\varphi_2(x_2)}{\varphi_2(x_2)} + \dots + \frac{\hat{h}(x_N)\varphi_N(x_N)}{\varphi_N(x_N)} = E. \quad (10)$$

The first term depends only on x_1 , the second only on x_2 , and so on, but the sum must be the constant E . This is only possible if each and every term is constant:

$$\hat{h}_1\varphi_1 = \varepsilon_1\varphi_1, \quad \hat{h}_2\varphi_2 = \varepsilon_2\varphi_2, \quad \dots, \quad \hat{h}_N\varphi_N = \varepsilon_N\varphi_N, \quad (11)$$

with

$$E = \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_N. \quad (12)$$

Functions such as $\varphi_i(x)$, obtained by solving a one-electron Schrödinger equation of the form $\hat{h}\varphi_i = \varepsilon_i\varphi_i$, are called *one-electron orbitals* or *one-electron energy eigenfunctions*.

Although Ψ is not antisymmetric, we can easily construct an antisymmetric linear combination of solutions with the N electrons distributed among the N one-electron orbitals in different ways

$$\Psi(x_1, x_2, \dots, x_N) = \frac{1}{\sqrt{N!}} \sum_P (-1)^{\zeta_P} \varphi_{P1}(x_1) \varphi_{P2}(x_2) \dots \varphi_{PN}(x_N). \quad (13)$$

Every term in the linear combination is an eigenfunction of \hat{H} with the same eigenvalue E , so the linear combination is also an eigenfunction with eigenvalue E .

- The list P_1, P_2, \dots, P_N is a permutation P of the list $1, 2, \dots, N$. The sum is over all permutations, with ζ_P the total number of pair interchanges needed to build the permutation P . The value of $(-1)^{\zeta_P}$ is $+1$ when P is an even permutation and -1 when P is an odd permutation.
- The $1/\sqrt{N!}$ is a normalizing factor.
- When $N = 2$,

$$\Psi(x_1, x_2) = \frac{1}{\sqrt{2}} [\varphi_1(x_1)\varphi_2(x_2) - \varphi_2(x_1)\varphi_1(x_2)] = \frac{1}{\sqrt{2}} \begin{vmatrix} \varphi_1(x_1) & \varphi_1(x_2) \\ \varphi_2(x_1) & \varphi_2(x_2) \end{vmatrix}. \quad (14)$$

- Generally,

$$\Psi(x_1, x_2, \dots, x_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_1(x_1) & \varphi_1(x_2) & \dots & \dots & \varphi_1(x_N) \\ \varphi_2(x_1) & \varphi_2(x_2) & \dots & \dots & \varphi_2(x_N) \\ \vdots & \vdots & \dots & \dots & \vdots \\ \vdots & \vdots & \dots & \dots & \vdots \\ \varphi_N(x_1) & \varphi_N(x_2) & \dots & \dots & \varphi_N(x_N) \end{vmatrix}. \quad (15)$$

Wave functions of this type are called Slater determinants. (For bosons we can use analogous symmetrized sums of products called permanents.)

- If two or more of $\varphi_1, \varphi_2, \dots, \varphi_N$ are the same, two or more rows of the determinant are the same and the wave function is zero; this is how the Pauli exclusion principle follows from the antisymmetry. If two electrons of the same spin approach the same point in space, even when all of the φ_i are different, two columns of the determinant become the same and the wave function is again zero. The antisymmetry built in to the Slater determinant helps to keep spin-parallel electrons apart.
- If we add a component of φ_2 to φ_1 ,

$$\tilde{\varphi}_1(x) = \varphi_1(x) + c\varphi_2(x), \quad (16)$$

the Slater determinant is unchanged:

$$\begin{aligned} \tilde{D}(x_1, x_2, \dots, x_N) &= \frac{1}{\sqrt{N!}} \begin{vmatrix} \tilde{\varphi}_1(x_1) & \tilde{\varphi}_1(x_2) & \dots & \tilde{\varphi}_1(x_N) \\ \varphi_2(x_1) & \varphi_2(x_2) & \dots & \varphi_2(x_N) \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \varphi_N(x_1) & \varphi_N(x_2) & \dots & \varphi_N(x_N) \end{vmatrix} \\ &= \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_1(x_1) & \varphi_1(x_2) & \dots & \varphi_1(x_N) \\ \varphi_2(x_1) & \varphi_2(x_2) & \dots & \varphi_2(x_N) \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \varphi_N(x_1) & \varphi_N(x_2) & \dots & \varphi_N(x_N) \end{vmatrix} + \frac{1}{\sqrt{N!}} \begin{vmatrix} c\varphi_2(x_1) & c\varphi_2(x_2) & \dots & c\varphi_2(x_N) \\ \varphi_2(x_1) & \varphi_2(x_2) & \dots & \varphi_2(x_N) \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \varphi_N(x_1) & \varphi_N(x_2) & \dots & \varphi_N(x_N) \end{vmatrix} \\ &= D(x_1, x_2, \dots, x_N). \end{aligned} \quad (17)$$

This shows that no generality is lost by assuming that the one-electron orbitals are orthonormal.

Interacting electrons

In the interacting N -electron Schrödinger equation, the one-electron operator $V(x)$ (which depends on the electron density in DFT and the one-electron density matrix in Hartree-Fock theory) is replaced by the electron-electron interaction

$$\left(\sum_{i=1}^N \left(-\frac{1}{2} \nabla_i^2 + V_{\text{nuc}}(\mathbf{r}_i) \right) + \frac{1}{2} \sum_{i=1}^N \sum_{\substack{j=1 \\ (j \neq i)}}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \right) \Psi(x_1, x_2, \dots, x_N) = E \Psi(x_1, x_2, \dots, x_N). \quad (18)$$

I have assumed for simplicity that spin-orbit interactions can be neglected, so the Hamiltonian is independent of spin. Reintroducing the electron-electron interaction may look like a small change but it has large consequences: the Schrödinger equation is no longer separable and the many-electron wave functions are no longer Slater determinants. Unlike the non-interacting Schrödinger equation, the interacting version cannot be solved exactly for systems of more than a few electrons, even using the world's most powerful computers.

Slater determinants as basis functions

Although Slater determinants are not exact solutions of the many-electron Schrödinger equation with interactions, we can still use them as *basis functions*. Suppose that $\varphi_1(x), \varphi_2(x), \dots$ are a complete orthonormal basis for the one-particle Hilbert space. A common way to choose the $\varphi_i(x)$ is to solve a one-electron or mean-field Schrödinger equation, usually obtained from density functional or Hartree-Fock theory, and use the resulting one-electron orbitals.

Given a complete basis for the one-electron Hilbert space, the set of all products of the form

$$\varphi_{i_1}(x_1) \varphi_{i_2}(x_2) \dots \varphi_{i_N}(x_N) \quad (19)$$

is a complete basis for the N -particle Hilbert space. If the N particles are electrons or other fermions, only antisymmetrized products are required and we can express the wave function as a linear combination of Slater determinants

$$\Psi(x_1, x_2, \dots, x_N) = \sum_{\mathbf{i}} C_{\mathbf{i}} D_{\mathbf{i}}(x_1, x_2, \dots, x_N), \quad (20)$$

where the sum is over all distinct determinants, the vector index $\mathbf{i} = (i_1, i_2, \dots, i_N)$ identifies the N one-electron basis function $\varphi_{i_1}, \varphi_{i_2}, \dots, \varphi_{i_N}$, appearing in determinant $D_{\mathbf{i}}$, and the $C_{\mathbf{i}}$ are expansion coefficients. Interchanging any two basis functions leaves $D_{\mathbf{i}}$ unaltered (bar a sign), so we can restrict the summation to vector indices \mathbf{i} for which $i_1 < i_2 < \dots < i_N$.²

Another way to index Slater determinants is to use the occupation number representation, in which every determinant is defined by a list of binary numbers, one for each one-electron basis

²It is interesting to think about how this works in an infinite system. Even if the set of one-electron basis functions is countable, the set of all ordered subsets of the set of one-electron basis functions is not, implying that we cannot index the determinants using natural numbers. Perhaps attempting to describe large systems using wave functions is not such a good idea?

function. If basis function φ_i appears in the determinant (is “occupied”), the i ’th binary number is set to 1; otherwise, it is set to 0.

Slater determinants are the “building blocks” of many-fermion physics, including most of the approximate many-electron wave functions we will be looking at.

Slater determinants and second quantization

The properties of Slater determinants underlie the properties of the creation and annihilation operators, \hat{c}_p^\dagger and \hat{c}_p , which are defined by their actions in the determinantal basis:

$$\hat{c}_p^\dagger |D_{i_1, i_2, \dots, i_N}\rangle = |D_{p, i_1, i_2, \dots, i_N}\rangle, \quad \hat{c}_p |D_{p, i_1, i_2, \dots, i_N}\rangle = |D_{i_1, i_2, \dots, i_N}\rangle. \quad (21)$$

If the determinant $|D_{i_1, i_2, \dots, i_N}\rangle$ already contains orbital φ_p , acting with \hat{c}_p^\dagger produces a determinant $|D_{p, i_1, i_2, \dots, i_N}\rangle$ with two identical rows. The result is therefore equal to zero. Similarly, if $|D_{i_1, i_2, \dots, i_N}\rangle$ does not contain φ_p , then \hat{c}_p finds nothing to annihilate and $\hat{c}_p |D_{i_1, i_2, \dots, i_N}\rangle = 0$. Whenever you are working with creation and annihilation operators, you are in fact manipulating Slater determinants.

The fermion anti-commutation relations follow from the antisymmetry of the determinantal basis. For example, given any determinant containing φ_q in the k ’th row but not containing φ_p , where $p \neq q$, we have

$$\begin{aligned} \hat{c}_p^\dagger \hat{c}_q |D_{i_1, \dots, i_{k-1}, q, i_{k+1}, \dots, i_N}\rangle &= (-1)^{k-1} \hat{c}_p^\dagger \hat{c}_q |D_{q, i_1, \dots, i_{k-1}, i_{k+1}, \dots, i_N}\rangle \\ &= (-1)^{k-1} |D_{p, i_1, \dots, i_{k-1}, i_{k+1}, \dots, i_N}\rangle \end{aligned} \quad (22)$$

and

$$\begin{aligned} \hat{c}_q \hat{c}_p^\dagger |D_{i_1, \dots, i_{k-1}, q, i_{k+1}, \dots, i_N}\rangle &= \hat{c}_q |D_{p, i_1, \dots, i_{k-1}, q, i_{k+1}, \dots, i_N}\rangle \\ &= (-1)^k \hat{c}_q |D_{q, p, i_1, \dots, i_{k-1}, i_{k+1}, \dots, i_N}\rangle \\ &= (-1)^k |D_{p, i_1, \dots, i_{k-1}, i_{k+1}, \dots, i_N}\rangle, \end{aligned} \quad (23)$$

implying that

$$(\hat{c}_p^\dagger \hat{c}_q + \hat{c}_q \hat{c}_p^\dagger) |D_{i_1, \dots, i_{k-1}, q, i_{k+1}, \dots, i_N}\rangle = 0. \quad (24)$$

If D_{i_1, \dots, i_N} already contains φ_p or does not contain φ_q , the operators $\hat{c}_p^\dagger \hat{c}_q$ and $\hat{c}_q \hat{c}_p^\dagger$ annihilate it and Eq. (24) still holds. Since the basis of Slater determinants is complete, it follows that

$$\hat{c}_p^\dagger \hat{c}_q + \hat{c}_q \hat{c}_p^\dagger = 0, \quad p \neq q. \quad (25)$$

If you have never worked through a detailed explanation of how the properties of fermion creation and annihilation operators arise from the properties of the Slater determinants on which they act, try Chapter 1 of Negele and Orland [19] for a physicist’s perspective or Chapter 1 of Helgaker, Jorgensen and Olsen [20] for a chemist’s perspective.

Exchange and correlation

If we write the full many-electron Hamiltonian,

$$\hat{H} = \sum_{i=1}^N \left(-\frac{1}{2} \nabla_i^2 + V_{\text{nuc}}(\mathbf{r}_i) \right) + \frac{1}{2} \sum_{i=1}^N \sum_{\substack{j=1 \\ (j \neq i)}}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \quad (26)$$

in the form

$$\hat{H} = \sum_{i=1}^N \hat{h}(x_i) + \frac{1}{2} \sum_{i=1}^N \sum_{\substack{j=1 \\ (j \neq i)}}^N \hat{v}(x_i, x_j), \quad (27)$$

and evaluate its expectation value $E = \langle D | \hat{H} | D \rangle$ in a normalized Slater determinant D of orthonormal one-electron functions φ_i , it is in principle straightforward, although in practice tedious [7, 18, 20], to show that

$$E = \sum_{i=1}^N \langle \varphi_i | \hat{h} | \varphi_i \rangle + \frac{1}{2} \sum_{i=1}^N \sum_{\substack{j=1 \\ (j \neq i)}}^N \left(\langle \varphi_i \varphi_j | \hat{v} | \varphi_i \varphi_j \rangle - \langle \varphi_j \varphi_i | \hat{v} | \varphi_i \varphi_j \rangle \right), \quad (28)$$

where

$$\langle \varphi_i | \hat{h} | \varphi_j \rangle = \int \varphi_i^*(x) \hat{h}(x) \varphi_j(x) dx, \quad (29)$$

$$\langle \varphi_i \varphi_j | \hat{v} | \varphi_k \varphi_l \rangle = \iint \varphi_i^*(x) \varphi_j^*(x') \hat{v}(x, x') \varphi_k(x) \varphi_l(x') dx dx'. \quad (30)$$

In the case of Coulomb interactions, when $\hat{v}(x, x') = 1/|\mathbf{r} - \mathbf{r}'|$, the first contribution to the electron-electron interaction energy is

$$\frac{1}{2} \iint \frac{\sum_i |\varphi_i(x)|^2 \sum_{j(\neq i)} |\varphi_j(x')|^2}{|\mathbf{r} - \mathbf{r}'|} dx dx'. \quad (31)$$

This is known as the Hartree energy and is equal to the sum of the classical Coulomb interaction energies of the charge densities associated with the one-electron orbitals appearing in the determinant. We could have guessed it would appear. The second contribution to the interaction energy, which is known as the *exchange* energy because the order of the orbitals in the bra is reversed relative to their order in the ket, describes how the Pauli principle affects the electron-electron interactions. The antisymmetry built into the Slater determinant prevents spin-parallel electrons from getting close to each other, and this decreases the positive (*i.e.*, repulsive) Coulomb energy of the electrons. The exchange term is therefore negative.

If the spins in a solid line up, so that there are more electrons of one spin than the other, the exchange energy (which acts only between electrons of the same spin) becomes more negative and the total electron-electron interaction energy (which is positive) is lowered. At the same time, the one-electron contribution to the total energy rises because electrons have been promoted from lower-energy occupied minority-spin orbitals to higher-energy unoccupied majority-spin

orbitals. In some solids, the lowering of the Coulomb energy wins and the spins polarize spontaneously. This is more likely when the hopping kinetic energy is small because the atoms are far apart and the energy bands are narrow. Exchange interactions are the primary cause of magnetism.

Real electrons are also kept apart by the repulsive Coulomb interactions between them. This effect, called *correlation*, is not included in a simple Slater determinant wave function. The Hartree energy of the Slater determinant is low when the electronic charge distribution is smooth, which helps to keep electrons away from regions in which there are lots of other electrons *on average*, but does not keep individual pairs of electrons apart. In a real solid, electrons are kept apart both by the Pauli principle (exchange), which is included in the Slater determinant, and by the Coulomb interaction (correlation), which is not.

One way to understand exchange and correlation is in terms of the pair density $n(\mathbf{r}, \sigma; \mathbf{r}', \sigma')$, defined such that $n(\mathbf{r}, \sigma; \mathbf{r}', \sigma') d^3r d^3r'$ is proportional to the probability of finding a spin σ electron in the volume element d^3r at \mathbf{r} and a (different) spin σ' electron in the volume element d^3r' at \mathbf{r}' . The closely related pair-correlation function, $g(\mathbf{r}, \sigma; \mathbf{r}', \sigma')$, is defined by

$$n(\mathbf{r}, \sigma; \mathbf{r}', \sigma') = n(\mathbf{r}, \sigma)g(\mathbf{r}, \sigma; \mathbf{r}', \sigma')n(\mathbf{r}', \sigma'). \quad (32)$$

If the volume elements d^3r and d^3r' are far apart, the numbers of electrons in d^3r and d^3r' are statistically independent; one therefore expects $\lim_{|\mathbf{r}'-\mathbf{r}|\rightarrow\infty} g(\mathbf{r}, \sigma; \mathbf{r}', \sigma') = 1$.

Figure 1 shows the pair-correlation functions of pairs of spin-parallel and spin-antiparallel electrons in a uniform electron gas, calculated assuming that the wave function is a Slater determinant of plane waves. The way in which antisymmetry keeps pairs of spin-parallel electrons apart is clear, as is the failure of pairs of spin-antiparallel electrons to avoid each other. Antisymmetry alone is not sufficient to correlate (in the statistical sense) the positions of spin-antiparallel electrons.

Basic notions of probability theory tell us that

$$n(\mathbf{r}, \sigma; \mathbf{r}', \sigma') d^3r d^3r' = n(\mathbf{r}, \sigma | \mathbf{r}', \sigma') d^3r \times n(\mathbf{r}', \sigma') d^3r', \quad (33)$$

where $n(\mathbf{r}, \sigma | \mathbf{r}', \sigma') d^3r$ is proportional to the conditional probability of finding a spin σ electron in d^3r given that there is a spin σ' electron in d^3r' . Since there are $N-1$ electrons in the system, excluding the one frozen at \mathbf{r}' , the conditional density satisfies the sum rule

$$\sum_{\sigma} \int n(\mathbf{r}, \sigma | \mathbf{r}', \sigma') d^3r = N - 1. \quad (34)$$

Describing statistical correlations in terms of conditional probabilities is very natural, but it is often easier to think about the exchange-correlation hole, $n_{xc}(\mathbf{r}, \sigma | \mathbf{r}', \sigma')$, defined by

$$n(\mathbf{r}, \sigma | \mathbf{r}', \sigma') = n(\mathbf{r}, \sigma) + n_{xc}(\mathbf{r}, \sigma | \mathbf{r}', \sigma'). \quad (35)$$

The density of spin σ electrons at \mathbf{r} would be $n(\mathbf{r}, \sigma)$ in the absence of the frozen spin σ' electron at \mathbf{r}' , so the exchange-correlation hole provides a very direct picture of the change (normally a

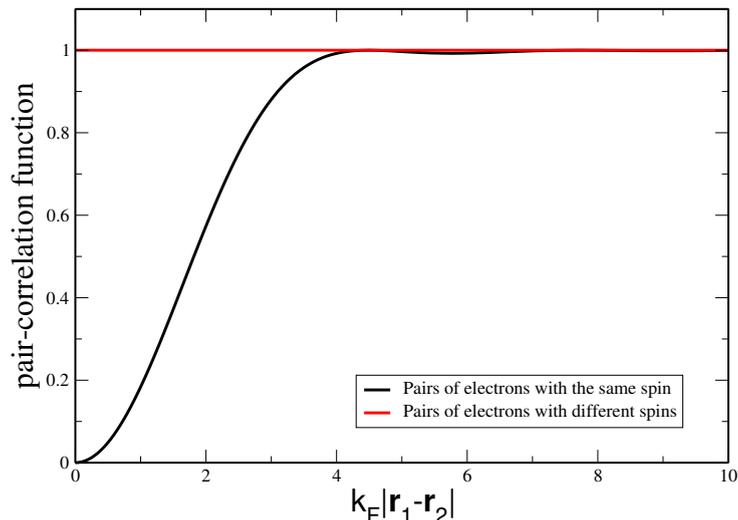


Fig. 1: Pair-correlation functions of pairs of spin-parallel and spin-antiparallel electrons in a uniform electron gas, calculated assuming that the wave function is a single Slater determinant of plane waves. The antisymmetry of the wave function helps to keep pairs of spin-parallel electrons apart but does not affect the pair-correlation function of pairs of spin-antiparallel electrons.

reduction) in electron density caused by the presence of the frozen electron. It follows from the definition of the exchange-correlation hole and the conditional-probability sum rule that

$$\sum_{\sigma} \int n_{xc}(\mathbf{r}, \sigma | \mathbf{r}', \sigma') d^3r = -1. \quad (36)$$

Every electron of charge $-e$ is thus surrounded by a hole of charge $+e$. If the hole is close to the electron, which is not always the case, the entire *quasi-particle* — the electron plus its hole — is charge neutral and might be expected to have only short-ranged interactions. This helps explain why the models of non-interacting electrons used in undergraduate solid-state physics courses work so well.

Figure 2 shows two views of the spin-summed exchange-correlation hole, $\sum_{\sigma} n_{xc}(\mathbf{r}, \sigma | \mathbf{r}', \sigma')$, around a single electron frozen at the point \mathbf{r}' in the middle of a bond in silicon [21]. The graphs were calculated using the VMC method and a Slater-Jastrow wave function (see later), which includes both exchange and correlation effects.

3 The Hartree-Fock approximation

Although the many-electron eigenfunctions of real molecules and solids can in principle be written as linear combinations of (huge numbers of) Slater determinants, we will see later that the number of determinants required rises exponentially with the number of electrons. This forces us to try something less ambitious. In the Hartree-Fock approximation, the variational principle is used to find the *single* Slater determinant that best approximates the many-electron ground state [18, 20, 6]. It turns out that the one-electron orbitals appearing in the best possible

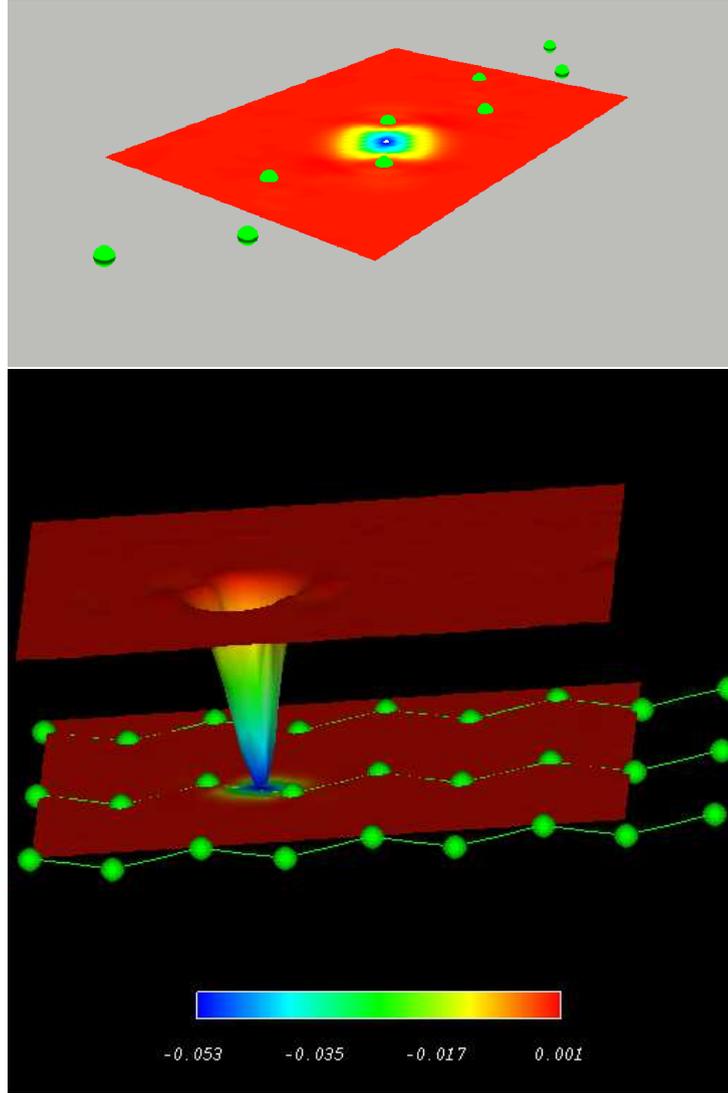


Fig. 2: Two views of the exchange-correlation hole around an electron in the middle of a bond in silicon. The zig-zag chains of atoms lie in the Si (110) plane. From Ref. [21].

single determinant obey a mean-field Schrödinger-like equation,

$$\begin{aligned} \hat{h}(x) \varphi_n(x) + \sum_{\substack{j=1 \\ (j \neq n)}}^N \int dx' \varphi_j^*(x') \hat{v}(x, x') \varphi_j(x') \varphi_n(x) \\ - \sum_{\substack{j=1 \\ (j \neq n)}}^N \int dx' \varphi_j^*(x') \hat{v}(x, x') \varphi_n(x') \varphi_j(x) = \lambda_n \varphi_n(x), \end{aligned} \quad (37)$$

known as the Hartree-Fock equation. The electron-electron interactions have been replaced by an effective potential with two contributions: the first summation describes the action of the Hartree potential on orbital $\varphi_n(x)$ and the second the action of the exchange potential. Notice that the exchange potential is actually an integral operator. The $j=n$ terms in both summations cancel if they are included, so the form of the Hartree-Fock differential equation is

independent of n . Another partial explanation for the success of one-electron concepts such as band-structures, atomic orbitals, and π and σ bonds, in interacting systems where these ideas appear to make little sense, is that the Hartree-Fock approximation is often reasonably accurate. Other mean-field-like methods, such as DFT, work better than Hartree-Fock for many purposes, but the wave function in DFT is an artificial construct introduced to help calculate the kinetic energy of a fictitious system of non-interacting electrons with the same position-dependent number density as the interacting system and has little to do with the true many-electron wave function [4–6]. Our aim here is to devise approximate many-electron wave functions for atoms, molecules and solids, so Hartree-Fock theory is a better starting point.

The Hartree-Fock Hamiltonian depends on the one-electron orbitals obtained by solving the Hartree-Fock equation, so we have a chicken and egg problem: we cannot find the orbitals until we know the Hamiltonian; but we cannot work out the Hamiltonian until we know the orbitals. As usual in mean-field theories, we have to iterate until the inputs and outputs of the mean-field equation are consistent with each other:

1. Guess the set of one-electron orbitals φ_j ($j = 1, \dots, N$) and construct the corresponding Hartree-Fock Hamiltonian.
2. Solve the Hartree-Fock equation to find a new set of one-electron orbitals. (This is quite tricky because the exchange term is an integral operator with a Coulomb kernel that diverges as $|\mathbf{r} - \mathbf{r}'| \rightarrow 0$, but it can be done.)
3. Use the new set of one-electron orbitals to construct a new Hartree-Fock Hamiltonian.
4. Repeat steps 2 and 3 until the set of one-electron orbitals no longer changes from cycle to cycle.

There is no guarantee that iterative algorithms of this type are stable, and clever tricks are sometimes required to make them converge, but self-consistent solutions can be found within a reasonable amount of computer time for systems of up to a few hundred electrons.

Until about 20 years ago, Hartree-Fock theory was widely used to study molecules, even though it is far from being able to reach chemical accuracy. DFT has now become dominant, partly because DFT calculations are easier to do and normally more accurate, and partly because our imperfect knowledge of the exchange-correlation functional leaves more scope for tweaking the calculations to make them give the right answers! A common tweak is to mix fractions of the Hartree-Fock exchange energy into the DFT exchange-correlation functional, making DFT calculations more similar to Hartree-Fock calculations. “Hybrid” density functionals including a portion of exact exchange are not consistently able to achieve chemical accuracy, but are often accurate enough to provide useful results. Hartree-Fock methods were not much used in solids until fairly recently because the calculations were difficult; their main use now is in wide band-gap insulators, where the results are not too bad. For metals, Hartree-Fock is something of a disaster, but hybrid exchange-correlation functionals including screened exchange interactions are popular and successful.

4 Configuration expansions

Configuration-interaction methods

We saw in Sec. 2 that any many-electron wave function can be expressed as a linear combination of Slater determinants (also known as configurations):

$$\Psi(x_1, x_2, \dots, x_N) = \sum_{\mathbf{i}} C_{\mathbf{i}} D_{\mathbf{i}}(x_1, x_2, \dots, x_N). \quad (38)$$

This expansion underlies most of the approximate wave functions used in traditional quantum chemistry. In the full-configuration-interaction (FCI) method [20], the sum over determinants is made finite by choosing a finite set of M ($\geq N$) one-electron basis functions (normally Hartree-Fock or DFT one-electron orbitals) and approximating Ψ as a linear combination of the ${}^M C_N$ distinct N -electron Slater determinants that can be built using them. The vector of expansion coefficients that minimizes the energy expectation value,

$$E = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \frac{\sum_{\mathbf{i}, \mathbf{j}} C_{\mathbf{i}}^* \langle D_{\mathbf{i}} | \hat{H} | D_{\mathbf{j}} \rangle C_{\mathbf{j}}}{\sum_{\mathbf{k}} C_{\mathbf{k}}^* C_{\mathbf{k}}}, \quad (39)$$

is easily shown to be the lowest eigenvector of the matrix eigenvalue problem

$$\sum_{\mathbf{j}} H_{\mathbf{ij}} C_{\mathbf{j}} = E C_{\mathbf{i}}, \quad (40)$$

where $H_{\mathbf{ij}} = \langle D_{\mathbf{i}} | \hat{H} | D_{\mathbf{j}} \rangle$ is an ${}^M C_N \times {}^M C_N$ matrix. The use of the variational principle is exactly as in one-electron quantum theory, but the basis functions are many-electron Slater determinants rather than one-electron orbitals and it is more difficult to work out the Hamiltonian matrix elements.

Unfortunately, the number of determinants required to approximate the ground state to a given accuracy rises exponentially with the system size, making FCI calculations impractical for anything but the smallest molecules. Suppose (very optimistically) that you can obtain a reasonably good description of the ground state of a single helium atom using a one-electron basis set containing $1s$ and $2s$ orbitals only. Since every atom holds four spin-orbitals ($1s \uparrow$, $1s \downarrow$, $2s \uparrow$, $2s \downarrow$), the FCI basis for a system of $N/2$ helium atoms and N electrons contains

$${}^M C_N = {}^{2N} C_N = \frac{(2N)!}{N! N!} \quad (41)$$

determinants. Using Stirling's approximation, $\ln(n!) \approx n \ln n - n$ for large n , gives

$${}^{2N} C_N \approx e^{(2 \ln 2)N}, \quad (42)$$

which rises exponentially with N . Even in this minimal and inaccurate basis set, calculating the many-electron ground state of a system of 5 helium atoms and 10 electrons requires finding the lowest eigenvector of a matrix with more than a million rows and columns. Dealing with 10 helium atoms requires a Hamiltonian matrix with over 10^{12} rows and columns and 10^{24}

elements. The size of the matrix can often be reduced using symmetry arguments, and the H_{ij} matrix is sparse because applying the Hamiltonian to a determinant changes at most two of the one-electron orbitals, but increasing the system size soon makes the state vector too large to store and manipulate.

It seems reasonable to hope that most of the Slater determinants in the vast FCI basis set are unnecessary and can be neglected. Even when the He atoms are far enough apart to be independent, however, simple truncation schemes do not work well.

The exact (within the FCI basis) $S_z = 0$ ground state of a single He atom is a linear combination of four determinants,

$$\begin{aligned} \Psi(x, x') = & C_1 \begin{vmatrix} \varphi_{1s}(\mathbf{r})\chi_{\uparrow}(\sigma) & \varphi_{1s}(\mathbf{r}')\chi_{\uparrow}(\sigma') \\ \varphi_{1s}(\mathbf{r})\chi_{\downarrow}(\sigma) & \varphi_{1s}(\mathbf{r}')\chi_{\downarrow}(\sigma') \end{vmatrix} + C_2 \begin{vmatrix} \varphi_{2s}(\mathbf{r})\chi_{\uparrow}(\sigma) & \varphi_{2s}(\mathbf{r}')\chi_{\uparrow}(\sigma') \\ \varphi_{2s}(\mathbf{r})\chi_{\downarrow}(\sigma) & \varphi_{2s}(\mathbf{r}')\chi_{\downarrow}(\sigma') \end{vmatrix} \\ & + C_3 \begin{vmatrix} \varphi_{1s}(\mathbf{r})\chi_{\uparrow}(\sigma) & \varphi_{1s}(\mathbf{r}')\chi_{\uparrow}(\sigma') \\ \varphi_{2s}(\mathbf{r})\chi_{\downarrow}(\sigma) & \varphi_{2s}(\mathbf{r}')\chi_{\downarrow}(\sigma') \end{vmatrix} + C_4 \begin{vmatrix} \varphi_{1s}(\mathbf{r})\chi_{\downarrow}(\sigma) & \varphi_{1s}(\mathbf{r}')\chi_{\downarrow}(\sigma') \\ \varphi_{2s}(\mathbf{r})\chi_{\uparrow}(\sigma) & \varphi_{2s}(\mathbf{r}')\chi_{\uparrow}(\sigma') \end{vmatrix}, \quad (43) \end{aligned}$$

where $\chi_{\uparrow}(\sigma) = \delta_{\sigma,\uparrow}$ and $\chi_{\downarrow}(\sigma) = \delta_{\sigma,\downarrow}$ are the usual up-spin and down-spin S_z eigenstates. The exact (within the basis) ground state of a system of 10 well separated He atoms is an antisymmetrized product of the ground states of each individual atom:

$$\begin{aligned} \Psi(x_1, x_2, \dots, x_{20}) = & \mathcal{N} \hat{\mathcal{A}} [\Psi_1(x_1, x_2) \Psi_2(x_3, x_4) \dots \Psi_{10}(x_{19}, x_{20})] \\ = & \frac{\mathcal{N}}{N!} \sum_P (-1)^{\zeta_P} \Psi_1(x_{P1}, x_{P2}) \Psi_2(x_{P3}, x_{P4}) \dots \Psi_{10}(x_{P19}, x_{P20}), \quad (44) \end{aligned}$$

where $\hat{\mathcal{A}}$ is the antisymmetrization operator and \mathcal{N} is a normalizing constant. The atomic ground states $\Psi_1(x_1, x_2)$, $\Psi_2(x_3, x_4)$, \dots , $\Psi_{10}(x_{19}, x_{20})$ are all of the same form but translated with respect to one another because they are centered on different atoms. Every atom has a finite probability, $p_{\text{excited}} = |C_2|^2 + |C_3|^2 + |C_4|^2$, of being found in an excited configuration in which at least one of the two electrons is occupying a $2s$ orbital. If the states of all 10 atoms were measured repeatedly, the average number found in excited configurations would be $10p_{\text{excited}}$. If the number of He atoms in the system were doubled, the expected number found in excited configurations would also double.

Suppose that you try to describe the $S_z = 0$ ground state of the system of 10 isolated He atoms using an FCI singles and doubles basis set, consisting of the Hartree-Fock ground-state determinant, in which all 20 electrons are occupying $1s$ orbitals, plus the 14,250 distinct $S_z = 0$ determinants with one or two of the twenty electrons occupying $2s$ orbitals. Three of the 118 determinants occurring in the $S_z = 0$ FCI singles and doubles basis set for a system of three He atoms are illustrated in Fig. 3. In one of these configurations an electron has been transferred from one atom to another; this excitation is very unlikely to happen when the atoms are far apart but might become more important as they approach each other.

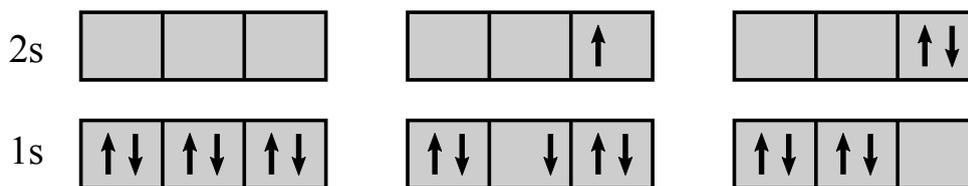


Fig. 3: Three of the 118 $S_z = 0$ configurations appearing in the FCI singles and doubles wave function for three He atoms. Only the 1s and 2s orbitals on each atom are included in the one-electron basis set. The configuration on the left is the Hartree-Fock ground state; the one in the middle contains a single electron-hole excitation; and the one on the right contains two excitations.

No configuration within the singles and doubles basis set contains more than two excited electrons, so no more than two of the 10 He atoms can be found in (charge neutral) excited configurations at any one time. If the expected number of excited atoms ($10p_{\text{excited}}$) is substantially less than 2 that might be sufficient, but what happens as the system size is increased? For large enough systems (and large enough is often very small), the expected number of atoms in excited configurations will exceed 2 and the basis set will be inadequate.

All naive attempts to truncate the FCI basis set introduce analogues of this problem, the most obvious symptom of which is that the calculated energy of a system of N well-separated atoms is greater than N times the energy of one atom. (Remember that FCI is a variational method, so lower energies are better energies.) In the language of quantum chemistry, the truncated FCI method is not size consistent; in the language of condensed matter physics, the results are not extensive. The problem gets worse as the number of atoms increases, with the fraction of the correlation energy recovered reducing to zero as the system size tends to infinity. If you try to fix the problem by increasing the maximum number of excited electrons in proportion to the system size, the truncated FCI method becomes exponentially scaling.

Even when the FCI eigenvalue problem is too large to solve using the conventional methods of linear algebra, molecules with up to a few tens of electrons can often be treated using the FCI QMC method [22,23], in which the contributions to the numerator and denominator of Eq. (39) are sampled stochastically without ever storing the complete eigenvector. This approach has also been used to study very small solid-state simulation cells subject to periodic boundary conditions [24]. Unfortunately, the fermion sign problem [25] imposes another limitation on the number of electrons and the maximum system size remains disappointingly small. Another way to increase the system size is to use one of several selected-CI approaches [26–29], which iteratively identify determinants that make important contributions to the ground state, neglecting the rest. The selection can reduce the rate at which the basis set increases with system size, but selected-CI methods remain exponentially scaling.

To summarize, although FCI methods can produce extraordinarily accurate results for light atoms and small molecules, they are of little value for large molecules or solids.

Coupled-cluster methods

The coupled-cluster method [20, 30] offers a better way of truncating the FCI Hilbert space. To motivate the idea, consider two well-separated He atoms, A and B , at positions \mathbf{d}_A and \mathbf{d}_B . The four-electron Hamiltonian is

$$\hat{H} = \sum_{i=1}^4 \left(-\frac{1}{2} \nabla_{\mathbf{r}_i}^2 + V_{\text{nuc}}(\mathbf{r}_i - \mathbf{d}_A) + V_{\text{nuc}}(\mathbf{r}_i - \mathbf{d}_B) \right) + \frac{1}{2} \sum_{i=1}^4 \sum_{\substack{j=1 \\ (j \neq i)}}^4 \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}. \quad (45)$$

What happens when we apply this Hamiltonian to the product $\Psi_A(x_1, x_2)\Psi_B(x_3, x_4)$ of the two atomic ground states? (The product is not a valid four-electron wave function because it is not antisymmetric on exchange of electrons between atoms, but we will antisymmetrize it later on.) If the two atoms are far enough apart, all interactions involving pairs of charged particles on different atoms can be neglected and we get

$$\begin{aligned} \hat{H}\Psi_A(x_1, x_2)\Psi_B(x_3, x_4) &\approx \left[\hat{H}_A(x_1, x_2) + \hat{H}_B(x_3, x_4) \right] \Psi_A(x_1, x_2)\Psi_B(x_3, x_4) \\ &= 2E_{\text{atom}} \Psi_A(x_1, x_2)\Psi_B(x_3, x_4), \end{aligned} \quad (46)$$

where H_A and H_B are the two-electron Hamiltonians for the two separate atoms. Applying the antisymmetrization operator \hat{A} to both sides of this equation gives

$$\hat{A}\hat{H}\Psi_A(x_1, x_2)\Psi_B(x_3, x_4) \approx 2E_{\text{atom}} \hat{A}\Psi_A(x_1, x_2)\Psi_B(x_3, x_4). \quad (47)$$

The Hamiltonian is totally symmetric on exchange of particles, so it commutes with \hat{A} to leave

$$\hat{H} \left[\hat{A}\Psi_A(x_1, x_2)\Psi_B(x_3, x_4) \right] \approx 2E_{\text{atom}} \left[\hat{A}\Psi_A(x_1, x_2)\Psi_B(x_3, x_4) \right]. \quad (48)$$

We have reached the obvious result: the energy of two well-separated atoms is the sum of the atomic energies and the wave function is an antisymmetrized product of the atomic wave functions.

Within an FCI expansion, the ground states on atoms A and B can be expanded in Slater determinants constructed using orbitals on those two atoms:

$$\Psi_A(x_1, x_2) = \sum_{\mathbf{i}_A} C_{\mathbf{i}_A} D_{\mathbf{i}_A}(x_1, x_2), \quad \Psi_B(x_3, x_4) = \sum_{\mathbf{i}_B} C_{\mathbf{i}_B} D_{\mathbf{i}_B}(x_3, x_4), \quad (49)$$

where \mathbf{i}_A lists the occupied orbitals on atom A and \mathbf{i}_B lists the occupied orbitals on atom B . Applying the antisymmetrization operator to a product of two Slater determinants, $D_{\mathbf{i}_A}(x_1, x_2)D_{\mathbf{i}_B}(x_3, x_4)$, produces a single larger determinant containing all of the orbitals involved, so

$$\Psi_{AB}(x_1, x_2, x_3, x_4) = \sum_{\mathbf{i}_A, \mathbf{i}_B} C_{\mathbf{i}_A} C_{\mathbf{i}_B} D_{\mathbf{i}_A, \mathbf{i}_B}(x_1, x_2, x_3, x_4). \quad (50)$$

An analogous result holds for any system consisting of well-separated fragments.

An easier way to reach the same conclusions is to use second-quantized notation, where

$$\hat{H} = \sum_{i,j} \hat{c}_i^\dagger h_{ij} \hat{c}_j + \frac{1}{2} \sum_{i,j,k,l} \hat{c}_i^\dagger \hat{c}_j^\dagger V_{ijkl} \hat{c}_l \hat{c}_k, \quad (51)$$

with

$$h_{ij} = \int \varphi_i^*(x) \left(-\frac{1}{2} \nabla^2 + V_{\text{nuc}}(\mathbf{r}-\mathbf{d}_A) + V_{\text{nuc}}(\mathbf{r}-\mathbf{d}_B) \right) \varphi_j(x) dx,$$

$$V_{ijkl} = \iint \varphi_i^*(x) \varphi_j^*(x') \frac{1}{|\mathbf{r}-\mathbf{r}'|} \varphi_k(x) \varphi_l(x') dx dx'. \quad (52)$$

When the two atoms are far enough apart, h_{ij} is negligible unless φ_i and φ_j are on the same atom and V_{ijkl} is negligible unless $\varphi_i, \varphi_j, \varphi_k,$ and φ_l are all on the same atom. Furthermore, if φ_i and φ_j are both on atom A , the $V_{\text{nuc}}(\mathbf{r}-\mathbf{d}_B)$ contribution to h_{ij} is negligible and vice versa. Under these assumptions,

$$\hat{H} \approx \hat{H}_A + \hat{H}_B. \quad (53)$$

Let us write the two-electron ground states of atoms A and B , treated separately, as

$$|\Psi_A\rangle = \hat{\Psi}_A^\dagger |\text{VAC}\rangle, \quad |\Psi_B\rangle = \hat{\Psi}_B^\dagger |\text{VAC}\rangle. \quad (54)$$

For the $S_z = 0$ ground state of Eq. (44), for example, we would have

$$\begin{aligned} \hat{\Psi}_A^\dagger &= C_1 \hat{c}_{A,1s,\uparrow}^\dagger \hat{c}_{A,1s,\downarrow}^\dagger + C_2 \hat{c}_{A,2s,\uparrow}^\dagger \hat{c}_{A,2s,\downarrow}^\dagger + C_3 \hat{c}_{A,1s,\uparrow}^\dagger \hat{c}_{A,2s,\downarrow}^\dagger + C_4 \hat{c}_{A,1s,\downarrow}^\dagger \hat{c}_{A,2s,\uparrow}^\dagger, \\ \hat{\Psi}_B^\dagger &= C_1 \hat{c}_{B,1s,\uparrow}^\dagger \hat{c}_{B,1s,\downarrow}^\dagger + C_2 \hat{c}_{B,2s,\uparrow}^\dagger \hat{c}_{B,2s,\downarrow}^\dagger + C_3 \hat{c}_{B,1s,\uparrow}^\dagger \hat{c}_{B,2s,\downarrow}^\dagger + C_4 \hat{c}_{B,1s,\downarrow}^\dagger \hat{c}_{B,2s,\uparrow}^\dagger. \end{aligned} \quad (55)$$

Since \hat{H}_A commutes with $\hat{\Psi}_B^\dagger$, \hat{H}_B commutes with $\hat{\Psi}_A^\dagger$, and $\hat{\Psi}_A^\dagger$ commutes with $\hat{\Psi}_B^\dagger$ (they would anti-commute if $\hat{\Psi}_A^\dagger$ and $\hat{\Psi}_B^\dagger$ both created odd numbers of electrons, but the argument below is easily generalized to the anti-commuting case), we find that $\hat{\Psi}_A^\dagger \hat{\Psi}_B^\dagger |\text{VAC}\rangle$ is the approximate four-electron ground state:

$$\begin{aligned} \hat{H} \hat{\Psi}_A^\dagger \hat{\Psi}_B^\dagger |\text{VAC}\rangle &\approx (\hat{H}_A + \hat{H}_B) \hat{\Psi}_A^\dagger \hat{\Psi}_B^\dagger |\text{VAC}\rangle \\ &= \hat{\Psi}_B^\dagger \hat{H}_A \hat{\Psi}_A^\dagger |\text{VAC}\rangle + \hat{\Psi}_A^\dagger \hat{H}_B \hat{\Psi}_B^\dagger |\text{VAC}\rangle \\ &= \hat{\Psi}_B^\dagger E_{\text{atom}} \hat{\Psi}_A^\dagger |\text{VAC}\rangle + \hat{\Psi}_A^\dagger E_{\text{atom}} \hat{\Psi}_B^\dagger |\text{VAC}\rangle = 2E_{\text{atom}} \hat{\Psi}_A^\dagger \hat{\Psi}_B^\dagger |\text{VAC}\rangle. \end{aligned} \quad (56)$$

The problem with the FCI singles and doubles wave function is that it neglects the contributions made by strings of four creation operators, two on atom A and two on atom B , appearing in $\hat{\Psi}_A^\dagger \hat{\Psi}_B^\dagger$. The coupled-cluster trial wave function adopts a product form from the beginning, cleverly ensuring that $\hat{\Psi}_{AB}^\dagger$ reduces to $\hat{\Psi}_A^\dagger \hat{\Psi}_B^\dagger$ when fragments A and B are far enough apart.

Instead of creating states from the vacuum, it is convenient to start from a single N -electron determinant, D_0 , normally taken to be the Hartree-Fock ground state. Once D_0 has been chosen, we can separate the orbitals (and corresponding creation and annihilation operators) into two types: the N orbitals appearing in D_0 are denoted φ_i ; and the $M-N$ orbitals not appearing in D_0 (known as ‘‘virtuals’’) are denoted φ_a . The choice of suffix (i, j, k, \dots for occupied orbitals;

a, b, c, \dots for unoccupied orbitals) indicates the type of any given orbital, creation operator, or annihilation operator.

Any determinant in the FCI basis set may be created from D_0 by making a number (ranging from 1 to N) of electron-hole excitations of the form $\hat{X}_i^a = \hat{c}_a^\dagger \hat{c}_i$. The excitation operator \hat{X}_i^a replaces the orbital φ_i appearing in D_0 by the previously unoccupied orbital φ_a . We can also define operators that create multiple electron-hole pairs, such as the double excitation $\hat{X}_{ij}^{ab} = \hat{c}_a^\dagger \hat{c}_b^\dagger \hat{c}_i \hat{c}_j = -\hat{c}_a^\dagger \hat{c}_i \hat{c}_b^\dagger \hat{c}_j = -\hat{X}_i^a \hat{X}_j^b$. Because the excitation operators are constructed using annihilation operators for orbitals in D_0 and creation operators for orbitals not in D_0 , creation and annihilation operators for the same orbital never occur. The excitation operators therefore commute with one another. No orbital in D_0 can be annihilated more than once and no orbital not in D_0 can be created more than once, so products of excitation operators are often zero. For example, $\hat{X}_i^a \hat{X}_i^a = 0$.

In the coupled-cluster method, the many-electron wave function is written in the product form

$$|\Psi\rangle = \left[\prod_{a,i} \left(1 + t_i^a \hat{X}_i^a \right) \right] \left[\prod_{b>a,j>i} \left(1 + t_{ij}^{ab} \hat{X}_{ij}^{ab} \right) \right] \dots |D_0\rangle, \quad (57)$$

where the coupled-cluster amplitudes t_i^a and t_{ij}^{ab}, \dots , are variational parameters. If the system consists of two well-separated fragments, A and B , all amplitudes involving orbitals on both fragments will be zero. After moving the terms involving fragment A to the front of the product of operators, the coupled-cluster wave function takes the separable form $\hat{\Psi}_A^\dagger \hat{\Psi}_B^\dagger |D_0\rangle$, where $\hat{\Psi}_A^\dagger$ and $\hat{\Psi}_B^\dagger$ are just as they would be for an isolated fragment. The coupled-cluster approach is therefore size consistent.

The product of any excitation operator \hat{X} with itself is zero, so

$$1 + t\hat{X} = 1 + t\hat{X} + \frac{1}{2!} (t\hat{X})^2 + \dots = e^{t\hat{X}}. \quad (58)$$

This allows us to rewrite the coupled-cluster Ansatz in the more commonly encountered exponential form

$$\begin{aligned} |\Psi\rangle &= \left[\prod_{a,i} e^{t_i^a \hat{X}_i^a} \right] \left[\prod_{b>a,j>i} e^{t_{ij}^{ab} \hat{X}_{ij}^{ab}} \right] \dots |D_0\rangle \\ &= \exp \left(\sum_{a,i} t_i^a \hat{X}_i^a + \sum_{a>b,i>j} t_{ij}^{ab} \hat{X}_{ij}^{ab} + \dots \right) |D_0\rangle = \exp(\hat{T}) |D_0\rangle. \end{aligned} \quad (59)$$

Truncating the exponent at the single excitation level, including only the $\sum_{i,a} t_i^a \hat{X}_i^a$ term, leads to the coupled-cluster singles (CCS) method; truncating at double excitations yields the coupled-cluster singles and doubles (CCSD) method; and so on. Regardless of the truncation level, the expectation value of the energy of n well-separated molecules is n times the expectation value of the energy of one molecule. The exponential form ensures that the wave function always includes determinants (but not *all* determinants) with up to N electron-hole pair excitations. When the maximum excitation level reaches N , the coupled-cluster method becomes equivalent to FCI.

method	$R = R_{\text{ref}}$	$R = 2R_{\text{ref}}$
HF	0.217822	0.363954
CCSD	0.003744	0.022032
CCSDT	0.000493	-0.001405
CCSDTQ	0.000019	-0.000446

Table 1: *Difference between the energy of a water molecule calculated using the coupled-cluster method and the FCI energy in the same basis set. The convergence is very rapid at the equilibrium bond length, $R_{\text{ref}} = 1.84345a_0$, but slower when the bond length is doubled. The bond angle is fixed at 110.565° and energies are in Hartree atomic units. Note that the coupled-cluster energy may lie below the FCI energy. Coupled-cluster results are not necessarily variational. Data from Ref. [31].*

Table 1 shows how the calculated energy of a water molecule converges as a function of the excitation level. The convergence is very rapid at the equilibrium bond length of $1.184345 a_0$ but slower when the bond length is doubled. Despite running into difficulties if there are multiple very different determinants with similar energies, as is often the case when bonds are being broken, the CCSD method hits a sweet spot between the demands of computational efficiency and accuracy. Adding the effect of triples perturbatively yields the CCSD(T) method, often known as the “gold standard” of quantum chemistry, which frequently produces excellent results. Straightforward implementations of the CCSD and CCSD(T) methods scale steeply with system size, the effort being proportional to N^6 for CCSD and N^7 for CCSD(T), but correlations are local and it is possible to do better than this. Coupled-cluster methods are even beginning to become useful in solids.

It would be nice if it were possible to treat the CC wave function variationally, evaluating the corresponding energy expectation value and adjusting the amplitudes to minimize the total energy. Unfortunately, however, the presence of arbitrary numbers of electron-hole pairs (at all truncation levels) makes the computational effort scale factorially with system size. To motivate a more practical approach, let us suppose for the time being that the cluster operator \hat{T} has not been truncated and that $|\psi\rangle = e^{\hat{T}}|D_0\rangle$ is the exact ground-state wave function:

$$\left(\hat{H} - E_0\right) e^{\hat{T}}|D_0\rangle = 0. \quad (60)$$

Multiplying by $e^{-\hat{T}}$ gives

$$\left(e^{-\hat{T}}\hat{H}e^{\hat{T}} - E_0\right) |D_0\rangle = 0. \quad (61)$$

We can therefore view the coupled-cluster method as a search for the operator $e^{\hat{T}}$ that makes the reference determinant $|D_0\rangle$ the ground state of the similarity-transformed Hamiltonian $\hat{H}_T = e^{-\hat{T}}\hat{H}e^{\hat{T}}$. Note that \hat{H}_T is not Hermitian, so its left and right eigenstates need not be the same; $|D_0\rangle$ is a right eigenstate.

In the CCSD method, we replace the cluster operator by a truncated approximation,

$$\hat{T} \approx \sum_{a,i} t_i^a \hat{X}_i^a + \sum_{a>b, i>j} t_{ij}^{ab} \hat{X}_{ij}^{ab}, \quad (62)$$

so that Eq. (61) is only approximately correct. To make it as correct as possible, we insist that it is satisfied in the subspace consisting of $|D_0\rangle$ and the determinants

$$|D_i^a\rangle = \hat{X}_i^a |D_0\rangle = \hat{c}_a^\dagger \hat{c}_i |D_0\rangle, \quad (63)$$

$$|D_{ij}^{ab}\rangle = \hat{X}_{ij}^{ab} |D_0\rangle = \hat{c}_a^\dagger \hat{c}_b^\dagger \hat{c}_i \hat{c}_j |D_0\rangle. \quad (64)$$

(Since $e^{\hat{T}}$ also contains products of excitation operators, this does not ensure that Eq. (61) is satisfied exactly.) The result is the CCSD equations:

$$\langle D_0 | \hat{H}_T - E_0 | D_0 \rangle = \langle D_0 | \hat{H}_T | D_0 \rangle - E_0 = 0, \quad (65)$$

$$\langle D_i^a | \hat{H}_T - E_0 | D_0 \rangle = \langle D_i^a | \hat{H}_T | D_0 \rangle = 0, \quad (66)$$

$$\langle D_{ij}^{ab} | \hat{H}_T - E_0 | D_0 \rangle = \langle D_{ij}^{ab} | \hat{H}_T | D_0 \rangle = 0, \quad a>b, i>j. \quad (67)$$

The second and third lines provide exactly as many equations as there are amplitudes, allowing us to find the t_i^a and t_{ij}^{ab} . The first line then determines the approximate ground-state energy. Furthermore, because the determinants appearing in the bras contain no more than two electron-hole pairs, the matrix elements can all be evaluated with an effort that scales as a power of the system size.

5 Slater-Jastrow wave functions

Cusps

Quantum chemists often find it useful to divide correlation effects into two separate types. The division is not clear or absolute, but helpful nevertheless. *Static correlation* arises when the ground state contains substantial components of several significantly different determinants with similar energy expectation values. For small molecules, static correlation can in principle be dealt with by including all of the important determinants in the basis set, although the number required may grow exponentially with system size. Configuration-expansion methods such as single-reference CCSD, that make excitations from a single determinant, often find strong static correlations difficult to deal with. *Dynamic correlation* arises from the Coulomb repulsions between nearby electrons. An important contribution is made by the non-analytic cusps in the many-electron wave function at points where pairs of electrons coalesce [32, 33].

The simplest case of a wave function cusp occurs when an electron approaches a nucleus; a good example is provided by the $1s$ energy eigenfunction of a hydrogen atom, $\varphi_{1s}(r) = \frac{1}{\sqrt{\pi}} e^{-r}$. The electron-nucleus cusps can easily be built into the one-electron orbitals (although not when the orbitals are expanded in a smooth analytic basis set of Gaussians or plane waves, as is often done for computational reasons). The electron-electron cusps are difficult to represent in a basis

of Slater determinants and cause FCI expansions to converge very slowly. This is the problem that Slater-Jastrow wave functions solve.

The forms of the cusps can be understood in general [32, 33]. Consider what happens when particles 1 and 2, with masses m_1 and m_2 and charges Z_1 and Z_2 , approach each other. Transforming into center-of-mass and difference coordinates, $\mathbf{r}_{\text{cm}} = (m_1\mathbf{r}_1 + m_2\mathbf{r}_2)/(m_1+m_2)$ and $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$, the N -particle Schrödinger equation becomes

$$\left(-\frac{1}{2\mu}\nabla_{\mathbf{r}}^2 - \frac{1}{2M}\nabla_{\mathbf{r}_{\text{cm}}}^2 - \sum_{i=3}^N \frac{1}{2m_i}\nabla_{\mathbf{r}_i}^2 + \frac{1}{2} \sum_{i=1}^N \sum_{\substack{j=1 \\ (j \neq i)}}^N \frac{Z_i Z_j}{r_{ij}} \right) \Psi(\mathbf{r}, \mathbf{r}_{\text{cm}}, \sigma_1, \sigma_2, x_3, \dots, x_N) = E\Psi(\mathbf{r}, \mathbf{r}_{\text{cm}}, \sigma_1, \sigma_2, x_3, \dots, x_N), \quad (68)$$

where $\mu = m_1 m_2 / (m_1 + m_2)$ is the reduced mass and $M = m_1 + m_2$ is the total mass of the two particles involved. As $\mathbf{r}_1 \rightarrow \mathbf{r}_2$ with $\mathbf{r}_{\text{cm}}, \mathbf{r}_3, \mathbf{r}_4, \dots, \mathbf{r}_N$ held fixed, the $Z_1 Z_2 / r$ divergence in the potential energy must be cancelled by a corresponding divergence in the $\nabla_{\mathbf{r}}^2$ part of the kinetic energy. In other words, we require

$$\frac{1}{\Psi(\mathbf{r})} \left(-\frac{1}{2\mu}\nabla_{\mathbf{r}}^2 + \frac{Z_1 Z_2}{r} \right) \Psi(\mathbf{r}) \quad (69)$$

to remain finite as $r \rightarrow 0$, where $\Psi(\mathbf{r})$ is shorthand for $\Psi(\mathbf{r}, \mathbf{r}_{\text{cm}}, \sigma_1, \sigma_2, x_3, \dots, x_N)$.

Near the origin, we can use the familiar representation of $\Psi(\mathbf{r})$ as a linear combination of spherical harmonics and radial functions,

$$\begin{aligned} \Psi(\mathbf{r}) &= \sum_{l=0}^{\infty} \sum_{m=-l}^l c_{l,m} Y_{l,m}(\vartheta, \varphi) r^l R_l(r) \\ &= \sum_{l=0}^{\infty} \sum_{m=-l}^l c_{l,m} Y_{l,m}(\vartheta, \varphi) r^l \left(1 + b_1^{(l)} r + b_2^{(l)} r^2 + \dots \right), \end{aligned} \quad (70)$$

where the $c_{l,m}$ and $b_i^{(l)}$ are expansion coefficients that depend on $\mathbf{r}_{\text{cm}}, \sigma_1, \sigma_2, x_3, \dots, x_N$. Starting from this representation, a few lines of algebra show that

$$\begin{aligned} \frac{1}{\Psi} \left(-\frac{1}{2\mu}\nabla_{\mathbf{r}}^2 + \frac{Z_1 Z_2}{r} \right) \Psi &= \frac{1}{\Psi} \sum_{l,m} c_{l,m} Y_{l,m} \left(-\frac{1}{2\mu r^2} \frac{\partial}{\partial r} \frac{1}{r} \frac{\partial}{\partial r} + \frac{l(l+1)}{2\mu r^2} + \frac{Z_1 Z_2}{r} \right) R_l \\ &= \frac{\sum_{l,m} c_{l,m} Y_{l,m} \left(\left(Z_1 Z_2 - \frac{l+1}{\mu} b_1^{(l)} \right) r^{l-1} + \dots \right)}{\sum_{l,m} c_{l,m} Y_{l,m} (r^l + \dots)}. \end{aligned} \quad (71)$$

The largest terms in the denominator at small r are the ones corresponding to the smallest angular momentum for which $c_{l,m}$ is non-zero. Denoting this angular momentum by l_0 , we see the right-hand side of Eq. (71) diverges as $r \rightarrow 0$ unless

$$Z_1 Z_2 - \frac{l_0+1}{\mu} b_1^{(l_0)} = 0. \quad (72)$$

We can therefore express $\Psi(\mathbf{r})$ for small r as

$$\Psi(\mathbf{r}) = r^{l_0} \sum_{m=-l_0}^{l_0} c_{l_0,m} Y_{l_0,m} \left(1 + \frac{\mu Z_1 Z_2}{l_0+1} r \right) + r^{l_0+1} \sum_{m=-(l_0+1)}^{l_0+1} c_{l_0+1,m} Y_{l_0+1,m} + \mathcal{O}(r^{l_0+2}). \quad (73)$$

Any wave function that describes particles interacting via Coulomb forces must be of this form. We are interested in three specific cases of this general result.

Electron-nucleus cusps: When an electron (mass $m = 1$, charge -1) approaches a nucleus (mass M , charge $+Z$), symmetry imposes no restrictions on the value of l and one generally expects $l_0 = 0$. Equation (73) then reduces to

$$\Psi(\mathbf{r}) = c_{0,0} Y_{0,0} (1 - Zr) + r \sum_{m=-1}^1 c_{1,m} Y_{1,m} + \mathcal{O}(r^2), \quad (74)$$

where we have noted that $M \gg 1$ and hence that $\mu \approx 1$. Since

$$Y_{1,-1} \propto \frac{x-iy}{r}, \quad Y_{1,0} \propto \frac{z}{r}, \quad Y_{1,1} \propto \frac{x+iy}{r},$$

we can rewrite Eq. (74) as

$$\Psi(\mathbf{r}) = a_{\text{en}} (1 - Zr) + \mathbf{b}_{\text{en}} \cdot \mathbf{r} + \mathcal{O}(r^2), \quad (75)$$

where $a_{\text{en}} = c_{0,0} Y_{0,0}$ and the \mathbf{r} -independent vector \mathbf{b}_{en} depends on $c_{1,-1}$, $c_{1,0}$ and $c_{1,1}$. The $-Zr$ contribution to the first term provides the cusp at the origin.

Electron-electron cusps, antiparallel spins: If a spin-up electron meets a spin-down electron, the wave function has singlet and triplet components in general. As long as the singlet component is non-zero, the spatial wave function need not be antisymmetric on exchange of \mathbf{r}_1 and \mathbf{r}_2 and l_0 is again 0 in the general case. Proceeding as for the electron-nucleus cusp, setting $\mu = 1/2$ and $Z = -1$, yields:

$$\Psi(\mathbf{r}) = a_{ee}^{\uparrow\downarrow} \left(1 + \frac{1}{2} r \right) + \mathbf{b}_{ee}^{\uparrow\downarrow} \cdot \mathbf{r} + \mathcal{O}(r^2). \quad (76)$$

Electron-electron cusps, parallel spins: If two electrons of the same spin meet, the spin wave function must be a triplet, implying that the spatial wave function must be antisymmetric. Only terms with odd values of l can appear in Eq. (70) and one expects l_0 to be 1 in the general case. The $l = 2$ term vanishes because it is even, so Eq. (73) becomes:

$$\begin{aligned} \Psi(\mathbf{r}) &= r \sum_{m=-1}^1 c_{1,m} Y_{1,m} \left(1 + \frac{1}{4} r \right) + \mathcal{O}(r^3) \\ &= (\mathbf{b}_{ee}^{\uparrow\uparrow} \cdot \mathbf{r}) \left(1 + \frac{1}{4} r \right) + \mathcal{O}(r^3). \end{aligned} \quad (77)$$

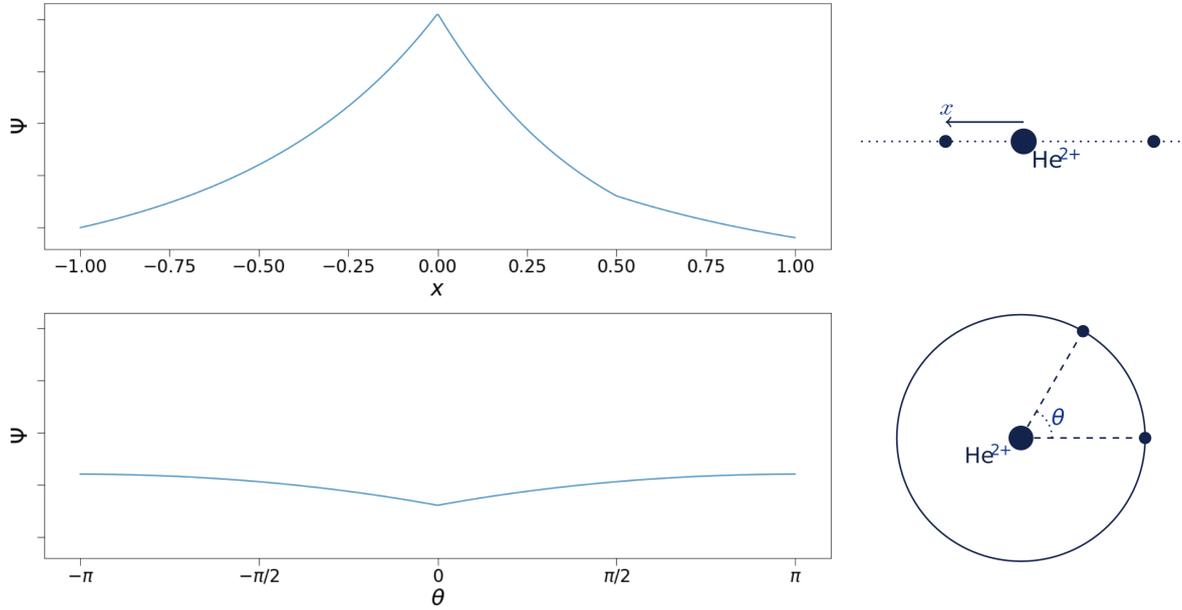


Fig. 4: *The cusps in the $S=0$ ground-state wave function of He, as learnt by a deep neural network. The cusps were not built in to the wave function in advance but discovered by the network in its attempts to minimize the total energy. According to Eqs. (75) and (76), the coefficients of the nuclear-electron and antiparallel electron-electron cusp terms should be -2 and 0.5 ; the learnt values were $-1.9979(4)$ and $0.4934(1)$. From Ref. [34].*

Figure 4 shows the electron-electron and electron-nucleus cusps for a helium atom. The wave function was represented as a deep neural network [34] (see later), the parameters of which were adjusted to minimize the variational ground-state energy. No attempt was made to force the network to generate the correct cusps; it discovered them spontaneously in its attempts to minimize the total energy. Its success in doing so confirms that the cusps have a significant effect on the total energy and adds weight to the assertion that it is a bad idea to use trial wave functions without cusps.

The Jastrow factor

We have explained that Slater determinants behave smoothly as electrons approach each other and cannot easily represent the electron-electron cusps. This failure increases the energy expectation value and slows down the convergence of configuration expansions. Cusp-related errors often limit the accuracy of otherwise well-converged FCI and coupled-cluster calculations.

A good way to add cusps to a Slater determinant is to use a Jastrow factor [35, 11, 7]. The determinant $D(x_1, x_2, \dots, x_N)$ is replaced by a Slater-Jastrow wave function,

$$\Psi_{\text{SJ}}(x_1, x_2, \dots, x_N) = e^{J(x_1, x_2, \dots, x_N)} D(x_1, x_2, \dots, x_N), \quad (78)$$

where J is a totally symmetric function of the electron coordinates. The Jastrow factor affects the normalization in a manner that is not easy to calculate, so we have made no attempt to normalize Ψ_{SJ} . Fortunately, QMC methods do not require normalized trial wave functions.

The simplest approximation assumes that

$$J(x_1, x_2, \dots, x_N) = -\frac{1}{2} \sum_{i=1}^N \sum_{\substack{j=1 \\ (j \neq i)}}^N u(x_i, x_j) \quad (79)$$

is a sum of two-electron terms. In a typical example, $u(x_i, x_j)$ increases in value as $|\mathbf{r}_i - \mathbf{r}_j|$ decreases, suppressing the value of the wave function when pairs of electrons approach each other. The determinant incorporates the antisymmetry that helps to keep like-spin electrons apart; the Jastrow factor adds a partial description of the correlations caused by the repulsive Coulomb interactions, which keep both like- and unlike-spin electrons apart.

By making sure that $u(x_i, x_j)$ has the right behavior as $|\mathbf{r}_i - \mathbf{r}_j| \rightarrow 0$, we can also use the Jastrow factor to make Ψ_{SJ} satisfy the cusp conditions. As in our previous discussion of cusps, we express Ψ_{SJ} as a function of \mathbf{r} , \mathbf{r}_{cm} , $\sigma_1, \sigma_2, x_3, \dots, x_N$, where $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ and $\mathbf{r}_{\text{cm}} = \frac{1}{2}(\mathbf{r}_1 + \mathbf{r}_2)$. We then consider how the wave function depends on \mathbf{r} at fixed $\sigma_1, \sigma_2, x_3, \dots, x_N$. Writing $\Psi(\mathbf{r}, \mathbf{r}_{\text{cm}}, \sigma_1, \sigma_2, x_3, \dots, x_N)$ as $\Psi(\mathbf{r})$ to simplify the notation, we have $\Psi_{\text{SJ}}(\mathbf{r}) = e^{-u(\mathbf{r})} D(\mathbf{r})$.

Antiparallel spins: The antiparallel cusp condition,

$$\Psi(\mathbf{r}) = a_{ee}^{\uparrow\downarrow} \left(1 + \frac{1}{2}r\right) + \mathbf{b}_{ee}^{\uparrow\downarrow} \cdot \mathbf{r} + \mathcal{O}(r^2), \quad (80)$$

can be imposed using a spherical Jastrow function, $u(\mathbf{r}) \rightarrow u(r)$.

Expanding $u(r)$ and $D(\mathbf{r})$ about the origin,

$$u(r) = u(0) + u'(0)r + \dots, \quad \text{and} \quad D(\mathbf{r}) = D(\mathbf{0}) + \mathbf{r} \cdot \nabla_{\mathbf{r}} D|_{\mathbf{r}=\mathbf{0}} + \dots, \quad (81)$$

we get

$$\begin{aligned} \Psi_{\text{SJ}}(\mathbf{r}) &= e^{-u(r)} D(\mathbf{r}) = \left(1 - u'(0)r + \dots\right) e^{-u(0)} \left(D(\mathbf{0}) + \mathbf{r} \cdot \nabla_{\mathbf{r}} D|_{\mathbf{r}=\mathbf{0}} + \dots\right) \\ &= a_{ee}^{\uparrow\downarrow} (1 - u'(0)r) + \mathbf{b}_{ee}^{\uparrow\downarrow} \cdot \mathbf{r} + \mathcal{O}(r^2), \end{aligned} \quad (82)$$

where $a_{ee}^{\uparrow\downarrow} = e^{-u(0)} D(\mathbf{0})$ and $\mathbf{b}_{ee}^{\uparrow\downarrow} = e^{-u(0)} \nabla_{\mathbf{r}} D|_{\mathbf{r}=\mathbf{0}}$.

In order to satisfy the spin-antiparallel cusp condition, all we require is that

$$\left. \frac{\partial u}{\partial r} \right|_{r=0} = -\frac{1}{2}. \quad (83)$$

Parallel spins: The parallel cusp condition,

$$\Psi(\mathbf{r}) = (\mathbf{b}_{ee}^{\uparrow\uparrow} \cdot \mathbf{r}) \left(1 + \frac{1}{4}r\right) + \mathcal{O}(r^3), \quad (84)$$

can also be imposed using a spherical Jastrow function. $D(\mathbf{0})$ is now zero, so (82) becomes

$$\Psi_{\text{SJ}}(\mathbf{r}) = \left(1 - u'(0)r + \dots\right) e^{-u(0)} \left(\mathbf{r} \cdot \nabla_{\mathbf{r}} D|_{\mathbf{r}=\mathbf{0}} + \dots\right) = (\mathbf{b}_{ee}^{\uparrow\uparrow} \cdot \mathbf{r}) (1 - u'(0)r) + \mathcal{O}(r^3), \quad (85)$$

with $\mathbf{b}_{ee}^{\uparrow\uparrow} = e^{-u(0)} \nabla_{\mathbf{r}} D|_{\mathbf{r}=\mathbf{0}}$. This has the correct cusp if

$$\left. \frac{\partial u}{\partial r} \right|_{r=0} = -\frac{1}{4}. \quad (86)$$

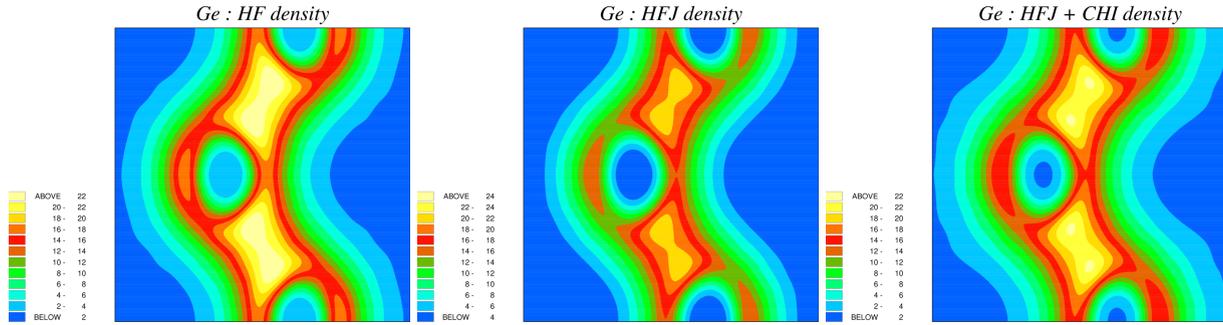


Fig. 5: The left-hand figure shows the pseudo-valence electron density of germanium as calculated using a determinant of DFT orbitals. The middle figure shows the effect of introducing a Jastrow factor containing spherical two-electron $u(|\mathbf{r}_i - \mathbf{r}_j|)$ terms. The cusps (which cannot be seen here) are improved and the energy expectation value is lowered, but the electron density is changed significantly. The right-hand figure shows the effect of including one-electron $\chi(\mathbf{r}_i)$ terms in addition to the spherical $u(|\mathbf{r}_i - \mathbf{r}_j|)$ terms. The cusps remain correct but the density is pushed back towards the original HF density, which was quite accurate.

Writing the spherical Jastrow function $u(x_1, x_2) = u(\mathbf{r}_1, \sigma_1, \mathbf{r}_2, \sigma_2)$ as $u_{\sigma_1\sigma_2}(r)$, we can summarize these results as follows:

$$\left. \frac{\partial u_{\sigma_1\sigma_2}(r)}{\partial r} \right|_{r=0} = \begin{cases} -\frac{1}{2}, & \sigma_1 = -\sigma_2, \\ -\frac{1}{4}, & \sigma_1 = \sigma_2. \end{cases} \quad (87)$$

If the one-electron orbitals are expanded in a basis of smooth functions such as plane waves or Gaussians, it is often best to incorporate the electron-nucleus cusps into the Jastrow factor too. This requires adding terms dependent on $|\mathbf{r}_i - \mathbf{d}_I|$ to J , where \mathbf{d}_I is the position of nucleus I . The large- r behavior of $u(r)$ in solids can be derived within the random phase approximation [35]. The result is that $u(r) \sim 1/\omega_p r$ as $r \rightarrow \infty$, where n is the average electron density and $\omega_p = \sqrt{4\pi n}$ is the plasma frequency of a uniform electron gas of that density.

The Jastrow function does not have to be pairwise or spherical. We can, for example, add any smooth function of \mathbf{r} to the spherical pairwise term $u(r)$ without affecting the cusps. We can also add to J a totally symmetric one-electron contribution of the form $\sum_i \chi(x_i)$, which can provide a convenient way to optimize the electron spin density $n(x) = n(\mathbf{r}, \sigma)$. An example is shown in Fig. 5. Finally, we can add terms that depend on the positions of more than two charged particles. The usual practice in QMC simulations [36] is to choose a fairly general parametrized Jastrow factor incorporating the cusps and adjust the parameters to minimize the energy expectation value.

Although one-determinant Slater-Jastrow wave functions do not achieve chemical accuracy, they are easy to use and often account for 80–90% or more of the correlation energy missed by Hartree-Fock theory. The $\mathcal{O}(N^3)$ system-size scaling of Slater-Jastrow based variational QMC simulations is favorable enough that they can be used to study periodic supercells containing of order 1000 electrons.

When studying small molecules, it is straightforward to carry out Slater-Jastrow QMC simulations with large linear combinations of Slater determinants, all multiplied by a Jastrow factor,

thus accounting for static correlation as well as dynamic correlation. Dealing with static correlation in solids is an unsolved problem in general, as the number of configurations required rises exponentially with system size, but one-determinant Slater-Jastrow VMC calculations are often used and often produce surprisingly good results. Jastrow factors are not so easy to combine with conventional quantum chemical methods such as coupled cluster, but alternative “R12” approaches are available [37].

6 Beyond Slater determinants

Attempting to describe the wave functions of solids as linear combinations of Slater determinants, with or without a Jastrow factor, is a futile task because the number of determinants required rises exponentially with system size. As a result, the vast majority of solid-state QMC simulations have used single-determinant trial wave functions. Single-determinant VMC results are not very much more accurate than DFT, but DMC simulations [11, 7] with one-determinant Slater-Jastrow trial wave functions have produced much of the most accurate data available for weakly correlated bulk solids, including the electron gas data used to parametrize the local density approximation. Until recently, it has been difficult to ascertain the quality of DMC results for solids because experiments are of limited accuracy and no better methods were known, but auxiliary-field QMC (which also requires a trial wave function) is now producing slightly better results for some molecules and solids and coupled-cluster approaches are making progress.

Given the limitations of multi-determinant expansions, it is no surprise that efforts have been made to find better starting points. An old idea that still has value is the backflow transformation introduced by Richard Feynman in 1954 [38]. Feynman’s approach was inspired by the way a classical fluid flows around an obstruction and his application was to liquid ^4He , but the idea is general enough to work for electrons too. It was first used in a QMC simulation of the interacting electron gas in 1994 [39].

In a backflow wave function, the electron positions in the Slater determinant are replaced by “quasi-particle” coordinates that depend on the positions of other nearby electrons

$$\mathbf{q}_i = \mathbf{r}_i + \boldsymbol{\xi}_i(x_1, x_2, \dots, x_N) = \mathbf{r}_i + \sum_{j (\neq i)} \eta_{ij} \mathbf{r}_{ij}, \quad (88)$$

where $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ and η_{ij} depends on $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ and on the relative spins of electrons i and j . The Slater matrix is otherwise unchanged, but every entry in every row and column now depends on the coordinates of every electron, which slows down QMC calculations by a factor of N . Although the backflow wave function is still a determinant, it is *not* a Slater determinant and cannot be expanded as a linear combination of a small number of Slater determinants; it is something new.

One advantage of the use of parametrized backflow transformations is that they provide a convenient way to adjust the *nodal surface* of the trial wave function. Given a choice of the spin coordinates $(\sigma_1, \sigma_2, \dots, \sigma_N)$, the nodal surface is the $(3N-1)$ -dimensional surface in the $3N$ -dimensional space of positions $(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ on which the wave function is zero. It matters

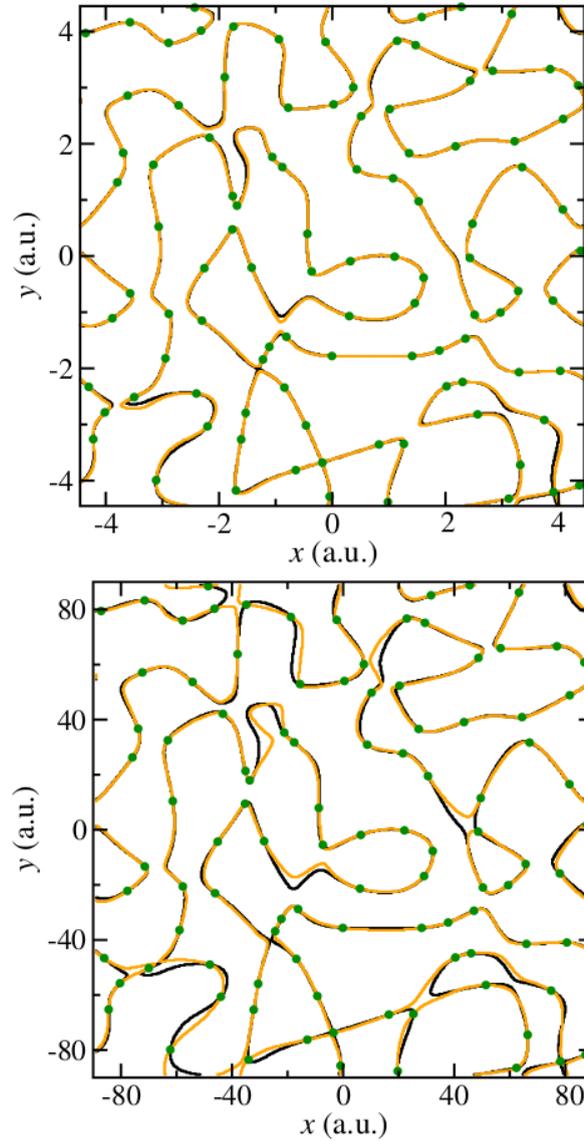


Fig. 6: Nodes encountered when moving one of the electrons in a two-dimensional homogeneous gas of 101 like-spin electrons. The positions of the other 100 electrons, indicated by the green circles, are held fixed. The Hartree-Fock and backflow nodes are in black and orange, respectively. The top panel shows results for a weakly correlated electron gas with density parameter $r_s = 0.5$; the bottom panel shows results for a less dense and more strongly correlated electron gas with $r_s = 10$. From Ref. [40].

because the quality of the nodal surface is the only factor that limits the quality of DMC results: if the nodal surface is exact, DMC gives the exact ground-state energy. Figure 6 shows how the nodes of an optimized backflow wave function differ from those of the Hartree-Fock determinant [40]. The differences are subtle but improve the quality of the results substantially. As far as can be ascertained, energies calculated using backflow DMC simulations of electron gases at densities comparable to those found in most solids are almost exact. Results for light atoms, where we know the ground-state energy almost exactly, are somewhat less impressive but Fig. 7 shows that backflow remains useful, reducing the error in the total energy of a Slater-Jastrow wave function by more than 50%.

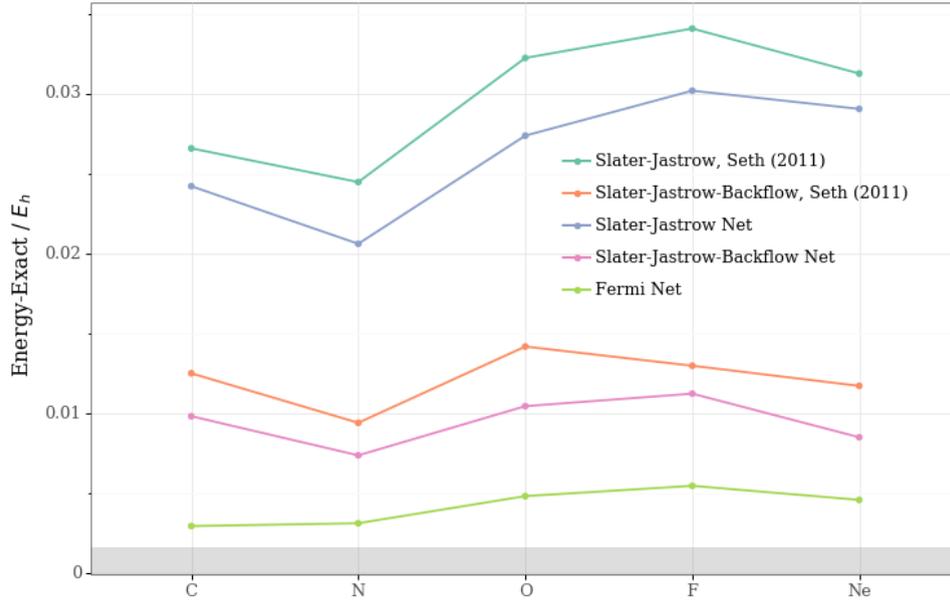


Fig. 7: Errors in the energies of various atoms as calculated using a variety of different one-determinant VMC trial wave functions. Hartree atomic units are used. The data used to plot the cyan and orange lines comes from Ref. [41]; the other three lines show data from Ref. [34].

Backflow is a good idea but we can take it much further [34]. Nothing requires the orbitals in a Slater determinant to be functions of the coordinates of a single electron only, nor need they be functions of a single three-dimensional (plus spin) quasi-particle coordinate, as in a backflow wave function. The only requirement is that interchanging any two coordinates, x_i and x_j , exchanges the corresponding columns of the determinant and thus changes the sign of the wave function. This freedom allows us to replace the one-electron orbitals $\varphi_i(x_j)$ by multi-electron functions of the form

$$\varphi_i(x_j; x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_N) = \varphi_i(x_j; \{x_{/j}\}), \quad (89)$$

where $x_{/j}$ is shorthand for all of the coordinates except x_j . As long as $\varphi_i(x_j; \{x_{/j}\})$ is invariant under any change in the order of the arguments after x_j , the resulting wave function,

$$D = \begin{vmatrix} \varphi_1(x_1, \{x_{/1}\}) & \varphi_1(x_2, \{x_{/2}\}) & \dots & \varphi_1(x_N, \{x_{/N}\}) \\ \varphi_2(x_1, \{x_{/1}\}) & \varphi_2(x_2, \{x_{/2}\}) & \dots & \varphi_2(x_N, \{x_{/N}\}) \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \varphi_N(x_1, \{x_{/1}\}) & \varphi_N(x_2, \{x_{/2}\}) & \dots & \varphi_N(x_N, \{x_{/N}\}) \end{vmatrix}, \quad (90)$$

is totally antisymmetric.

Surprisingly, it can be shown that *any* totally antisymmetric wave function can be represented as a *single* generalized determinant of this type [34]. The proof does not explain how to construct the generalized determinant we need, but the fact that it exists is reassuring: the exponential wall that makes expanding many-electron wave functions in conventional Slater determinants so difficult might not apply when generalized determinants are used. The NP hardness of the

many-electron problem is bound to rear its ugly head somewhere, but it is going to take a different form and the exponential rise in difficulty with system size might be less steep.

Trial wave functions of the type introduced in Eq. (90) are so general that it is hard to see how to parametrize them. This is a common problem with trial wave function construction: the more general the functional form, the more parameters you need and the less you know about which ones to choose. Until a few years ago, this would have ruled out using generalized determinants, but recent progress in machine learning [42, 43] has made it much easier to find parametrized representations of extremely complicated functions without having to choose the functional form or parameters explicitly. Deep neural networks [44] are general function approximators, able in principle to represent any function in any number of dimensions. Furthermore, via the magic of automatic differentiation and back propagation, optimizing the parameters that define a neural network is remarkably efficient.

In Ref. [34], generalized determinants were represented using a new neural network architecture, the *Fermi Net*, designed to guarantee the necessary exchange symmetries but otherwise being very general. The method used was VMC, conventional in all respects except for the wave function. The Fermi Net *was* the wave function and was used to calculate the value and first and second derivatives of the wave function at arbitrary points in coordinate space. This is all that is required to implement the Metropolis algorithm and allow the calculation of the local energy. The parameters of the network were adjusted to minimize the energy expectation value, exactly as prescribed by the variational principle. Learning from the variational principle differs from the more frequently encountered concept of learning from data; the Fermi-Net optimization generates its own input on the fly and can never run out of training data.

Using neural networks to represent wave functions is a fashionable idea and several other approaches are being explored [45–47]. Here I concentrate on Fermi Net because I played a minor role in helping to develop it and because it is in some way the most accurate approach proposed so far. The field is so young, however, that I would not be surprised to see better approaches come along soon.

The Fermi Net takes electron spins σ_i , positions \mathbf{r}_i , the vectors between electrons $\mathbf{r}_i - \mathbf{r}_j$, and electron-nucleus vectors $\mathbf{r}_i - \mathbf{d}_I$ as input. The network is only capable of representing smooth analytic functions of position (this is on purpose; the VMC method runs into difficulties if the gradient or value of the trial wave function is discontinuous), so it is unable to represent the cusps exactly. To circumvent this problem, the distances $|\mathbf{r}_i - \mathbf{r}_j|$ and $|\mathbf{r}_i - \mathbf{d}_I|$ are also provided as inputs. The distance function $|\mathbf{r}|$ has its own cusp at the origin, enabling the network to represent the electron-electron and electron-nucleus cusps as smooth functions of the interparticle distances. As was shown in Fig. 4, it accomplishes this very effectively. Note that the Fermi-Net wave function takes positions as input and returns values of the wave function as output. No one-electron or many-electron basis set is required.

Although this approach to machine learning wave functions has only been applied to atoms and small molecules to date, the results have been spectacularly good. Fermi-Net wave functions are clearly much better than any other known type of VMC trial wave function used in such systems.

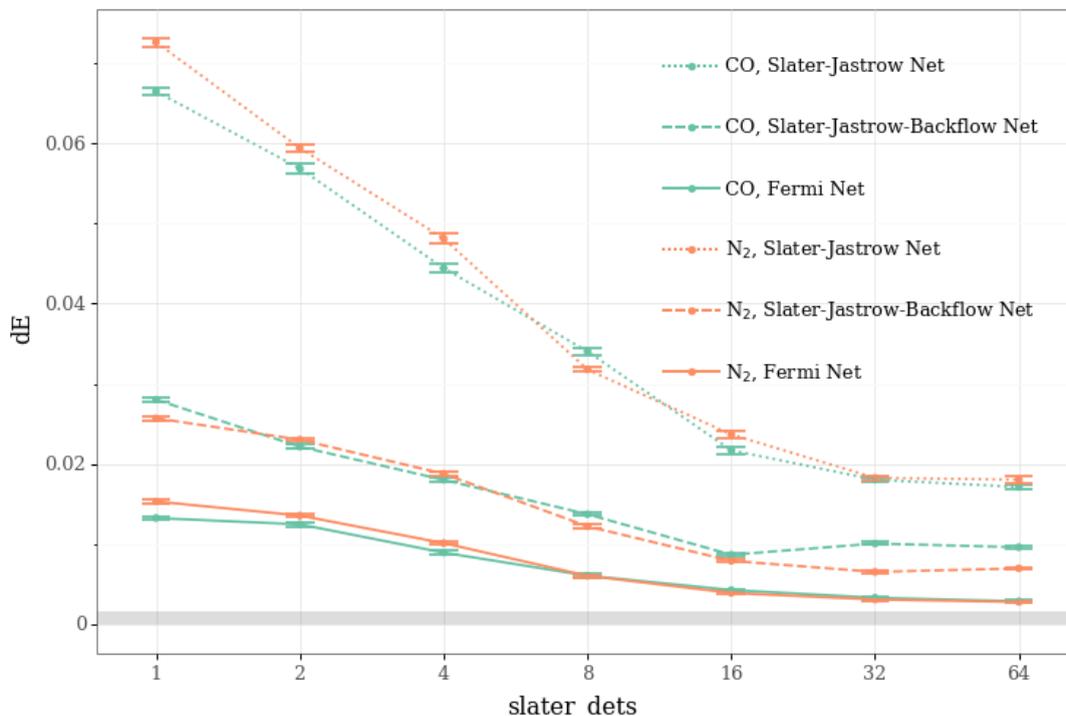


Fig. 8: Convergence with the number of determinants of the total energy (in Hartrees) of the CO and N₂ molecules. The Slater-Jastrow, Slater-Jastrow-backflow, and Fermi-Net wave functions were all represented as neural networks. From Ref. [34].

Figure 7 compares the accuracies of several different one-determinant VMC methods. The dark blue and cyan lines show results obtained with a single-determinant Slater-Jastrow wave function. They differ only because the dark blue results were generated using a neural-network representation of the Slater-Jastrow wave function whereas the cyan line used an explicitly parametrized functional form [41]. The generality and freedom inherent in the neural network allows it to slightly outperform the conventional implementation. The orange and pink lines are related in the same way. Both show results for a single-determinant Slater-Jastrow-backflow wave function, but the orange line used a conventional representation and the pink line a neural network representation. The green line shows the Fermi-Net results. Even though only one determinant was used, they are quite close to chemical accuracy indicated by the grey bar.

Figure 8 illustrates the convergence of the total energy of two molecules, CO and N₂, as a function of the number of determinants used. All results were obtained using neural-network wave functions of the corresponding type, so these are fair comparisons. The results obtained with 16 Fermi-Net determinants are close to chemical accuracy. From now on, all of the calculations reported used 16 Fermi-Net determinants and a neural network with approximately 700,000 variational parameters. This may sound extreme, but the number of amplitudes required for comparably accurate CCSD(T) calculations is even larger.

Figure 9 shows results for various molecules with up to 30 electrons. All of the methods investigated are less accurate for larger systems, but the Fermi Net is again the best and the growth of the Fermi Net errors with system size appears to be more systematic.

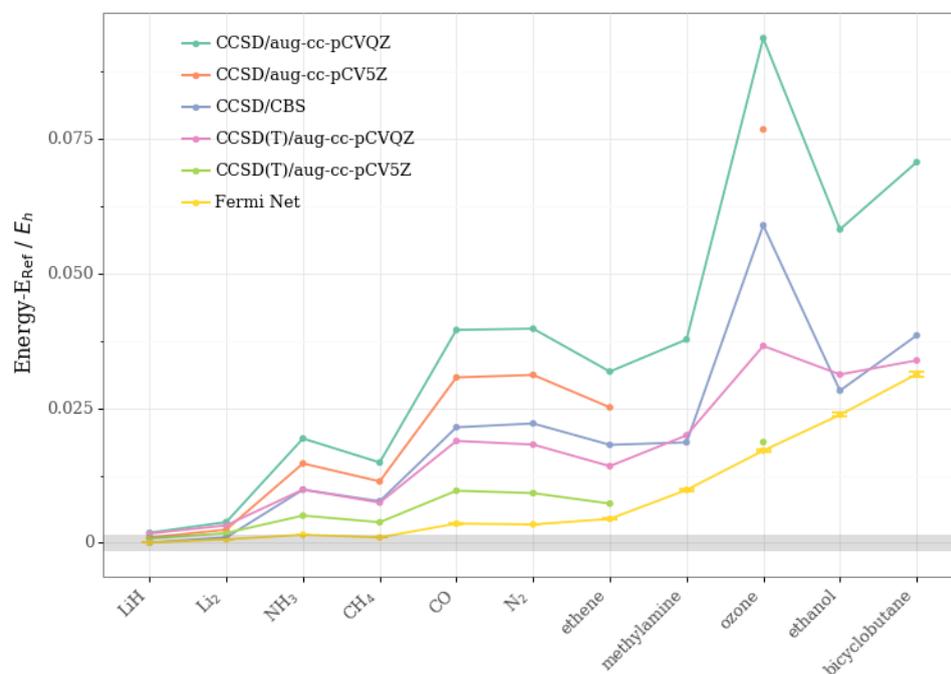


Fig. 9: Errors in the total energies (in Hartrees) of various molecules with up to 30 electrons calculated using Fermi Net, CCSD, and CCSD(T). High-quality QZ and 5Z basis sets were used for the coupled-cluster calculations. From Ref. [34].

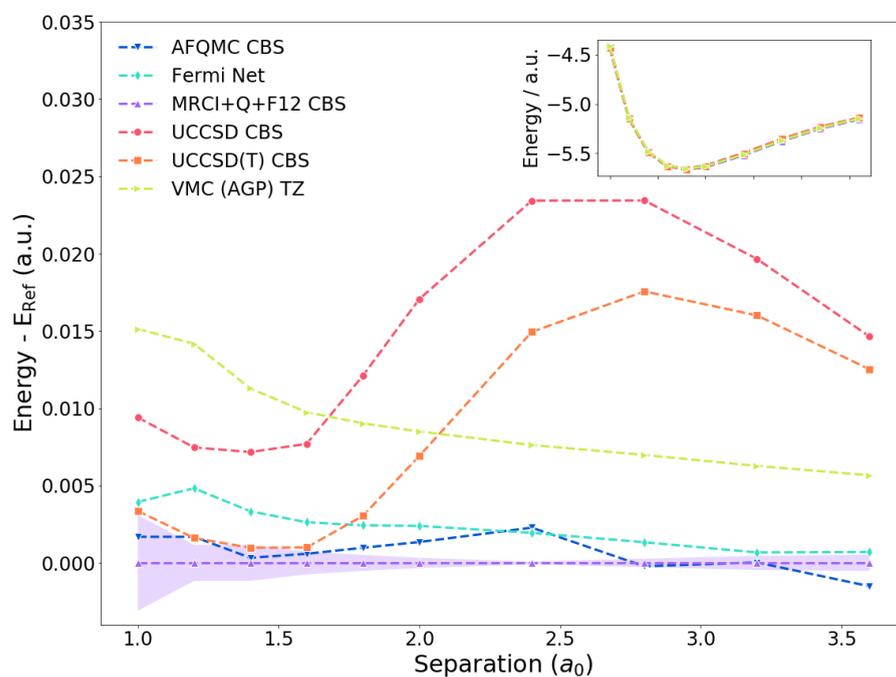


Fig. 10: Energy (in Hartrees) of the H_{10} chain calculated using a wide variety of different methods for a range of inter-atomic distances (in Bohr radii). All energies are measured relative to the MRCI + Q + F12 CBS energy. The mauve shaded region indicates the estimated uncertainty in the reference result. Fermi Net results from Ref. [34]; all other results from Ref. [48].

The H_{10} chain was the subject of an important recent benchmark paper [48] comparing practically every accurate method known at the time. Figure 10 shows the total energy as a function of the inter-atomic spacing. At each spacing, all energies are measured relative to a reference obtained by extrapolating the MRCI + Q + F12 result to the complete basis set (CBS) limit. The mauve shaded region indicates the estimated uncertainty in the reference energy. All of the data except for the Fermi-Net results came from Ref. [48]. Despite the newness and conceptual simplicity of our neural-net based approach, it comfortably outperforms most of its rivals.

References

- [1] M. Troyer and U.J. Wiese, *Phys. Rev. Lett.* **94**, 170201 (2005)
- [2] S. Azadi, B. Monserrat, W.M.C. Foulkes, and R.J. Needs, *Phys. Rev. Lett.* **112**, 165501 (2014)
- [3] N.D. Drummond, R.J. Needs, A. Sorouri, and W.M.C. Foulkes, *Phys. Rev. B* **78**, 125106 (2008)
- [4] R.G. Parr and W. Yang: *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, 1989)
- [5] J. Kohanoff: *Electronic Structure Calculations for Solids and Molecules: Theory and Computational Methods* (Cambridge University Press, 2006)
- [6] R.M. Martin: *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, 2008)
- [7] R.M. Martin, L. Reining, and D.M. Ceperley: *Interacting Electrons: Theory and Computational Approaches* (Cambridge University Press, 2016)
- [8] X.W. Guan, M.T. Batchelor, and C. Lee, *Rev. Mod. Phys.* **85**, 1633 (2013)
- [9] J. Bardeen, L.N. Cooper, and J.R. Schrieffer, *Phys. Rev.* **108**, 1175 (1957)
- [10] R.B. Laughlin in R.E. Prange and S.M. Girvin (eds.): *The Quantum Hall Effect* (Springer, New York, 1990), chap. 7, p. 233.
- [11] W.M.C. Foulkes, L. Mitas, R.J. Needs, and G. Rajagopal, *Rev. Mod. Phys.* **73**, 33 (2001)
- [12] S. Zhang in E. Pavarini, E. Koch, and S. Zhang (eds.): *Many-Body Methods for Real Materials* (Forschungszentrum Jülich, 2019), Ch. 6
www.cond-mat.de/events/correl19/manuscripts/correl19.pdf
- [13] M. Casula and S. Sorella, *J. Chem. Phys.* **119**, 6500 (2003)
- [14] M. Bajdich, L. Mitas, L.K. Wagner, and K.E. Schmidt, *Phys. Rev. B* **77**, 115112 (2008)
- [15] C. Genovese, T. Shirakawa, K. Nakano, and S. Sorella, arXiv:2002.03347 (2020)
- [16] U. Schollwöck, *Rev. Mod. Phys.* **77**, 259 (2005)
- [17] R. Orús, *Ann. Phys.* **349**, 117 (2014)
- [18] N. Ashcroft and N.D. Mermin: *Solid State Physics* (Brooks Cole, 1976)
- [19] J.W. Negele and H. Orland: *Quantum Many-Particle Systems* (Perseus, 1998)

- [20] T. Helgaker, P. Jorgensen, and J. Olsen: *Molecule Electronic-Structure Theory* (Wiley, 2014)
- [21] R.Q. Hood, M.Y. Chou, A.J. Williamson, G. Rajagopal, R.J. Needs, and W.M.C. Foulkes, *Phys. Rev. Lett.* **78**, 3350 (1997)
- [22] G.H. Booth, A.J.W. Thom, and A. Alavi, *J. Chem. Phys.* **131**, 054106 (2009)
- [23] D. Cleland, G.H. Booth, and A. Alavi, *J. Chem. Phys.* **132**, 041103 (2010)
- [24] G.H. Booth, A. Grüneis, G. Kresse, and A. Alavi, *Nature* **493**, 365 (2013)
- [25] J.S. Spencer, N.S. Blunt, and W.M.C. Foulkes, *J. Chem. Phys.* **136**, 054110 (2012)
- [26] E. Giner, A. Scemama, and M. Caffarel, *Can. J. Chem.* **91**, 879 (2013)
- [27] F.A. Evangelista, *J. Chem. Phys.* **140**, 124114 (2014)
- [28] N.M. Tubman, J. Lee, T.Y. Takeshita, M. Head-Gordon, and K.B. Whaley, *J. Chem. Phys.* **145**, 044112 (2016)
- [29] A.A. Holmes, N.M. Tubman, and C.J. Umrigar, *J. Chem. Theory Comput.* **12**, 3674 (2016)
- [30] I. Shavitt and R.J. Bartlett: *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory* (Cambridge University Press, 2009)
- [31] W. Klopper and D.P. Tew: *Coupled cluster theory: Fundamentals* (2006)
www.ipc.kit.edu/theochem/download/Kapitel13.pdf
- [32] T. Kato, *Commun. Pure Appl. Math.* **10**, 151 (1957)
- [33] R.T. Pack and W.B. Brown, *J. Chem. Phys.* **45**, 556 (1966)
- [34] D. Pfau, J.S. Spencer, A.G.d.G. Matthews, and W.M.C. Foulkes, arXiv:1909.02487 (2019)
- [35] D.M. Ceperley, *Phys. Rev. B* **18**, 3126 (1978)
- [36] P. López-Ríos, P. Seth, N.D. Drummond, and R.J. Needs, *Phys. Rev. E* **86**, 036703 (2012)
- [37] W. Klopper, F.R. Manby, S. Ten-No, and E.F. Valeev, *Int. Rev. Phys. Chem.* **25**, 427 (2007)
- [38] R.P. Feynman, *Phys. Rev.* **94**, 262 (1954)
- [39] Y. Kwon, D.M. Ceperley, and R.M. Martin, *Phys. Rev. B* **48**, 12037 (1993)
- [40] P. López Ríos, A. Ma, N.D. Drummond, M.D. Towler, and R.J. Needs, *Phys. Rev. E* **74**, 066701 (2006)
- [41] P. Seth, P. López-Ríos, and R.J. Needs, *J. Chem. Phys.* **134**, 084105 (2011)

-
- [42] A. Krizhevsky, I. Sutskever, and G.E. Hinton: In *NIPS* (2012), Vol. 25, pp. 1097–1105
- [43] D. Silver, A. Huang, C.J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot et al., *Nature* **529**, 485 (2016)
- [44] I. Goodfellow, Y. Bengio, and A. Courville: *Deep Learning* (MIT Press, 2017)
- [45] D. Luo and B.K. Clark, *Phys. Rev. Lett.* **122**, 226401 (2019)
- [46] J. Hermann, Z. Schätzle, and F. Noé, arXiv:1909.08423 (2019)
- [47] K. Choo, A. Mezzacapo, and G. Carleo, *Nat. Commun.* **11**, 2368 (2020)
- [48] M. Motta, D.M. Ceperley, G.K.L. Chan, J.A. Gomez, E. Gull, S. Guo, C.A. Jiménez-Hoyos, T.N. Lan, J. Li, F. Ma, A.J. Millis, N.V. Prokof'ev, U. Ray, G.E. Scuseria, S. Sorella, E.M. Stoudenmire, Q. Sun, I.S. Tupitsyn, S.R. White, D. Zgid, and S. Zhang, *Phys. Rev. X* **7**, 031059 (2017)