

# Neural network variational Monte Carlo for quantum phase transitions and multi-component systems

Imperial College London

Department of Physics

Submitted for the degree of Doctor of Philosophy (PhD)

Gino William Cassella

# Statement of originality

All work contained in this thesis is the original work of the author. All quotations or excerpts from other works are clearly indicated in the text, alongside appropriate references. Chapters 4 and 5 describe results which have previously been published by the author. Where figures and tables are closely reproduced from the published works, explicit references to those published works are included in accordance with the policies of the respective publications. These policies can be found online at:

- <https://journals.aps.org/copyrightFAQ.html> (chapter 4)

*[...] the author has the right to use the article or a portion of the article in a thesis or dissertation without requesting permission from APS, provided the bibliographic citation and the APS copyright credit line are given on the appropriate pages.*

- <https://www.nature.com/nature-portfolio/reprints-and-permissions/permissions-requests>

*Authors have the right to reuse their article's Version of Record, in whole or in part, in their own thesis. Additionally, they may reproduce and make available their thesis, including Springer Nature content, as required by their awarding academic institution.*

Moreover, both works are published 'open access' under the Creative Commons Attribution 4.0 International License, which "permits unrestricted use, distribution, and reproduction in any medium, provided attribution to the author(s) and the published article's title, journal citation, and DOI are maintained".

# Copyright declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

# Abstract

Computational solutions of the time-independent Schrödinger equation are now a commodity research tool in the natural sciences. However, advancing our understanding of novel states of matter requires the development of methods with greater accuracy, efficiency, or applicability to systems where current approaches fall short.

Recently, deep learning has emerged as a revolutionary paradigm in the computational sciences, driving state-of-the-art developments in computer vision and natural language processing. Now, a race is afoot to extend this revolution to quantum chemistry: neural network representations of many-body electronic wavefunctions, combined with the variational Monte Carlo (VMC) method, have been shown to produce highly accurate solutions to the time-independent Schrödinger equation. Still a nascent field, neural network VMC (NNVMC) promises an ideal tradeoff between accuracy and asymptotic efficiency.

This thesis explores novel advantages of NNVMC, enabled by the lack of dependence on a set of basis functions. We study two applications of the method: totally *ab initio* treatment of quantum phase transitions in solid-state systems, exemplified by the Wigner transition in the homogeneous electron gas; and the accurate description of multicomponent systems, exemplified by positronic bound states of ordinary molecules, which are challenging to describe within electronic basis sets. We demonstrate that the accuracy of our method rivals or exceeds, with minimal inductive bias, existing state-of-the-art methods.

Finally, we unify these themes to study nuclear quantum effects without resorting to the Born-Oppenheimer approximation. Our proof-of-principle suggests that neural network trial wavefunctions offer a route to an unambiguous determination of the phases of solid hydrogen at high pressure. More generally, we believe that NNVMC offers a universal method, and expect that neural-network-based methods will become a powerful tool in the study of systems where the fundamental physics is not yet fully understood.

# Acknowledgements

I would like to offer my deepest gratitude to my supervisor, Professor Matthew Foulkes, for his compassionate guidance, straightforward insight, and the generous donation of a (small portion) of his scientific taste. I am also deeply indebted to my co-supervisors, Dr James Spencer and Dr David Pfau, for patient mentorship in technical matters and the realities of modern science. I cannot thank these three enough for providing me with the opportunities which set me upon my current path.

I am fortunate to have had many other great mentors over the years; I am indebted to Dr Gøran Nilsen (and the rest of the gang at ISIS, particularly Mark Devonport and Dr Ross Stewart), Professor Johannes Knolle, and Dr Anton Souslov.

I would also like to offer thanks to all the following people.

To all the students who have passed through the CMTH group at Imperial, who were true comrades. In no particular order, to Chris, Tom, Milan, Tonny, Halvard, Mathieu, Krishnan, David, Emanuele, Ignacio, Andres, and Henry. I reserve special thanks for Tom Hodson and Peru d'Ornellas, who showed me great compassion in some of the hardest days of my life.

To my parents, for constant encouragement, advice, and the occasional subsidy of beer money.

To the boys, for lifting my spirits. Especially Chris, Dan S, Max, and Dan C. Coming home to a house filled with good friends meant the world to me.

To Thierry and Nathaniel, without whom I would never have developed the oversized ego required to make it this far.

Finally, to Anna, who always encourages the best in me, and puts up with the worst.

For Grandad. I know how a capacitor works now.

# Contents

|          |                                                                 |           |
|----------|-----------------------------------------------------------------|-----------|
| <b>1</b> | <b>Introduction</b>                                             | <b>11</b> |
| <b>2</b> | <b>Quantum chemistry and electronic structure theory</b>        | <b>14</b> |
| 2.1      | Hartree-Fock theory . . . . .                                   | 16        |
| 2.1.1    | The Hartree-Fock equations . . . . .                            | 17        |
| 2.1.2    | Numerical Hartree-Fock theory . . . . .                         | 19        |
| 2.1.3    | Post-Hartree-Fock theory . . . . .                              | 21        |
| 2.2      | Density functional theory . . . . .                             | 25        |
| 2.2.1    | The Hohenberg-Kohn theorems . . . . .                           | 25        |
| 2.2.2    | Kohn-Sham DFT . . . . .                                         | 29        |
| 2.3      | Quantum Monte Carlo . . . . .                                   | 32        |
| <b>3</b> | <b>Variational Monte Carlo and neural network wavefunctions</b> | <b>33</b> |
| 3.1      | Variational Monte Carlo . . . . .                               | 33        |
| 3.1.1    | Energy evaluation . . . . .                                     | 34        |
| 3.1.2    | Wavefunction optimization . . . . .                             | 36        |
| 3.1.3    | Diffusion Monte Carlo . . . . .                                 | 39        |
| 3.2      | The Slater-Jastrow(-Backflow) Ansatz . . . . .                  | 41        |
| 3.2.1    | Slater determinants and fermionic anti-symmetry . . . . .       | 41        |
| 3.2.2    | Jastrow factors and the Kato cusp conditions . . . . .          | 44        |
| 3.2.3    | The backflow transformation . . . . .                           | 45        |
| 3.3      | Neural network variational Monte Carlo . . . . .                | 46        |
| 3.3.1    | Deep neural networks . . . . .                                  | 47        |
| 3.3.2    | Neural quantum states . . . . .                                 | 49        |
| 3.3.3    | Continuum neural network wavefunctions . . . . .                | 49        |
| 3.3.4    | FermiNet and PsiFormer . . . . .                                | 52        |
| 3.3.5    | Implementation of NNVMC . . . . .                               | 62        |
| 3.3.6    | Applications of NNVMC . . . . .                                 | 70        |
| <b>4</b> | <b>The homogeneous electron gas</b>                             | <b>79</b> |
| 4.1      | Periodic boundary conditions . . . . .                          | 80        |

|          |                                                             |            |
|----------|-------------------------------------------------------------|------------|
| 4.1.1    | Many-body eigenstates under PBCs . . . . .                  | 83         |
| 4.1.2    | Implementing PBCs in NNVMC . . . . .                        | 86         |
| 4.2      | Results and discussion . . . . .                            | 91         |
| 4.2.1    | Unpolarized simple cubic 14-HEG . . . . .                   | 92         |
| 4.2.2    | Polarized body-centered cubic 27-HEG . . . . .              | 94         |
| 4.3      | Related work and future directions . . . . .                | 97         |
| 4.4      | Conclusion . . . . .                                        | 102        |
| <b>5</b> | <b>Positronic chemistry</b>                                 | <b>103</b> |
| 5.1      | Traditional approaches to positron binding . . . . .        | 105        |
| 5.2      | Multicomponent calculations in NNVMC . . . . .              | 108        |
| 5.3      | Results . . . . .                                           | 109        |
| 5.3.1    | Positron binding energies . . . . .                         | 110        |
| 5.3.2    | Annihilation rates . . . . .                                | 116        |
| 5.4      | Discussion and conclusion . . . . .                         | 119        |
| <b>6</b> | <b>Nuclear quantum effects</b>                              | <b>121</b> |
| 6.1      | The Born-Oppenheimer approximation redux . . . . .          | 122        |
| 6.2      | Quantum nuclei in NNVMC . . . . .                           | 128        |
| 6.3      | Results and discussion . . . . .                            | 131        |
| 6.3.1    | The hydrogen dimer . . . . .                                | 132        |
| 6.3.2    | Small molecules . . . . .                                   | 135        |
| 6.3.3    | Eight protons in a box . . . . .                            | 137        |
| 6.4      | Conclusion . . . . .                                        | 143        |
| <b>7</b> | <b>Conclusion</b>                                           | <b>145</b> |
|          | <b>Bibliography</b>                                         | <b>149</b> |
| <b>A</b> | <b>Notational conventions</b>                               | <b>170</b> |
| <b>B</b> | <b>Monte Carlo integration and the Metropolis algorithm</b> | <b>172</b> |
| <b>C</b> | <b>Gradient of the energy functional</b>                    | <b>175</b> |
| <b>D</b> | <b>Variance matching procedure for benzene</b>              | <b>178</b> |
| <b>E</b> | <b>Data tables</b>                                          | <b>180</b> |

# List of Figures

|     |                                                                                                              |     |
|-----|--------------------------------------------------------------------------------------------------------------|-----|
| 3.1 | Schematic of the FermiNet input featurization. . . . .                                                       | 53  |
| 3.2 | Schematic of the FermiNet architecture. . . . .                                                              | 56  |
| 3.3 | Schematic depiction of the PsiFormer architecture. . . . .                                                   | 61  |
| 4.1 | Ground-state energy and order parameter for 27-HEG. . . . .                                                  | 95  |
| 4.2 | Comparison of the ground-state energy of 14-HEG with related works. . . . .                                  | 101 |
| 5.1 | Potential energy surface of bare LiH molecule and its positronic complex. . . . .                            | 112 |
| 5.2 | Properties of the positronic bound state of BeO with varying inter-atomic separation. . . . .                | 114 |
| 5.3 | Orthographic projections of the ground-state density for positronic dilithium and benzene molecules. . . . . | 115 |
| 6.1 | Protonic density of the orthohydrogen dimer. . . . .                                                         | 133 |
| 6.2 | Two-particle protonic pair density of the parahydrogen and orthohydrogen dimers. . . . .                     | 134 |
| 6.3 | Protonic density of water, ammonia, and methane. . . . .                                                     | 136 |
| 6.4 | Protonic density as a function of Wigner-Seitz radius for the 8-protons-in-a-box system. . . . .             | 138 |
| 6.5 | Nearest-neighbour separation as a function of Wigner-Seitz radius for the 8-protons-in-a-box-system. . . . . | 140 |
| 6.6 | Electronic and structural properties of eight-protons-in-a-box as a function of Wigner-Seitz radius. . . . . | 141 |
| D.1 | Demonstration of variance-matching procedure for positronic benzene. . . . .                                 | 179 |

# List of Tables

|     |                                                                                              |     |
|-----|----------------------------------------------------------------------------------------------|-----|
| 3.1 | Summary of symbols and parameters used in defining the FermiNet architecture. . . . .        | 57  |
| 3.2 | Summary of symbols and parameters used in defining the PsiFormer architecture. . . . .       | 60  |
| 3.3 | Summary of NNVMC hyperparameters. . . . .                                                    | 68  |
| 4.1 | Ground-state energy benchmarks for 14-HEG. . . . .                                           | 93  |
| 5.1 | Ground-state energy comparison for LiH and its positronic bound state.                       | 111 |
| 5.2 | Ground-state energy comparison for BeO and its positronic bound state. . . . .               | 112 |
| 5.3 | Comparison of $2\gamma$ annihilation rates between computational methods.                    | 118 |
| 6.1 | Ground-state properties of small molecules containing quantum protons.                       | 135 |
| E.1 | Ground-state energy as a function of bondlength for bare and positronic LiH. . . . .         | 180 |
| E.2 | Ground-state energy as a function of bondlength for bare and positronic BeO. . . . .         | 181 |
| E.3 | Ground-state energies and variances used in variance matching for benzene. . . . .           | 182 |
| E.4 | Ground-state energy as a function of Wigner-Seitz radius for eight protons in a box. . . . . | 183 |

# Chapter 1

## Introduction

“The fundamental laws necessary for the mathematical treatment of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation.”

*P.A.M. Dirac, 1929 [1]*

Nearly a century after Dirac’s declaration of victory in the fundamental treatment of atomic physics by the laws of quantum mechanics, we are still heeding his call to action. In the intervening years, applied quantum mechanics – under the guise of quantum chemistry and electronic structure theory – has become a vast field of study. Today, any scientist with access to a reasonably modern personal computer can use free and open software to solve the laws of quantum mechanics for modest atomic systems ‘without too much computation’.

The main features of atoms and small molecules are now thoroughly understood. However, our appetite to understand larger and more complex atomic systems continues to grow. Physicists push the boundaries of our understanding of challenging strongly correlated electron systems like high-temperature superconductors and exotic topological matter; chemists strive to unravel complex catalytic processes and develop advanced energy materials to meet the challenges of the modern world; and biologists seek to deepen our control of fundamental biological processes to revolutionize medicine and healthcare. All of these pursuits exist at the frontier of our ability to (approximately) solve the laws of quantum mechanics for complex atomic systems.

At the time of writing, the best methods for producing accurate solutions to the

laws of quantum mechanics exhibit a rapid asymptotic increase of computational cost with system size that hinders their application to cutting edge problems. At the other end of the spectrum, methods that offer the most palatable scaling with system size do not treat the correlated motion of electrons in atomic matter with sufficient detail to describe the most complex systems. Thus, expanding the size and complexity of quantum systems we can treat requires the development of new algorithms that push the frontiers of efficiency and accuracy.

Developing flexible computational methods with a very high degree of quantitative accuracy remains a necessity for (even *qualitatively*) understanding the physics of strongly correlated matter, or making accurate predictions for highly multireferential chemical systems, such as those found during the breaking of chemical bonds. Under these circumstances, one cannot rely on lower-level techniques: the accuracy required to correctly predict the behaviour of these systems is too great to rely on fortunate cancellation of systematic errors. Furthermore, highly accurate methods have recently become an essential requirement for the development and validation of computationally efficient, yet less precise, computational methods – for example, machine-learned interatomic potentials.

In recent years, the computational sciences have undergone a revolution due to the development of ‘deep neural networks’. These are highly flexible schemes for approximating unknown functions by optimizing some objective, or ‘loss function’; a process known as ‘deep learning’. Deep neural networks are now central to state-of-the-art techniques in fields like computer vision and natural language processing. Naturally, practitioners of applied quantum mechanics – being, in many ways, a computational discipline – have sought to replicate the success of deep learning in their own pursuits. Here, we will join them.

In this thesis, we will explore the recently developed combination of the variational Monte Carlo method with a deep neural network-based representation of the many-body wavefunction. The resulting algorithm – neural network variational Monte Carlo – has been shown to produce extremely accurate solutions to the many-body Schrödinger equation (the central equation of quantum mechanics) while incurring only a cubic scaling of computational cost with system size. We will explore the features of the method that render it distinct from other quantum chemical methods. In particular, we will demonstrate that the lack of dependence on a set of basis functions enables novel applications and radically expands our ability to describe novel quantum systems.

The thesis is structured as follows. Chapter 2 introduces the general goal of quantum chemistry and electronic structure theory – solving the time-independent Schrödinger equation – and provides a high-level overview of the most commonly used techniques. In chapter 3, we introduce the variational Monte Carlo technique

and the Slater-Jastrow-Backflow ansatz, the most commonly used form of the many-body wavefunction. This is followed by a brief introduction to the principles of deep neural networks, a historical review of the development of neural network variational Monte Carlo, and a detailed description of the neural network wavefunction architectures employed in the later chapters: ‘FermiNet’ and ‘PsiFormer’. We conclude the chapter with a thorough review of recent and emerging research directions in neural network variational Monte Carlo.

The novel advantages of neural network variational Monte Carlo are demonstrated through three specific applications in chapters 4 to 6. In chapter 4 we develop the theory and technical implementation of periodic boundary conditions in neural network variational Monte Carlo, enabling the study of solid systems. We apply this method to the homogeneous electron gas, and demonstrate that the flexibility of neural networks allows a totally *ab initio* ‘discovery’ of the Wigner transition. In chapter 5, we study the ability of neural network wavefunctions to describe ‘multicomponent’ systems – those containing multiple species of quantum particle – by calculating the binding energy and annihilation rate of positronic bound states with ordinary molecules. Finally, in chapter 6, we synthesize the technical developments of chapters 4 and 5 to apply the neural network variational Monte Carlo method to the study of nuclear quantum effects, and observe the Wigner-Huntington transition in solid hydrogen with a completely quantum-mechanical treatment of the protonic motion. We conclude in chapter 7 by discussing promising directions for future research based on our results.

Notational conventions employed throughout can be found in appendix A.

## Chapter 2

# Quantum chemistry and electronic structure theory

Quantum chemistry and electronic structure theory are the disciplines concerned with obtaining properties of chemical systems by the application of the laws of quantum mechanics. This is, for the most part, the business of solving the time-independent Schrödinger equation (TISE).

The TISE can be written as an eigenvalue problem,

$$\mathcal{H} |\Psi_i\rangle = E_i |\Psi_i\rangle, \quad (2.1)$$

where  $\mathcal{H}$  is the Hamiltonian operator, with eigenvalues  $E_i$  and eigenstates  $|\Psi_i\rangle$ . Written in a basis of the spin and spatial co-ordinates of particles in a many body system, this becomes,

$$\mathcal{H}\Psi_i(\mathbf{r}_1, \sigma_1, \dots, \mathbf{r}_N, \sigma_N) = E_i \Psi_i(\mathbf{r}_1, \sigma_1, \dots, \mathbf{r}_N, \sigma_N), \quad (2.2)$$

where  $\sigma_i$  and  $\mathbf{r}_i$  are the spin and spatial co-ordinates of the  $i^{\text{th}}$  particle, respectively. Now that we are working in a continuous basis, the Hamiltonian is understood (for each choice of spin co-ordinates) to be a differential operator on the space of the particle co-ordinates.

Everyday matter is an assembly of atoms. In a non-relativistic treatment, this is described by the ‘Born-Oppenheimer molecular Hamiltonian’, referred to from here as *the molecular Hamiltonian*. For a system of atoms containing  $N$  electrons and  $M$  atomic nuclei interacting via the Coulomb potential, the molecular Hamiltonian is written [2],

$$\mathcal{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i=1}^M \sum_{j>i}^M \frac{Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j|} - \sum_{i=1}^N \sum_{j=1}^M \frac{Z_i}{|\mathbf{r}_i - \mathbf{R}_j|}, \quad (2.3)$$

where  $\mathbf{r}_i$  and  $\mathbf{R}_i$  are the co-ordinates of the  $i^{\text{th}}$  electron and atomic nucleus, respectively, and  $Z_i$  is the charge (atomic number) of the  $i^{\text{th}}$  atomic nucleus. In writing the molecular Hamiltonian, we have invoked the Born-Oppenheimer approximation<sup>1</sup> by treating the atomic nuclei as infinitely massive point particles. Thus, we omit the kinetic energy of the nuclei from the Hamiltonian, and treat the nuclear co-ordinates,  $\mathbf{R}_i$ , as fixed parameters rather than dynamical degrees of freedom. Here, and throughout the remainder of the thesis, we are utilizing atomic units, where the reduced Planck's constant,  $\hbar$ , the rest-mass of the electron,  $m_e$ , and the elementary charge,  $e$ , are taken equal to unity.

In addition to the molecular Hamiltonian, we must impose the fermionic exchange statistics of electrons. That is, the many-body electronic wavefunction is a totally antisymmetric function under the exchange of the spin and spatial co-ordinates of two electrons,

$$\Psi(\dots, \mathbf{r}_i, \sigma_i, \dots, \mathbf{r}_j, \sigma_j, \dots) = -\Psi(\dots, \mathbf{r}_j, \sigma_j, \dots, \mathbf{r}_i, \sigma_i, \dots). \quad (2.4)$$

An important consequence of the quantum statistics of electrons is the ‘Pauli exclusion principle’. In a real-space basis this manifests as the property that the wavefunction must vanish at the coalescence of two like-spin electrons,

$$\Psi(\dots, \mathbf{r}, \sigma, \dots, \mathbf{r}, \sigma, \dots) = 0. \quad (2.5)$$

Finally, we are prepared to begin tackling Dirac's proposed endeavor. Solving eq. (2.2) for many interacting electrons is a formidable task: for a system containing  $N$  electrons, this becomes an ODE in  $3N$  variables. Thus, obtaining the full spectrum of the molecular Hamiltonian is impossible for systems of practical interest. Fortunately, nature extends an olive branch in the form of the ‘variational principle’.

The spectrum of a physical Hamiltonian must be bounded from below such that there exists a ‘ground-state energy’  $E_0 \leq E_i$  for all  $i$ . Thus, solving eq. (2.2) for the ‘ground-state wavefunction’ can be reformulated as a variational problem,

$$|\Psi_0\rangle = \underset{\Psi}{\operatorname{argmin}} E[\Psi], \quad E[\Psi] = \frac{\langle \Psi | \mathcal{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle}. \quad (2.6)$$

We will refer to  $E[\Psi]$  as the ‘energy’ associated with a wavefunction  $\Psi$ . The energy is defined for any  $\Psi$ , even when  $\Psi$  is not an eigenstate of the Hamiltonian. The variational principle offers a powerful, general strategy for finding the ground state solution to the TISE: successively refine guesses (or ansätze) for the form of the ground-state wavefunction until the lowest possible energy is achieved.

---

<sup>1</sup>In fact, we have invoked a stronger approximation – the ‘clamped nuclei’ approximation. These approximations will be discussed in detail in chapter 6.

It might appear that, in the face of an overwhelming adversary, we have immediately given up and resigned ourselves to consider only ground-state solutions of the TISE. While this is a fair criticism, the situation is not as dire as it might seem. In many cases of interest, the energy difference between the ground-state solution of the TISE and the first excited-state solution is larger than the thermal energy scale at room temperature.<sup>2</sup> Thus, in considering ‘every day’ chemistry, treating systems as occupying their electronic ground states is a very good approximation, and one can make a lot of progress equipped only with a procedure for determining the ground-state solution of the TISE.

In the following sections we will introduce several numerical methods that rely on the variational principle to approximately solve the Schrödinger equation for the ground-state wavefunction.

## 2.1 Hartree-Fock theory

The simplest ansatz for the electronic many-body wavefunction satisfying fermionic exchange statistics is the ‘Slater determinant’. To construct a Slater determinant for a system of  $N$  electrons, we define  $N$  orthonormal ‘spin orbitals’,  $\phi_i(\mathbf{r}_j, \sigma_j)$  – functions of the spin and spatial co-ordinates of a single electron. A Slater determinant is the determinant of a matrix whose columns are the spin orbitals evaluated at the co-ordinates of each particle [3],

$$\Psi(\mathbf{r}, \boldsymbol{\sigma}) = D(\mathbf{r}, \boldsymbol{\sigma}) = \begin{vmatrix} \phi_1(\mathbf{r}_1, \sigma_1) & \phi_1(\mathbf{r}_2, \sigma_2) & \dots & \phi_1(\mathbf{r}_N, \sigma_N) \\ \phi_2(\mathbf{r}_1, \sigma_1) & \phi_2(\mathbf{r}_2, \sigma_2) & \dots & \phi_2(\mathbf{r}_N, \sigma_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_N(\mathbf{r}_1, \sigma_1) & \phi_N(\mathbf{r}_2, \sigma_2) & \dots & \phi_N(\mathbf{r}_N, \sigma_N) \end{vmatrix}. \quad (2.7)$$

Exchanging the co-ordinates of two electrons exchanges two columns of the determinant, flipping the sign. Here we have adopted the compact notation  $\mathbf{r} \equiv \mathbf{r}_1, \dots, \mathbf{r}_N$  (and  $\boldsymbol{\sigma}$ , defined equivalently for the spin co-ordinates) that will be utilized throughout the thesis.

Taking a single Slater determinant as the wavefunction ansatz yields ‘Hartree-Fock theory’. The solutions of Hartree-Fock theory in a particular computational basis, and their use in devising ‘post Hartree-Fock’ methods, are at the heart of modern quantum chemistry. We will provide a brief exposition of the main concepts of Hartree-Fock theory in order to contextualize results obtained using these methods later.

---

<sup>2</sup>Even in systems where this is not the case, such as metals, the variational approach often works well. In these cases, the low-lying excited states of the system are usually qualitatively similar to the ground state.

### 2.1.1 The Hartree-Fock equations

Substituting the Slater determinant ansatz, eq. (2.7), into the energy functional, eq. (2.6), one obtains

$$\begin{aligned}
E[\Psi] = & \sum_{i=1}^N \int d\mathbf{x}_1 \phi_i^*(\mathbf{x}_1) \mathcal{H}_0(\mathbf{x}_1) \phi_i(\mathbf{x}_1) \\
& + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \int d\mathbf{x}_1 d\mathbf{x}_2 \phi_i^*(\mathbf{x}_1) \phi_j^*(\mathbf{x}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \phi_i(\mathbf{x}_1) \phi_j(\mathbf{x}_2) \\
& - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \int d\mathbf{x}_1 d\mathbf{x}_2 \phi_i^*(\mathbf{x}_1) \phi_j^*(\mathbf{x}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \phi_i(\mathbf{x}_2) \phi_j(\mathbf{x}_1),
\end{aligned} \tag{2.8}$$

where  $\mathcal{H}_0$  is the ‘non-interacting’ portion of the molecular Hamiltonian,

$$\mathcal{H}_0(\mathbf{x}_i) = -\frac{1}{2} \nabla_i^2 - \sum_{j=1}^M \frac{Z_j}{|\mathbf{r}_i - \mathbf{R}_j|}. \tag{2.9}$$

Here we have combined the spin and spatial co-ordinates of the electrons into the single co-ordinate  $\mathbf{x}_i = (\mathbf{r}_i, \sigma_i)$ , and omitted the Coulomb interaction between atomic nuclei, as this contributes only an overall constant to the energy independent of the wavefunction ansatz.

The only degrees of freedom in the Slater determinant ansatz are the spin orbitals. We minimize the energy functional with respect to the spin orbitals while maintaining orthonormality. That is, we need to find the spin orbitals that make stationary the functional,

$$\mathcal{L}[\{\phi_i\}] = E[\{\phi_i\}] - \sum_{i=1}^N \sum_{j=1}^N \epsilon_{ij} \left( \int d\mathbf{x} \phi_i^*(\mathbf{x}) \phi_j(\mathbf{x}) - \delta_{ij} \right), \tag{2.10}$$

where  $\epsilon_{ij}$  is a matrix of Lagrange multipliers for enforcing the orthonormality constraint. Applying the calculus of functional variations, we obtain [2],

$$\mathcal{F}(\mathbf{x}_i) \phi_i(\mathbf{x}_i) = \sum_{j=1}^N \epsilon_{ij} \phi_j(\mathbf{x}_j), \tag{2.11}$$

where  $\mathcal{F}(\mathbf{x}_i)$  is the *Fock* operator,

$$\mathcal{F}(\mathbf{x}_i) = \mathcal{H}_0(\mathbf{x}_i) + \mathcal{J}(\mathbf{x}_i) - \mathcal{K}(\mathbf{x}_i). \tag{2.12}$$

The *Coulomb* operator,  $\mathcal{J}(\mathbf{x}_i)$ , is written,

$$\mathcal{J}(\mathbf{x}_i) = \sum_{j=1}^N \int d\mathbf{x} \frac{|\phi_j(\mathbf{x})|^2}{|\mathbf{r}_i - \mathbf{r}_1|}, \quad (2.13)$$

and represents the electrostatic potential at  $\mathbf{x}_i$  due to the charge density of the spin orbitals. The *exchange* operator,  $\mathcal{K}(\mathbf{x}_i)$ , is an *integral* operator, that must be written in terms of its operation on the spin orbital  $\phi_i(\mathbf{x}_j)$ ,

$$\mathcal{K}(\mathbf{x}_j)\phi_i(\mathbf{x}_j) = \sum_{k=1}^N \phi_k(\mathbf{x}_j) \int d\mathbf{x} \frac{\phi_k^*(\mathbf{x})\phi_i(\mathbf{x})}{|\mathbf{r}_j - \mathbf{r}|}. \quad (2.14)$$

Unlike the Coulomb operator, the exchange operator has no classical interpretation as an external potential. Its contribution represents the correlated motion of electrons due to the fermionic antisymmetry of the wavefunction.

The ‘Hartree-Fock equations’, eq. (2.11), can be brought into the form of an eigenvalue problem by the application of an appropriate unitary transformation. For any set of spin orbitals, we have the freedom to apply a unitary transformation,

$$\phi'_i = \sum_{j=1}^N U_{ij}\phi_j, \quad (2.15)$$

that diagonalizes the matrix of Lagrange multipliers,  $\epsilon_{ij}$ . Such a transformation leaves the value of the Slater determinant unaltered. The precise form of this transformation is immaterial. From the simple fact that the transformation is possible, we are permitted to write the ‘canonical Hartree-Fock equations’,

$$\mathcal{F}(\mathbf{x})\phi_i(\mathbf{x}) = \epsilon_i\phi_i(\mathbf{x}). \quad (2.16)$$

This is quite an unusual eigenvalue problem, in that the form of the Fock operator itself depends upon the full set of eigenfunctions,  $\{\phi_i\}$ . To solve this problem requires a ‘self-consistent’ procedure. The eigenvalue problem is solved for a Fock operator defined by an initial set of spin orbitals. Then, these solutions are used to produce a new Fock operator. This process is repeated until the change in the solutions between iterations of the procedure falls below some predetermined threshold.

To obtain some insight into the meaning of the Lagrange multipliers,  $\epsilon_i$ , we close eq. (2.16) on the left to obtain,

$$\begin{aligned} \int d\mathbf{x} \phi_j^*(\mathbf{x}) \mathcal{F}(\mathbf{x}) \phi_i(\mathbf{x}) &= \int d\mathbf{x} \phi_j^*(\mathbf{x}) \epsilon_i \phi_i(\mathbf{x}), \\ &= \epsilon_i \delta_{ij}. \end{aligned} \quad (2.17)$$

Setting  $j = i$  and expanding the Fock operator on the left-hand side of eq. (2.17), we obtain,

$$\begin{aligned}\epsilon_i &= \int d\mathbf{x}_1 \phi_i^*(\mathbf{x}_1) \mathcal{H}_0(\mathbf{x}_1) \phi_i(\mathbf{x}_1), \\ &+ \sum_{j=1}^N \int d\mathbf{x}_1 d\mathbf{x}_2 \phi_i^*(\mathbf{x}_1) \phi_j^*(\mathbf{x}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \phi_i(\mathbf{x}_1) \phi_j(\mathbf{x}_2), \\ &- \sum_{j=1}^N \int d\mathbf{x}_1 d\mathbf{x}_2 \phi_i^*(\mathbf{x}_1) \phi_j^*(\mathbf{x}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \phi_i(\mathbf{x}_2) \phi_j(\mathbf{x}_1).\end{aligned}\tag{2.18}$$

We find that the sum of the Lagrange multipliers over  $i$  is very nearly identical to the energy functional, eq. (2.8). The only difference is that the contribution from the Coulomb and exchange terms has been double counted (*i.e.*, we are missing a factor of 1/2). Thus, after the solution of eq. (2.16), we can write the Hartree-Fock approximation of the ground-state energy from the Lagrange multipliers as,

$$\begin{aligned}E_{\text{HF}} &= \sum_{i=1}^N \left[ \epsilon_i - \frac{1}{2} \sum_{j=1}^N \int d\mathbf{x}_1 d\mathbf{x}_2 \phi_i^*(\mathbf{x}_1) \phi_j^*(\mathbf{x}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \phi_i(\mathbf{x}_1) \phi_j(\mathbf{x}_2), \right. \\ &\quad \left. + \frac{1}{2} \sum_{j=1}^N \int d\mathbf{x}_1 d\mathbf{x}_2 \phi_i^*(\mathbf{x}_1) \phi_j^*(\mathbf{x}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \phi_i(\mathbf{x}_2) \phi_j(\mathbf{x}_1) \right].\end{aligned}\tag{2.19}$$

This leads to the interpretation of the Lagrange multipliers  $\epsilon_i$  as ‘energies’ corresponding to the spin-orbitals  $\phi_i$ . Whilst this is a useful mental model,<sup>3</sup> it is important to remember that the Lagrange multipliers themselves do not, in general, have any physical meaning. Regardless, we will refer to the Lagrange multipliers as the ‘orbital energies’ from here on out.

From the preceding discussion, it is clear why Hartree-Fock theory is known as a ‘mean-field’ theory. Equation (2.16) takes the form of a single-particle TISE where the external potential is given by the mean field created by the other electrons.

## 2.1.2 Numerical Hartree-Fock theory

Before proceeding, we are going to leave spin by the wayside. The treatment of spin within Hartree-Fock theory is not overly complicated, but will muddy the following discussion. Instead, we wish to focus on the computational aspects of solving the Hartree-Fock equations and how the results are used to derive post-Hartree-Fock methods. Thus, we will omit the spin co-ordinates from here on out and continue

---

<sup>3</sup>For example, this mental picture helps in understanding results like Koopmans’ theorem, that states that the ionization potential and electron affinity for removing / adding an electron to the spin orbital  $\phi_i$  is given by the negative of the Lagrange multiplier  $\epsilon_i$ .

our discussion considering only the spatial co-ordinates,  $\mathbf{r}_i$ , of the electrons. Correspondingly, we will refer to the spin orbitals as simply ‘orbitals’. For a full treatment of spin restricted, unrestricted, and generalized Hartree-Fock, see Szabo and Ostlund [2].

To calculate approximate ground-state wavefunctions and energies with Hartree-Fock theory, we need to devise a prescription for the implementation of the Hartree-Fock equations on a computer. Typically, this is achieved by expressing the orbitals as a linear combination of basis functions,

$$\phi_i(\mathbf{r}) = \sum_j C_{ij} f_j(\mathbf{r}). \quad (2.20)$$

For historical reasons, we refer to the basis functions,  $f_j(\mathbf{r})$  as ‘atomic orbitals’; however, these are not necessarily given by solutions of the Schrödinger equation for lone atoms, and may include, *e.g.*, plane waves. Orbitals expressed as a linear combination of atomic orbitals are known as ‘molecular orbitals’.

Substituting this expansion into the canonical Hartree-Fock equations, eq. (2.16), and closing on the left transforms the integro-differential equation into a linear system of equations,

$$\mathbf{F}\mathbf{C} = \mathbf{S}\mathbf{C}\epsilon. \quad (2.21)$$

Here,  $\mathbf{F}$  is the *Fock matrix*: the Fock operator in the basis of the atomic orbitals,

$$F_{ij} = \int d\mathbf{r} f_i^*(\mathbf{r}) \mathcal{F}(\mathbf{r}) f_j(\mathbf{r}). \quad (2.22)$$

$\mathbf{C}$  is a matrix of coefficients which defines the spin orbitals in eq. (2.20), and  $\mathbf{S}$  is the *overlap matrix*, defined by

$$S_{ij} = \int d\mathbf{r} f_i^*(\mathbf{r}) f_j(\mathbf{r}). \quad (2.23)$$

The generalized eigenvalue problems, eq. (2.21), for all the molecular orbitals are known as the ‘Roothan equations’. The numerical solution of the Roothan equations is the primary way in which Hartree-Fock theory is used to solve the TISE on computers.

Selecting an appropriate basis of atomic orbitals is of tantamount importance to the accuracy and efficiency of the Hartree-Fock method. Within a basis of  $K$  atomic orbitals, a naive evaluation<sup>4</sup> of the Fock matrix demands the calculation of  $O(K^4)$

---

<sup>4</sup>The number of integrals can often be dramatically reduced by exploiting properties of the atomic basis, such as localization and symmetry.

two-electron integrals of the form

$$I_{ijkl} = \int d\mathbf{r}_1 d\mathbf{r}_2 f_i^*(\mathbf{r}_1) f_j(\mathbf{r}_1) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} f_k^*(\mathbf{r}_2) f_l(\mathbf{r}_2). \quad (2.24)$$

Evaluating the two-electron integrals is the primary computational bottleneck of Hartree-Fock theory. Clearly, then, it would be judicious to select a set of atomic orbitals that produce molecular orbitals that are a close approximation of the true ground-state wavefunction with as few elements as possible. Furthermore, one has to consider the computational cost of evaluating eq. (2.24) for the chosen form of the atomic orbitals.

In the spirit of the LCAO approach, an obvious choice of basis functions are the ‘Slater-type orbitals’ (STOs). These are the analytical solutions of the Schrödinger equation for a hydrogenic atom, and exhibit the correct limiting behavior both very close and very far from nuclei. However, evaluating eq. (2.24) for STOs is computationally expensive. A common alternative is to adopt generalized Gaussian basis functions, constructed by taking linear combinations of Gaussians – known individually as ‘primitive’ Gaussians – which approximate STOs. These approximations are known as ‘contracted’ Gaussian type orbitals (GTOs). Accurate calculations in a GTO basis usually require a larger set of basis functions, but evaluating eq. (2.24) for GTOs is *substantially* easier than with STOs. This is because products of Gaussians can be re-written as linear combinations of individual Gaussians, and integrals over the individual Gaussians can be evaluated in closed form.

Solving the Roothan equations in a basis of  $K$  atomic orbitals yields  $K$  molecular orbitals.<sup>5</sup> As our overall objective is to minimize the energy, the  $N$  molecular orbitals associated with the smallest orbital energies are those which populate the Hartree-Fock Slater determinant. The  $N$  orbitals with the lowest energy are known as the *occupied* orbitals. The remaining  $(K - N)$  orbitals are not used in constructing the ground-state solution, and are known as the *virtual* or *unoccupied* orbitals.

Increasing the size of the atomic orbital basis increases the size of the variational space for the molecular orbitals. Thus, increasing the size of the basis set (generally) lowers the energy obtained in a converged Hartree-Fock calculation. In the ‘complete basis set’ (CBS) limit,  $K \rightarrow \infty$ , the lowest possible Hartree-Fock energy – the ‘Hartree-Fock limit’ – is obtained.

### 2.1.3 Post-Hartree-Fock theory

The Hartree-Fock approximation provides a firm footing from which to further analyze solutions to the TISE. A single Slater determinant does not provide a general

---

<sup>5</sup>When accounting for spin, this becomes  $2K$  *spin* orbitals.

solution to the TISE because the motion of the electrons is not correlated beyond the exchange interaction. However, applying the Hartree-Fock approximation through the machinery of the Roothan equations has left us with excess electronic orbitals to work with. The Hartree-Fock determinant is only one of the combinatorial number of possible determinants that can be constructed using the  $K$  orbitals. There are

$$\binom{K}{N} = \frac{K!}{N!(K-N)!} \quad (2.25)$$

such determinants.<sup>6</sup> As we shall see shortly, we can use these determinants to construct successively more accurate solutions to the TISE.

Before proceeding, it is helpful to define a shorthand notation for Slater determinants that does not require the explicit enumeration of all the orbital indices entering the determinant. We take the Hartree-Fock ground state as a *reference* determinant,  $|\Psi_{\text{HF}}\rangle$ . Then, all other determinants are defined by specifying the orbitals that have been substituted relative to the reference state. For example, a determinant in which the orbital  $\phi_a$  has been replaced with the orbital  $\phi_r$  is written as  $|\Psi_a^r\rangle$ . This is a ‘single excitation’, because only one orbital has been altered relative to the reference state. To make this abundantly clear, let us write the real-space representation of these two determinants for a system of two electrons and concrete values of  $a = 1$  and  $r = 3$ :

$$\langle \mathbf{r} | \Psi_{\text{HF}} \rangle = \begin{vmatrix} \phi_1(\mathbf{r}_1) & \phi_2(\mathbf{r}_1) \\ \phi_1(\mathbf{r}_2) & \phi_2(\mathbf{r}_2) \end{vmatrix} \xrightarrow[\text{excitation}]{\text{single}} \langle \mathbf{r} | \Psi_1^3 \rangle = \begin{vmatrix} \phi_3(\mathbf{r}_1) & \phi_2(\mathbf{r}_1) \\ \phi_3(\mathbf{r}_2) & \phi_2(\mathbf{r}_2) \end{vmatrix}. \quad (2.26)$$

That is, each excitation replaces the orbital populating a column of the Slater determinant. Multiply excited determinants follow on from this notation. For example, a ‘doubly excited’ determinant is written  $|\Psi_{ab}^{rs}\rangle$ , indicating that the orbital  $\phi_a$  has been replaced by the orbital  $\phi_r$ , and the orbital  $\phi_b$  has been replaced by the orbital  $\phi_s$ .

These excited determinants form a many-body basis for antisymmetric functions. If the Roothan equations were solved in a complete basis set, the resulting assembly of excited Slater determinants is a *complete basis for fermionic wavefunctions* [2]. Our approach then becomes obvious: write the exact wavefunction as an expansion in the basis of excited Slater determinants,

$$|\Psi\rangle = c_0 |\Psi_0\rangle + \sum_{r=1}^K \sum_{a=1}^K c_a^r |\Psi_a^r\rangle + \sum_{r=1}^K \sum_{s>r}^K \sum_{a=1}^K \sum_{b>a}^K c_{ab}^{rs} |\Psi_{ab}^{rs}\rangle + \dots \quad (2.27)$$

---

<sup>6</sup>Recall again that we are omitting the treatment of spin for simplicity. Accounting for spin replaces  $K \rightarrow 2K$  in this discussion.

By substituting this expansion into the TISE, and once again applying the calculus of variations, we can obtain a set of linear equations to solve for the set of coefficients  $\{c\}$  that define the exact ground-state wavefunction. It is worth pausing for a moment to appreciate this fact: the procedure for the exact solution of the TISE is not particularly complicated.

Of course, our enthusiasm should be short-lived. Solving the Roothan equations in a complete set of basis functions is not possible in reality. Even if it were, the situation becomes more dire when we consider the combinatorial explosion of terms in eq. (2.27) with increasing  $K$ .

We must invoke two approximations to devise a useful scheme for real calculations. The first is to adopt a finite basis of atomic orbitals and abandon the complete basis set limit. Solving the TISE by finding the coefficients that minimize the energy of eq. (2.27) for a finite basis set is known as the ‘full configuration interaction’ (FCI) method. Our second approximation is to adopt a scheme for truncating the expansion in eq. (2.27). This is the central idea in a broad variety of quantum chemistry methods, including the configuration interaction (CI) and coupled cluster (CC) methods.

In the CI method, the expansion of the wavefunction is typically truncated at a finite number of excitations. For example, minimizing the energy of an expansion in all single and doubly excited determinants yields the ‘CI singles-doubles’ (CISD) method. The computational complexity of CI truncated at the  $m^{\text{th}}$  level of excitations is  $O(K^{3m})$ , *i.e.*, the complexity of the CISD method is  $O(K^6)$ . The difference in the number of basis states between systems of different sizes is roughly linear in the number of electrons, such that we can substitute  $K \sim N$  when quoting the computational complexity.

One substantial drawback of the CI method is a lack of ‘size extensivity’: the energy of a system should scale linearly in proportion to the number of electrons,  $N$  [4]. A departure from size extensivity implies that there is a systematic error in energies obtained by the CI method relative to the true ground-state energy that scales with  $N$  [5]. This systematic error will poison any quantities that require the calculation of energy differences between systems with differing numbers of electrons; for example, dissociation energies of molecular fragments.

CC methods were developed to address the lack of size extensivity in the CI method [5–7]. In CC methods, the wavefunction ansatz is written in terms of an excitation operator,  $\mathcal{T}$ , acting on the Hartree-Fock reference state:

$$|\Psi\rangle = \exp(\mathcal{T}) |\Psi_0\rangle. \quad (2.28)$$

This is called the ‘cluster form’ of the wavefunction. The excitation operator is a

sum of operators which take the Hartree-Fock reference state to the  $m^{\text{th}}$  level of excitations,

$$\mathcal{T} = \sum_{m=1}^{\infty} \mathcal{T}_m, \quad \mathcal{T}_m |\Psi_0\rangle = \frac{1}{2^m} \sum_{\substack{r\dots s \\ a\dots b}}^K c_{a\dots b}^{r\dots s} |\Psi_{a\dots b}^{r\dots s}\rangle, \quad (2.29)$$

where  $r\dots s$  and  $a\dots b$  indicate tuples of  $m$  indices. The nomenclature for truncated levels of the cluster form of the wavefunction follows from CI. If  $\mathcal{T}$  is taken to contain only single and double excitation operators,  $\mathcal{T} = \mathcal{T}_1 + \mathcal{T}_2$ , one obtains the ‘CC singles-doubles’ or CCSD method.

To see how the CC method differs from the CI method, consider only the double excitation operator,  $\mathcal{T} = \mathcal{T}_2$ . This is the ‘CC doubles’ or CCD method. The cluster form of the wavefunction now reads,

$$\begin{aligned} |\Psi\rangle &= \exp(\mathcal{T}_2) |\Psi_0\rangle \\ &= (1 + \mathcal{T}_2 + \frac{1}{2}\mathcal{T}_2^2 + \dots) |\Psi_0\rangle \\ &= |\Psi_0\rangle + \frac{1}{4} \sum_{\substack{rs \\ ab}}^K c_{ab}^{rs} |\Psi_a^{rs}\rangle + \frac{1}{32} \sum_{\substack{rs \\ ab}}^K \sum_{\substack{tu \\ cd}}^K c_{ab}^{rs} c_{cd}^{tu} |\Psi_{abcd}^{rstu}\rangle + \dots \end{aligned} \quad (2.30)$$

Unlike CID, the CCD method includes terms in the truncation of eq. (2.27) corresponding to higher-than-double excitations. This is caveated by the fact that the coefficients of these excited determinants are restricted to be products of the coefficients of the double excitations. It can be shown that CC methods are size extensive [5], and they retain the  $O(K^{3m})$  complexity scaling of CI methods.

There is an important extension of the CC method that falls outside the framework of truncating the full CI expansion that we have relied on thus far. This is the inclusion of higher excitations via perturbation theory. For example, the CCSD(T) method – ‘coupled cluster singles-doubles with perturbative triples’ – adds a correction to the CCSD energy based upon a first-order perturbative expansion in the two-electron integrals between single/double and triply excited determinants appearing in the CCSD expansion [8]. The resulting method has a computational complexity of  $O(N^7)$ . We take pains to mention the CCSD(T) method because it is widely considered to be the modern ‘gold standard’ of quantum chemistry for medium-sized molecules, yielding an ideal trade-off between computational cost and accuracy.

Like the Hartree-Fock method, CC and CI methods are applied within a finite basis of atomic orbitals. It is typical to *extrapolate* to the complete basis set limit by carrying out calculations in increasingly large basis sets until one’s computational budget is exhausted [9–11]. The resulting energies are extrapolated to  $K \rightarrow \infty$  by fitting to empirical laws, obtained by observing the convergence of calculations

in very large basis sets for small systems. The development of hierarchical basis sets, which produce well-behaved extrapolations to the complete basis set limit for correlated post-Hartree-Fock methods, has been a significant theme in the historical development of accurate quantum chemistry [12].

The primary drawbacks of post-Hartree-Fock methods are the prohibitive computational scaling with electron number, and the heavy dependence on selecting an appropriate basis to achieve accurate and cost-effective calculations. Nonetheless, correlated post-Hartree-Fock methods like CCSD(T) are now commonplace tools for theoretical chemistry.

## 2.2 Density functional theory

Density functional theory (DFT) is a method for solving the TISE which departs from the framework we have considered thus far: the central object of DFT is *not* the many-body wavefunction. Instead, in DFT, one considers only the ‘one-electron density’,  $n(\mathbf{r})$ , related to the many-body wavefunction by

$$n(\mathbf{r}) = N \int d\mathbf{r}_2 \dots d\mathbf{r}_N |\Psi(\mathbf{r}, \dots, \mathbf{r}_N)|^2, \quad (2.31)$$

where we have again omitted the spin co-ordinates for simplicity. Two remarkable theorems due to Hohenberg and Kohn make it possible to obtain a variational principle in terms of the one-electron density [13].

### 2.2.1 The Hohenberg-Kohn theorems

The first Hohenberg-Kohn theorem states that there is a one-to-one correspondence between the ground-state one electron density and the external potential of a system. For the molecular Hamiltonian, the external potential is given by the Coulomb potential of the fixed nuclei. If we re-write the molecular Hamiltonian, eq. (2.3), as

$$\mathcal{H} = \hat{F} + \hat{V}_{\text{ext}}, \quad (2.32)$$

where

$$\hat{V}_{\text{ext}} = - \sum_{i=1}^N \sum_{j=1}^M \frac{Z_i}{|\mathbf{r}_i - \mathbf{R}_j|}, \quad (2.33)$$

we see that the remaining terms,  $\hat{F}$ , are all – up to an additive constant given by the nuclear repulsion term – independent of the external potential and identical for all  $N$ -electron systems. Thus, the Hamiltonian and, therefore, the ground-state wavefunction are both completely determined by the external potential and number

of electrons,  $N$ . Furthermore, from eq. (2.31), the one-electron density is itself a functional of the ground-state wavefunction and, therefore, also a functional of the external potential. We proceed to prove by contradiction that this functional is one-to-one.

Assume that there exists another external potential,  $\hat{V}'_{\text{ext}}$ , differing from  $\hat{V}_{\text{ext}}$  by more than an additive constant, which yields the same ground-state one-electron density,  $n_0(\mathbf{r})$ . Clearly, the ground-state solution of the TISE must be different for these two external potentials. Denote the two (normalized) ground-state solutions by  $\Psi_0$  and  $\Psi'_0$ , the two Hamiltonians by  $\mathcal{H}$  and  $\mathcal{H}'$ , and the corresponding ground-state energies by  $E_0$  and  $E'_0$ . Now, we have that

$$E'_0 = \langle \Psi'_0 | \mathcal{H}' | \Psi'_0 \rangle < \langle \Psi_0 | \mathcal{H}' | \Psi_0 \rangle = \langle \Psi_0 | \mathcal{H} + (\hat{V}'_{\text{ext}} - \hat{V}_{\text{ext}}) | \Psi_0 \rangle, \quad (2.34)$$

where the inequality arises from the variational principle. Developing the expectation value on the right-hand side, we obtain

$$\begin{aligned} E'_0 &< E_0 + \langle \Psi_0 | (\hat{V}'_{\text{ext}} - \hat{V}_{\text{ext}}) | \Psi_0 \rangle, \\ &= E_0 + \int d\mathbf{r} (V'_{\text{ext}}(\mathbf{r}) - V_{\text{ext}}(\mathbf{r})) n_0(\mathbf{r}), \end{aligned} \quad (2.35)$$

where we have used the fact that the external potential operator is a local one-electron operator to re-express its expectation value as an integral over the one-electron density.

We can carry out exactly the same analysis with the primed and unprimed quantities swapped to obtain,

$$\begin{aligned} E_0 &< E'_0 + \langle \Psi'_0 | (\hat{V}_{\text{ext}} - \hat{V}'_{\text{ext}}) | \Psi'_0 \rangle, \\ &= E'_0 + \int d\mathbf{r} (V_{\text{ext}}(\mathbf{r}) - V'_{\text{ext}}(\mathbf{r})) n_0(\mathbf{r}), \end{aligned} \quad (2.36)$$

where we have leveraged the assumption that both ground-state one-electron densities are identical. Addition of eq. (2.35) and eq. (2.36) yields the inconsistency,

$$E'_0 + E_0 < E'_0 + E_0. \quad (2.37)$$

Thus, our proof is complete.

The fact that the ground-state one-electron density is a one-to-one functional of the external potential implies also that the external potential is uniquely determined (up to an additive constant) by the ground-state one-electron density of a system. Furthermore, as previously stated, the Hamiltonian and, therefore, the ground-state wavefunction are both functionals of the external potential. Thus, fi-

nally, we conclude that the ground-state wavefunction is uniquely determined by the ground-state one-electron density. This is an astonishing fact: the  $3N$ -dimensional ground-state wavefunction which we fought so hard to unravel in the previous section is, in principle, determined by a comparatively pedestrian 3-dimensional function – the ground-state one-electron density.

The second Hohenberg-Kohn theorem produces a variational principle in terms of the one-electron density. From the first theorem, we know that some one-electron density (which is not necessarily the ground-state density of the Hamiltonian under consideration),  $n'(\mathbf{r})$ , specifies (up to an additive constant) some external potential  $V'_{\text{ext}}(\mathbf{r})$ . In turn, this defines a fictitious Hamiltonian with a ground-state wavefunction  $|\Psi'\rangle$ . Taking this fictitious ground-state wavefunction as a variational ansatz for the real Hamiltonian, we find that

$$\frac{\langle \Psi' | \mathcal{H} | \Psi' \rangle}{\langle \Psi' | \Psi' \rangle} \geq E_0, \quad (2.38)$$

where  $E_0$  is the ground-state energy of the true Hamiltonian. Equality is only achieved for non-degenerate ground states when  $|\Psi'\rangle$  is precisely equal to  $|\Psi_0\rangle$ , corresponding to the case where  $n'(\mathbf{r})$  is precisely the ground-state one-electron density in one-to-one correspondence with the real external potential,  $V_{\text{ext}}(\mathbf{r})$ . Thus, we define an energy functional in terms of the density, analogous to eq. (2.6),

$$E[n(\mathbf{r})] = \frac{\langle \Psi[n(\mathbf{r})] | \mathcal{H} | \Psi[n(\mathbf{r})] \rangle}{\langle \Psi[n(\mathbf{r})] | \Psi[n(\mathbf{r})] \rangle}, \quad (2.39)$$

to be minimized with respect to  $n(\mathbf{r})$  in order to solve the TISE. Here we have explicitly written the state,  $|\Psi[n(\mathbf{r})]\rangle$ , as a functional of the one-electron density.

The Hohenberg-Kohn DFT formulation, as presented, has several shortcomings. Equation (2.39) given in terms of the components of the molecular Hamiltonian is written,

$$E[n] = F[n] + \int d\mathbf{r} V_{\text{ext}}(\mathbf{r}) n(\mathbf{r}). \quad (2.40)$$

The second term due to the external potential is straightforward to evaluate in terms of the one-electron density, as it is given in terms of a local, one-electron operator. The first term is a *universal* functional, independent of the system under consideration. Unfortunately, the Hohenberg-Kohn theorems only establish that such a functional exists. They do not make any statement about the nature of this functional, or how to obtain it.

Furthermore, we have assumed so far that any given density,  $n(\mathbf{r})$ , corresponds to an acceptable ground-state density for some fictitious Hamiltonian. In fact, this is not the case (*i.e.*, there exist densities that do not correspond to Hamiltonians

[14]), and one must only consider the set of so-called ‘ $v$ -representable’ densities. For a density that is not  $v$ -representable, the universal functional  $F[n]$  is ill-defined. A naive application of the variational principle, eq. (2.39), may yield nonphysical densities. This problem is closely related to another: the above formulation is only useful in determining non-degenerate ground states.

These issues are resolved by Levy [15]. They consider a modification of the universal functional,  $Q[n]$ , which is equivalent to  $F[n]$  for  $v$ -representable densities, but does not become ill-defined for densities that are not  $v$ -representable. Instead,  $Q[n]$  imposes the weaker condition of ‘ $N$ -representability’: the requirement that the density is one which can be obtained from some antisymmetric wavefunction for  $N$  electrons.

From eq. (2.39), we see that the universal functional  $F[n]$  is defined as the expectation value of the operator  $\hat{F}$  with respect to the ground-state wavefunction of the fictitious Hamiltonian defined by the external potential corresponding to  $n$ ,

$$F[n] = \langle \Psi_0^{(n)} | \hat{F} | \Psi_0^{(n)} \rangle. \quad (2.41)$$

The  $v$ -representability problem becomes obvious here: this functional is only well-defined if the density,  $n$ , yields a well-defined Hamiltonian and associated ground-state wavefunction. Instead, the modified universal function is defined as,

$$Q[n] = \min_{|\Psi^{(n)}\rangle} \langle \Psi^{(n)} | \hat{F} | \Psi^{(n)} \rangle. \quad (2.42)$$

That is, the functional  $Q[n]$  chooses the minimum expectation value of  $\hat{F}$  with respect to *all antisymmetric wavefunctions consistent with the density  $n$* .

Re-defining the energy functional in terms of  $Q[n]$ , we see that the variational lower bound is still trivially obeyed,

$$\begin{aligned} E[n] &= Q[n] + \int d\mathbf{r} V_{\text{ext}}(\mathbf{r}) n(\mathbf{r}), \\ &= \min_{|\Psi^{(n)}\rangle} \langle \Psi^{(n)} | \hat{F} | \Psi^{(n)} \rangle + \int d\mathbf{r} V_{\text{ext}}(\mathbf{r}) n(\mathbf{r}), \\ &= \langle \Psi_{\text{min}}^{(n)} | \hat{F} + \hat{V}_{\text{ext}} | \Psi_{\text{min}}^{(n)} \rangle, \\ &= \langle \Psi_{\text{min}}^{(n)} | \mathcal{H} | \Psi_{\text{min}}^{(n)} \rangle \geq E_0, \end{aligned} \quad (2.43)$$

where  $\Psi_{\text{min}}^{(n)}$  is shorthand for the wavefunction consistent with  $n$  which minimizes the expectation value of  $\hat{F}$ . Now it remains to prove that the ground-state density of the real Hamiltonian,  $n_0$ , minimizes the energy functional in terms of  $Q[n]$  and

yields the ground-state energy. By the variational principle,

$$\langle \Psi_0 | \hat{F} | \Psi_0 \rangle + \int d\mathbf{r} V_{\text{ext}}(\mathbf{r}) n_0(\mathbf{r}) \leq \langle \Psi_{\min}^{(n_0)} | \hat{F} | \Psi_{\min}^{(n_0)} \rangle + \int d\mathbf{r} V_{\text{ext}}(\mathbf{r}) n_0(\mathbf{r}), \quad (2.44)$$

therefore,

$$\langle \Psi_0 | \hat{F} | \Psi_0 \rangle \leq \langle \Psi_{\min}^{(n_0)} | \hat{F} | \Psi_{\min}^{(n_0)} \rangle. \quad (2.45)$$

By the definition of  $\Psi_{\min}^{(n_0)}$ , we have that

$$\langle \Psi_0 | \hat{F} | \Psi_0 \rangle \geq \langle \Psi_{\min}^{(n_0)} | \hat{F} | \Psi_{\min}^{(n_0)} \rangle, \quad (2.46)$$

because  $\Psi_0$  must be amongst the wavefunctions which are consistent with the ground-state density  $n_0$ . Comparing these two inequalities, we arrive at the conclusion that

$$\langle \Psi_0 | \hat{F} | \Psi_0 \rangle = \langle \Psi_{\min}^{(n_0)} | \hat{F} | \Psi_{\min}^{(n_0)} \rangle, \quad (2.47)$$

and  $E[n_0] = E_0$  as required.

## 2.2.2 Kohn-Sham DFT

We now proceed with deriving a numerical procedure for the solution of the TISE within the DFT framework obtained in the previous section. The most common numerical procedure is the Kohn-Sham method [16]. While it is attractive to treat the interacting electron problem by considering the density alone, Kohn and Sham re-introduce electronic orbitals into the method. This results in a great simplification.

Derivation of the Kohn-Sham equations begins by re-expressing the universal functional  $Q[n]$  in terms of three constituents: the kinetic energy functional,  $T[n]$ ; the Coulomb functional,  $E_C[n]$ ; and the ‘exchange-correlation’ functional,  $E_{\text{XC}}[n]$ . The kinetic energy functional is taken as the kinetic energy of a system of non-interacting electrons with density  $n$ , and the Coulomb functional is simply given by the self-interaction energy of a charge density  $n$ ,

$$E_C[n] = \frac{1}{2} \int d\mathbf{r} d\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}. \quad (2.48)$$

The Coulomb term is analogous to the classical Coulomb operator we considered in deriving the Hartree-Fock equations. Likewise, the nomenclature of the *exchange-correlation* functional suggests that  $E_{\text{XC}}[n]$  describes the non-classical contributions to the energy arising from the correlated motion of the electrons. By definition, the exchange-correlation functional also includes the difference between the non-interacting kinetic energy and the true kinetic energy of the interacting system.

The benefit of explicitly separating out these contributions is that the kinetic and Coulomb functionals are easily evaluated. Substantial progress can then be made with fairly crude approximations for the form of the exchange-correlation functional, because its contribution is relatively small.

To apply the variational principle, we wish to make stationary the functional,

$$E[n] = T[n] + E_{\text{XC}}[n] + \int d\mathbf{r} \left[ \frac{1}{2} \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + V_{\text{ext}}(\mathbf{r}) \right] n(\mathbf{r}), \quad (2.49)$$

with respect to the density,  $n$ , subject to the constraint that

$$\int d\mathbf{r} n(\mathbf{r}) = N. \quad (2.50)$$

The variational problem can be written compactly as,

$$\frac{\delta T[n]}{\delta n} + V_{\text{KS}}(\mathbf{r}) = \lambda, \quad (2.51)$$

where  $\lambda$  is a Lagrange multiplier for the electron number constraint, and  $V_{\text{KS}}$  is the ‘Kohn-Sham potential’,

$$V_{\text{KS}}(\mathbf{r}) = \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + V_{\text{ext}}(\mathbf{r}) + V_{\text{XC}}(\mathbf{r}), \quad (2.52)$$

defined in terms of a fictitious potential corresponding to the exchange-correlation energy,

$$V_{\text{XC}}(\mathbf{r}) = \frac{\delta E_{\text{XC}}[n]}{\delta n}. \quad (2.53)$$

This variational problem, eq. (2.51), is equivalent to that which would be obtained by treating a system of *non-interacting* electrons subject to an external potential  $V_{\text{KS}}(\mathbf{r})$ . Thus, the ground-state energy and density can be obtained by solving the one-electron TISE

$$\left[ -\frac{1}{2} \nabla^2 + V_{\text{KS}}(\mathbf{r}) \right] \phi_i(\mathbf{r}) = \epsilon_i \phi_i(\mathbf{r}), \quad (2.54)$$

for a set of orbitals,  $\phi_i$ , from which we construct the density,<sup>7</sup>

$$n(\mathbf{r}) = \sum_{i=1}^N |\phi_i(\mathbf{r})|^2, \quad (2.55)$$

---

<sup>7</sup>The upper summation index is usually quoted as  $N/2$  here, for spin-unpolarized states where the spin-up and spin-down orbitals are taken to be identical. As we have thus far omitted spin from our discussion, we instead take  $N$  as the upper summand. This would correspond to a fully spin-polarized density.

and non-interacting kinetic energy,

$$T[n] = -\frac{1}{2} \sum_{i=1}^N \int d\mathbf{r} \phi_i^*(\mathbf{r}) \nabla^2 \phi_i(\mathbf{r}). \quad (2.56)$$

Equation (2.54), like the Hartree-Fock equations, must be solved in a self-consistent manner: the Kohn-Sham potential,  $V_{\text{KS}}$ , depends upon the solutions of the equation. Again, as in the case of the Hartree-Fock equations, the one-electron eigenvalues from eq. (2.54) can be used to obtain the total ground-state energy after compensating for double counting of the interaction energies,

$$E[n] = \sum_{i=1}^N \epsilon_i - \frac{1}{2} \int d\mathbf{r} d\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} - \int d\mathbf{r} n(\mathbf{r}) V_{\text{XC}}(\mathbf{r}) + E_{\text{XC}}[n]. \quad (2.57)$$

The numerical implementation of this scheme requires a computational basis in which to vary the orbitals,  $\phi_i$ . We will not bore the reader by re-treading much of the ground from our section on deriving the Roothan-Hall equations. Suffice to say, the introduction of a computational basis yields a set of eigenvalue problems.

What we have derived is, in principle, yet another procedure for *exactly* solving the TISE for a system of  $N$  interacting electrons. Remarkably, this procedure only requires the solution of a one-electron TISE. When the orbitals,  $\phi_i$ , are expressed in an LCAO basis, the computational complexity of solving this one-electron TISE becomes a meager  $O(N^3)$ .

Unfortunately, there is a catch. We still do not know what the exact form of the universal exchange-correlation functional should be. The simplest treatment of the exchange-correlation functional is the ‘local-density approximation’ (LDA), where it is assumed that,

$$E_{\text{XC}}^{\text{LDA}}[n] = \int d\mathbf{r} n(\mathbf{r}) \epsilon_{\text{XC}}(n(\mathbf{r})), \quad (2.58)$$

where  $\epsilon_{\text{XC}}(n)$  is the exchange and correlation energy per electron of a uniform electron gas of density  $n$ . In practice, this is implemented by fitting the functional form of  $\epsilon_{\text{XC}}(n)$  to more expensive, and highly accurate, reference calculations (for example, quantum Monte Carlo calculations, which we will discuss later).

Beyond the local-density approximation, one may consider the ‘generalized-gradient approximation’ (GGA). Under the GGA, one incorporates non-local effects in the exchange-correlation functional through the gradient of the one-electron density. This is written as a correction to the LDA functional,

$$E_{\text{XC}}^{\text{GGA}}[n] = E_{\text{XC}}^{\text{LDA}}[n] + \int d\mathbf{r} F_{\text{XC}}[n(\mathbf{r}), |\nabla n(\mathbf{r})|], \quad (2.59)$$

where  $F_{\text{XC}}$  is a scalar-valued functional of the one-electron density and its gradient.

Nature does not offer us any guidance in choosing the appropriate form of  $F_{\text{XC}}$ , and there are many variations of the functional in common usage. Choosing the correct functional for a given system can be challenging, and it is hard to draw quantitative conclusions about which functional yields results that are most physically accurate.

It seems unlikely that we will ever obtain the true, universal exchange-correlation functional. Practitioners of quantum chemistry and electronic structure theory persist nonetheless with the approximate implementations of DFT available to us. In fact, if CCSD(T) is the ‘gold-standard’ of quantum chemistry, then Kohn-Sham DFT is undoubtedly the ‘workhorse’ of quantum chemistry. For the vast majority of applied practitioners, Kohn-Sham DFT is *the* way in which the TISE is solved on a computer. As of 2014, several of the top 100 most cited journal articles of all time were related to the development of Kohn-Sham DFT or the derivation of specific DFT functionals [17].

## 2.3 Quantum Monte Carlo

Quantum Monte Carlo (QMC) is an umbrella term for many related quantum chemistry methods that use random numbers – through the framework of Monte Carlo integration – to handle the high-dimensional integrals that arise in solving the TISE.

Typically, very few approximations are made in QMC calculations compared to post-Hartree-Fock or DFT methods. As a result, QMC methods are used to produce extremely accurate reference calculations. For example, LDA-DFT functionals are produced by fitting to QMC calculations for the uniform electron gas [18]. In addition to their high accuracy, QMC methods exhibit favorable scaling with system size, typically  $O(N^3)$ . However, this accuracy and favorable scaling comes with an extremely large computational pre-factor when compared with post-Hartree-Fock or DFT methods. As a result of this large pre-factor, even methods with very poor *asymptotic* scaling like CCSD(T) are often preferable to quantum Monte Carlo for small- to medium-sized systems.

Different QMC methods are distinguished by their representation of quantum systems, and how that representation is used to produce solutions of the Schrödinger equation. A wide variety of QMC methods have been developed over the course of the last half-century [19]. In the next chapter, we will discuss two QMC methods in detail: the variational and diffusion Monte Carlo methods.

# Chapter 3

## Variational Monte Carlo and neural network wavefunctions

In this chapter, we will describe the basic principles of the variational Monte Carlo (VMC) method. We will primarily limit our discussion to the details of the neural-network VMC (NNVMC) method, with occasional reference to alternative schemes. After introducing the VMC framework, we discuss the commonly used Slater-Jastrow (-Backflow) (SJB) wavefunction ansatz. This ansatz encodes physical priors that motivate several of the design choices we will make in constructing neural network wavefunction ansätze. We conclude with a comprehensive review of recent developments in NNVMC, its context within the related sub-field of neural quantum states (NQS), and the architecture and numerical implementation of the two neural network wavefunction ansätze used in later chapters: the FermiNet and PsiFormer.

### 3.1 Variational Monte Carlo

The VMC method is a stochastic technique for obtaining approximate solutions to the variational formulation of the TISE. We define a wavefunction ansatz parameterized by a set of variational parameters  $\theta$ ,  $\Psi_\theta$ , referred to as the ‘trial wavefunction’. In the context of electronic structure, this trial wavefunction will be constrained to satisfy the fermionic antisymmetry constraint, eq. (2.4). Furthermore, in the present work, we will discuss ansätze for a fixed particle number and total spin (see section 3.2.1 for further discussion).<sup>1</sup> Thus, the parameters  $\theta$  define only the ‘orbital’ (*i.e.*, spatial) behaviour of the wavefunction.

The energy expectation value,  $E[\Psi_\theta]$  provides an upper-bound to the ground-state energy  $E_0$ . The trial wavefunction is iteratively refined to obtain the lowest possible energy, *i.e.*, the best possible approximation to the ground state. The choice

---

<sup>1</sup>In principle, the particle number or total spin constraint could be relaxed. However, they are greatly simplifying, and appropriate for the systems we will study here.

of ansatz is crucial as it defines the subset of the Hilbert space accessible through different values of  $\theta$ .

The *expressivity* of an ansatz is a loose characterization of the variety of states it can represent. To obtain a good approximation to the ground-state wavefunction, one must balance a sufficiently expressive parameterization against the computational complexity of its evaluation.

Early applications of VMC to fermionic systems include studies of zero-temperature properties of  $^3\text{He}$  [20], and pioneering work on the zero-temperature phase diagram of the homogeneous electron gas [21]. More recent applications of the VMC method are summarized in refs [22–26].<sup>2</sup>

### 3.1.1 Energy evaluation

For a system of  $N$  electrons in the continuum,  $E[\Psi]$  is a high-dimensional integral over the  $3N$  spatial and  $N$  spin co-ordinates of the electrons. This integral is analytically intractable, and cannot be evaluated by numerical quadrature on a regular grid for systems with more than a few quantum particles. For example, a coarse Newton-Cotes quadrature, with only 10 integrand evaluations in each spatial dimension, for a 10-electron system like the neon atom would require  $\sim 10^{30}$  evaluations of the integrand – a clear impossibility.

In the VMC method, we use Monte Carlo integration (see appendix B). Monte Carlo integration is well-suited for evaluating high-dimensional integrals. The numerical error decreases with the inverse square root of the number of point estimates of the integrand, regardless of the integral’s dimensionality. Thus, reducing the numerical error by a factor of  $x$  requires only  $x^2$  times more evaluations of the integrand, unlike the exponential scaling of runtime with dimension required by numerical quadrature on a regular grid.

To evaluate the energy expectation value, we transform  $E[\Psi]$  into a form which can be expressed as a Monte Carlo estimator:

$$\begin{aligned} E[\Psi] &= \frac{\int d\mathbf{r} d\boldsymbol{\sigma} \Psi^*(\mathbf{r}, \boldsymbol{\sigma}) \mathcal{H} \Psi(\mathbf{r}, \boldsymbol{\sigma})}{\int d\mathbf{r} d\boldsymbol{\sigma} |\Psi(\mathbf{r}, \boldsymbol{\sigma})|^2}, \\ &= \frac{\int d\mathbf{r} d\boldsymbol{\sigma} |\Psi(\mathbf{r}, \boldsymbol{\sigma})|^2 \frac{\mathcal{H} \Psi(\mathbf{r}, \boldsymbol{\sigma})}{\Psi(\mathbf{r}, \boldsymbol{\sigma})}}{\int d\mathbf{r} d\boldsymbol{\sigma} |\Psi(\mathbf{r}, \boldsymbol{\sigma})|^2}, \\ &= \int d\mathbf{r} d\boldsymbol{\sigma} p(\mathbf{r}, \boldsymbol{\sigma}) E_L(\mathbf{r}, \boldsymbol{\sigma}), \end{aligned} \tag{3.1}$$

---

<sup>2</sup>In actuality, most of these examples use the VMC method to generate a trial wavefunction as an input to the diffusion Monte Carlo method, which we will discuss later. Indeed, this is the primary use of the VMC method today. This is chiefly due to the fact that the wavefunction ansätze used are insufficiently expressive to closely approximate the true ground-state wavefunction.

where

$$p(\mathbf{r}, \boldsymbol{\sigma}) = \frac{|\Psi(\mathbf{r}, \boldsymbol{\sigma})|^2}{\int d\mathbf{r} d\boldsymbol{\sigma} |\Psi(\mathbf{r}, \boldsymbol{\sigma})|^2} \quad (3.2)$$

is the probability density function defined by the trial wavefunction,  $\Psi$ , and

$$E_L(\mathbf{r}, \boldsymbol{\sigma}) = \frac{\mathcal{H}\Psi(\mathbf{r}, \boldsymbol{\sigma})}{\Psi(\mathbf{r}, \boldsymbol{\sigma})} \quad (3.3)$$

is the ‘local energy’. Evaluation of the normalization constant in eq. (3.2) is not required in Monte Carlo integration: samples from  $p(\mathbf{r}, \boldsymbol{\sigma})$  can be generated with access to only the unnormalized distribution.

We evaluate eq. (3.1) by drawing samples  $(\mathbf{r}^{(i)}, \boldsymbol{\sigma}^{(i)})$  from  $p(\mathbf{r}, \boldsymbol{\sigma})$  and calculating the energy expectation value,

$$E[\Psi] = \langle E_L \rangle_p \simeq \frac{1}{M} \sum_{i=1}^M E_L(\mathbf{r}^{(i)}, \boldsymbol{\sigma}^{(i)}), \quad (3.4)$$

where we have introduced the notation  $\langle \cdot \rangle_p$  to indicate expectation values of local quantities taken with respect to  $p(\mathbf{r}, \boldsymbol{\sigma})$ . In the limit  $M \rightarrow \infty$ , the Monte Carlo estimate of the expectation value approaches the exact energy. Samples are drawn using Markov Chain Monte Carlo (MCMC) methods, typically the Metropolis algorithm [27].

The statistical error is obtained from eq. (3.4) as the standard error in the mean,

$$\sigma_M = \sqrt{\frac{\text{Var}_p(E_L)}{M}} \quad (3.5)$$

where the variance of the local energy,

$$\text{Var}_p(E_L) = \langle (E_L - \langle E_L \rangle_p)^2 \rangle_p, \quad (3.6)$$

is a  $\Psi$ -dependent constant. If  $\text{Var}_p(E_L)$  is too large, we cannot accurately estimate the integral with a computationally feasible number of samples. We can estimate  $\text{Var}_p(E_L)$  by considering the difference between the trial wavefunction and the ground-state wavefunction,

$$\delta\Psi = \Psi - \Psi_0, \quad (3.7)$$

and the difference between the energy expectation value and the true ground-state energy,

$$\delta E = \langle E_L \rangle_p - E_0. \quad (3.8)$$

Re-writing  $\delta E$  as,

$$\delta E = \frac{\langle \delta\Psi | (\mathcal{H} - E_0) | \delta\Psi \rangle}{\langle \Psi | \Psi \rangle}, \quad (3.9)$$

which can be verified by inspection. We see that the energy difference scales as  $\delta E \sim O(\delta\Psi^2)$ . Now, using the equality

$$E_L - \langle E_L \rangle_p = \frac{(\mathcal{H} - E_0)\delta\Psi}{\Psi} - \delta E, \quad (3.10)$$

we also have that  $\text{Var}_p(E_L) \sim O(\delta\Psi^2)$ . If  $\Psi$  is an eigenstate of  $\mathcal{H}$ , then the local energy is constant,

$$E_L(\mathbf{r}, \boldsymbol{\sigma}) = \frac{\mathcal{H}\Psi_i(\mathbf{r}, \boldsymbol{\sigma})}{\Psi_i(\mathbf{r}, \boldsymbol{\sigma})} = \frac{E_i\Psi_i(\mathbf{r}, \boldsymbol{\sigma})}{\Psi_i(\mathbf{r}, \boldsymbol{\sigma})} = E_i, \quad (3.11)$$

and  $\text{Var}_p(E_L) = 0$ . This is known as the ‘zero-variance principle’. As a consequence of the zero variance principle, we do not need many evaluations of the integrand to obtain a reasonable statistical error with an accurate trial wavefunction.

### 3.1.2 Wavefunction optimization

Let us turn our attention to the problem of minimizing  $E[\Psi_\theta]$  with respect to  $\theta$ . These parameters could, for example, be coefficients of a basis set expansion, or, as discussed later, the weights and biases of a neural network.

The zero-variance principle offers some freedom in our choice of optimization objective. We may choose to minimize the energy, the variance of the local energy, or some combination of the two. Variance minimization has the attractive property that the strict lower bound is known *a priori*, and can be rendered numerically stable in very large systems using modified importance sampling schemes [28]. However, energy minimization typically yields closer approximations to the ground-state wavefunction [29–31]. Furthermore, variance minimization may converge to an excited eigenstate, for which the variance of the local energy also vanishes. We will focus our discussion on energy minimization, and adopt that strategy later in the present work.

We wish to solve a (generically, non-linear) optimization problem with respect to  $\theta$ ,

$$\theta^* = \underset{\theta}{\operatorname{argmin}} E[\Psi_\theta]. \quad (3.12)$$

Gradients of the energy with respect to  $\theta$  can be written as a Monte Carlo estimator by the application of the product rule to eq. (3.1),

$$\nabla_\theta E[\Psi_\theta] = \langle E_L^* \nabla_\theta \log \Psi_\theta + E_L \nabla_\theta \log \Psi_\theta^* - 2\langle E_L \rangle_{p_\theta} \nabla_\theta \log |\Psi_\theta| \rangle_{p_\theta}, \quad (3.13)$$

where we have assumed a complex-valued wavefunction. The molecular Hamiltonian is time-reversal invariant, so the ground-state wavefunction can always be written

as a real valued function. For real valued wavefunctions, eq. (3.13) reduces to

$$\nabla_{\theta} E[\Psi_{\theta}] = 2\langle (E_L - \langle E_L \rangle_{p_{\theta}}) \nabla_{\theta} \log |\Psi_{\theta}| \rangle_{p_{\theta}}. \quad (3.14)$$

A detailed derivation of this form of the gradient, and a brief discussion of the real-valuedness of the ground states of time-reversal invariant Hamiltonians, is given in appendix C.

The zero-variance principle also implies that eq. (3.14) vanishes quadratically with the difference between the trial wavefunction and the ground-state wavefunction.<sup>3</sup> This is crucial to the success of the VMC method, allowing us to obtain accurate gradients with relatively few Monte Carlo walkers. We can solve the minimization problem, eq. (3.12), by gradient descent,

$$\theta_{i+1} = \theta_i + \text{update}(\theta_i, \nabla_{\theta_i} E[\Psi_{\theta_i}]), \quad (3.15)$$

such that

$$\lim_{i \rightarrow \infty} \theta_i = \theta^*. \quad (3.16)$$

We will make our choice of gradient descent rule concrete later in section 3.3.5.

In order to calculate highly accurate gradient estimates, we could accumulate a large ( $\sim 10^{5-6}$ ) number of samples from the probability distribution.<sup>4</sup> To make this scheme computationally efficient, the same samples should be used to re-calculate gradients after each update to the parameters  $\theta$  through re-weighting: given samples from a wavefunction defined by parameters  $\theta_0$ , Monte Carlo estimates of quantities taken with respect to a wavefunction defined by a different set of parameters  $\theta$  can be obtained as,

$$\langle x \rangle_{p_{\theta}} = \frac{\langle w(\theta) x \rangle_{p_0}}{\langle w(\theta) \rangle_{p_0}}, \quad w(\theta) = \left| \frac{\Psi_{\theta}}{\Psi_{\theta_0}} \right|^2. \quad (3.17)$$

This is known as ‘correlated sampling’ [19]. If the difference between  $\theta_0$  and  $\theta$  becomes too large, most of the weights will become close to zero and only those that are largest by happenstance will contribute to the Monte Carlo estimates, increasing the statistical error. This is handled by occasionally regenerating the samples during optimization. Highly accurate gradients can be treated as exact, and deterministic optimization methods (e.g. Newton’s method, the Levenberg-Marquadt algorithm) used to minimize the energy.

---

<sup>3</sup>Subject to the mild assumption that  $\nabla_{\theta} \log |\Psi_{\theta}|$  remains finite.

<sup>4</sup>The precise number of samples required to obtain ‘accurate’ gradients depends upon the variance of the local gradient estimator in eq. (3.14). This is sensitive to the form of the ansatz, and the degree to which the trial wavefunction has converged to the true ground state – becoming identically zero at the ground state where  $E_L = \langle E_L \rangle$ . However, this is difficult to determine in practice, and the number of samples required for accurate gradient estimates should be determined empirically.

Another approach, employed in the present work, is to generate smaller batches of samples at every parameter update step, obtain noisy estimates of the gradient, and take many more parameter update steps in the course of optimization. This is similar to ‘mini-batch’ stochastic gradient descent techniques in modern machine learning [32]. Using mini-batches reduces the amount of computation required per gradient descent update without substantially reducing convergence rates in very high-dimensional parameter spaces. Furthermore, the intrinsic noise of mini-batch approaches aids in escaping local minima and saddle points during optimization. With noisy gradients, we require optimization methods that are stable in the presence of statistical noise. The simplest is straightforward gradient descent,

$$\text{update}(\theta, \nabla_{\theta} E[\Psi_{\theta}]) = -\eta \nabla_{\theta} E[\Psi_{\theta}], \quad (3.18)$$

where  $\eta$  is a small positive constant (the ‘step size’ or ‘learning rate’). This approach is often referred to simply as stochastic gradient descent in machine learning literature.<sup>5</sup> More sophisticated methods include Newton’s method incorporating modifications to mitigate the impact of statistical noise [31, 33], and the ‘linear method’ [31, 34, 35] based on minimizing the energy for a linear expansion in the parameters of the wavefunction around the ground state. In the present work, we employ the natural gradient descent method [36], closely related to the ‘stochastic reconfiguration’ method used in traditional VMC calculations [37]. We defer our discussion of natural gradient descent to section 3.3.5.

We conclude by summarizing the general framework of VMC, specialized to energy minimization by stochastic gradient descent:

1. Choose a parameterized ansatz for the wavefunction,  $\Psi_{\theta}(\mathbf{r}, \boldsymbol{\sigma})$ .
2. Draw samples  $(\mathbf{r}^{(1)}, \boldsymbol{\sigma}^{(1)}), \dots, (\mathbf{r}^{(M)}, \boldsymbol{\sigma}^{(M)})$  for the particle co-ordinates from  $p_{\theta}(\mathbf{r}, \boldsymbol{\sigma}) = |\Psi_{\theta}(\mathbf{r}, \boldsymbol{\sigma})|^2$  via Markov Chain Monte Carlo.
3. Calculate Monte Carlo estimates of  $\langle E_L \rangle_{p_{\theta}}$  and  $\nabla_{\theta} \langle E_L \rangle_{p_{\theta}}$  using the drawn samples
4. Update the trial wavefunction parameters via some gradient descent update rule  $\theta_{\text{new}} = \theta_{\text{old}} + \text{update}(\theta_{\text{old}}, \nabla_{\theta} \langle E_L \rangle_{p_{\theta}})$ .
5. Repeat steps 2-4 until convergence.
6. Accumulate many samples to estimate the final energy expectation value, and its statistical uncertainty, with respect to the converged parameters.

---

<sup>5</sup>There is some ambiguity in definitions here, and ‘stochastic gradient descent’ is often used informally to refer to several similar algorithms.

We will refer to steps 2 through 5 as ‘training’ and step 6 as ‘inference’, in line with modern machine learning vernacular. Training is deemed to be converged after a fixed number of iterations, or once the decrease in  $\langle E_L \rangle_{p_\theta}$  averaged over a suitable number of iterations falls below a predetermined threshold.

### 3.1.3 Diffusion Monte Carlo

In traditional VMC calculations, the wavefunction ansatz is not sufficiently expressive to obtain an accurate estimate of the ground-state energy. To improve upon the accuracy of a VMC calculation, the optimized trial wavefunction is used as the input to a ‘diffusion Monte Carlo’ (DMC) calculation.

In the DMC method, the many-body wavefunction evolves under the imaginary time Schrödinger equation [23],

$$-\partial_t \Psi(\mathbf{r}, t) = (\mathcal{H} - E_T) \Psi(\mathbf{r}, t), \quad (3.19)$$

where  $t$  is the imaginary time, and  $E_T$  is a fixed energy offset. Here, and throughout this section, we have suppressed spin co-ordinates for simplicity. The wavefunction is represented by  $N_w$  Monte Carlo walkers moving in the  $3N$ -dimensional space of electron co-ordinates. Writing the wavefunction at some time  $t$  as a linear combination of Dirac deltas,<sup>6</sup>

$$\Psi(\mathbf{r}, t) = \sum_{i=1}^{N_w} \delta(\mathbf{r} - \mathbf{r}^{(i)}(t)) \quad (3.20)$$

and substituting into eq. (3.19), the wavefunction at a time  $\tau$  later can be written

$$\begin{aligned} \Psi(\mathbf{r}, t + \tau) &= \int d\mathbf{r}' G(\mathbf{r}, \mathbf{r}', \tau) \Psi(\mathbf{r}', t), \\ &= \sum_{i=1}^{N_w} G(\mathbf{r}, \mathbf{r}^{(i)}(t), \tau). \end{aligned} \quad (3.21)$$

Here,  $G(\mathbf{r}, \mathbf{r}', \tau)$  is the Green’s function of the imaginary-time Schrödinger equation. In the limit of small  $\tau$ , the kinetic energy operator in the Hamiltonian contributes a diffusive component to the Green’s function, and the potential energy operator contributes a growth component, i.e. a reweighting of the walkers. The dynamics is simulated by making a small  $\tau$  approximation to the Green’s function, sampling the diffusive term by Markov Chain Monte Carlo, and reweighting the walkers according to the growth term. The diffusive dynamics transforms each delta function into a Gaussian with variance  $\tau$ , and the reweighting term is most commonly implemented by resampling the number of walkers. The evolution over long time scales is enacted

---

<sup>6</sup>Offering appropriate apologies to any nearby mathematicians.

by repeated application of this process.

Expressing the wavefunction on the eigenbasis of  $\mathcal{H}$ ,

$$\Psi(\mathbf{r}, 0) = \sum_i c_i \psi_i(\mathbf{r}), \quad (3.22)$$

and considering the application of the imaginary-time evolution operator

$$\Psi(\mathbf{r}, t) = \sum_i e^{-(\mathcal{H}-E_T)t} c_i \psi_i(\mathbf{r}) = \sum_i e^{-(E_i-E_T)t} c_i \psi_i(\mathbf{r}), \quad (3.23)$$

we find that imaginary-time evolution projects out the ground-state wavefunction,

$$\Psi(\mathbf{r}, t \rightarrow \infty) = e^{-(E_0-E_T)t} c_0 \psi_0(\mathbf{r}). \quad (3.24)$$

We can estimate the ground-state energy by adjusting  $E_T$  to maintain a constant population of walkers. From eq. (3.23), we see that this is only achieved when  $E_T$  is chosen equal to the ground-state eigenvalue of  $\mathcal{H}$ . Given a reasonable initial estimate of the ground-state energy  $E_0$ , we can tune  $E_T$  around this value to obtain an approximation of the ground-state energy.

The DMC method as described – known as ‘pure DMC’ – is computationally useless for fermionic systems. Our representation of the wavefunction cannot vary in sign, as is required for enforcing fermionic anti-symmetry. Schemes for directly incorporating fermionic anti-symmetry suffer from the infamous fermion sign problem, and cannot produce accurate ground-state properties in any but the tiniest systems. Even when the ground-state wavefunction is strictly positive everywhere, pure DMC is inefficient for systems subject to divergent potentials. The walker population explodes in regions of configuration space where the potential diverges toward  $-\infty$ , prohibiting the accumulation of adequate statistics with finite sampling. Even if the divergence is removed with the use of, *e.g.*, model potentials, the walker population may fluctuate wildly from time-step to time-step.

Both problems are alleviated by incorporating an optimized trial wavefunction,  $\Psi_T$ . Multiplying eq. (3.19) by the trial wavefunction yields (after some rearrangement) [38],

$$-\partial_t f(\mathbf{r}, t) = -\frac{1}{2} \nabla^2 f(\mathbf{r}, t) + \nabla \cdot [\mathbf{v}_D(\mathbf{r}) f(\mathbf{r}, t)] + [E_L(\mathbf{r}) - E_T] f(\mathbf{r}, t), \quad (3.25)$$

where  $f(\mathbf{r}, t) = \Psi(\mathbf{r}, t) \Psi_T(\mathbf{r})$ , and  $\mathbf{v}_D(\mathbf{r})$  is the ‘drift velocity’,

$$\mathbf{v}_D(\mathbf{r}) = \nabla \log |\Psi_T(\mathbf{r})|. \quad (3.26)$$

The Green’s function of this modified equation includes a drift term with the walk-

ers moving at the drift velocity, in addition to the original diffusive component. This acts to drive walkers towards regions where the magnitude of the trial wavefunction is large, imposing an importance sampling transformation on the walker distribution. With a reasonably accurate trial wavefunction, this importance sampling vastly improves sampling efficiency in comparison to the pure DMC method. Additionally, we use the ‘nodal surface’ of the trial wavefunction, defined by the  $3N - 1$  dimensional manifold where  $\Psi_T(\mathbf{r}) = 0$ , to impose fermionic anti-symmetry: any move which takes a walker across the nodal surface during MCMC sampling is rejected. The DMC method augmented in these ways is referred to as ‘fixed-node DMC’ (FN-DMC). The energy obtained by an FN-DMC calculation is variational in the accuracy of the nodal surface of the trial wavefunction. The exact ground-state energy is obtained only when the nodal surface of the trial wavefunction coincides exactly with that of the true ground-state wavefunction.

In this brief discussion, we have omitted many of the important details of a realistic implementation of the FN-DMC method. Instead, we have introduced the main ideas to provide motivation for later discussion of traditional VMC calculations, which are almost always used in accompaniment with the FN-DMC method. Making such calculations accurate and efficient for molecular Hamiltonians requires a substantial number of numerical tricks. We refer the interested reader to refs [39] and [38].

## 3.2 The Slater-Jastrow(-Backflow) Ansatz

The Slater-Jastrow (SJ) wave function is the most commonly used ansatz for VMC calculations [26, 40]. Rather than taking a large linear combination of Slater determinants to incorporate electronic correlation into the wavefunction, a small linear combination of Slater determinants (or even a single Slater determinant) is multiplied by a permutation-invariant function of the electronic co-ordinates, known as a ‘Jastrow factor’, containing explicit many-body terms. We can write the Slater-Jastrow ansatz as

$$\Psi_{SJ}(\mathbf{r}, \boldsymbol{\sigma}) = \exp[J(\mathbf{r}, \boldsymbol{\sigma})] \sum_{k=1}^K c_k D_k(\mathbf{r}, \boldsymbol{\sigma}), \quad (3.27)$$

where  $J(\mathbf{r}, \boldsymbol{\sigma})$  is the Jastrow factor, and the  $D_k(\mathbf{r}, \boldsymbol{\sigma})$  are Slater determinants.

### 3.2.1 Slater determinants and fermionic anti-symmetry

We will only be considering Hamiltonians that do not contain any explicit terms operating on the spin co-ordinates of the electrons. Thus, the operators  $\mathcal{S}_z$  and  $\mathcal{S}^2$ ,

corresponding, respectively, to the axial projection of the total electronic spin, and the squared magnitude of the total electronic spin, commute with the Hamiltonian. This permits a set of simultaneous eigenstates, and defines good quantum numbers  $s_z$  (corresponding to the axial projection of the total spin) and  $s$  (corresponding to total spin). We can restrict our wavefunction ansatz to regions of the Hilbert space with a particular  $s_z$  by making a fixed choice of the electronic spin co-ordinates,  $\sigma_i$ . An ansatz with fixed electronic spin co-ordinates is known as a ‘spin-assigned wavefunction’.<sup>7</sup>

Electrons carry a total spin  $1/2$ , such that there are only two choices of the electron spin co-ordinate,  $\pm 1/2$ , which we will label as ‘up-spin’,  $\uparrow$ , and ‘down-spin’,  $\downarrow$ . Note that the projection axis for the  $\mathcal{S}_z$  operator can be chosen arbitrarily in the absence of an external magnetic field, and up/down do not refer to a particular spatial orientation. Which  $\mathcal{S}_z$  eigenstate the wavefunction represents is determined by the number,  $N_\uparrow$ , of  $\uparrow$  electrons, and the number,  $N_\downarrow$ , of  $\downarrow$  electrons. Different  $\mathcal{S}_z$  eigenstates are distinguished by the difference in the population of  $\uparrow$  and  $\downarrow$  electrons,  $s_z = (N_\uparrow - N_\downarrow)/2$ .

We cannot determine *a priori* the value of the ground-state spin quantum numbers; they can only be determined by direct calculation. In some systems, there are simple empirical rules which are usually effective for this determination, such as Hund’s rules for lone atoms. The ground-state spin quantum number for molecules can also often be inferred from Hund’s rules. Systems for which  $s = 0$  in the ground state are often referred to as ‘closed shell’ systems, and ‘open shell’ systems otherwise.<sup>8</sup> Most ordinary molecules are closed shell in their equilibrium geometry. A prototypical exception is the diatomic oxygen molecule,  $\text{O}_2$ : the ground-state wavefunction for this molecule has total spin quantum number  $s = 1$ .

For a given value of  $s$ , there are  $2s+1$  (referred to as the ‘multiplicity’ of the state) allowed values of  $s_z \in \{-s, -s+1, \dots, s-1, s\}$ . Thus, a state with  $s = 0$  is known as a ‘singlet’ state (having only a single allowed value,  $s_z = 0$ ), a state with  $s = \frac{1}{2}$  is known as a ‘doublet’ (having two allowed values,  $s_z \in \{-\frac{1}{2}, \frac{1}{2}\}$ ), and a state with  $s = 1$  is known as a ‘triplet’ (having three allowed values,  $s_z \in \{-1, 0, 1\}$ ). We can fix only  $s_z$  by spin assignment: the expectation value  $\langle \mathcal{S}^2 \rangle$  remains a variational degree of freedom. An important consequence of this is that we cannot target a specific multiplet state by spin assignment alone for an arbitrary wavefunction ansatz. In

---

<sup>7</sup>We may fix  $s_z$  without choosing a spin-assigned wavefunction by only fixing the total number of electrons with each choice of spin co-ordinate, rather than fixing the specific values  $\sigma_i$ . A spin-assigned wavefunction is distinguished by a mapping of electron indices  $i$  to fixed spin co-ordinates  $\sigma_i$ . It can be shown that spin assignment does not change the expectation value of any spin-independent operator [23].

<sup>8</sup>This nomenclature arises from the chemists’ view of Hartree-Fock molecular orbitals, where an excess of electrons of one spin species corresponds to unpaired electrons in the highest occupied molecular orbital – the outermost ‘shell’.

fact, because our chosen  $s_z$  value may be consistent with multiple multiplet states (for instance,  $s_z = 0$  is consistent with any integer value of  $s$ ), our wavefunction may suffer from ‘spin contamination’ [41], converging to a combination of multiple  $\mathcal{S}^2$  eigenstates. We will not explore spin contamination further in the present work; however, the variational freedom in  $\langle \mathcal{S}^2 \rangle$  does play a brief role in chapter 5 when studying the dissociation of beryllium oxide, exhibiting a transition from a singlet to triplet state as the interatomic distance is increased.

By convention, we write the co-ordinate dependence of the many-body wavefunction such that the first  $N_\uparrow$  spatial co-ordinates,  $\mathbf{r}^\uparrow = (\mathbf{r}_1, \dots, \mathbf{r}_{N_\uparrow})$ , correspond to  $\uparrow$  electrons, and the other  $N_\downarrow$  spatial co-ordinates,  $\mathbf{r}^\downarrow = (\mathbf{r}_{N_\uparrow+1}, \dots, \mathbf{r}_N)$ , correspond to  $\downarrow$  electrons. We retain the superscript notation indicating the electron spin for clarity. Under a fixed spin assignment, we only need to maintain fermionic anti-symmetry under the exchange of the spatial co-ordinates of electrons with the same spin. In a Slater-Jastrow wavefunction, this makes it possible to replace the full Slater determinant with a product of separate determinants for the  $\uparrow$  and  $\downarrow$  electrons,

$$D(\mathbf{r}) = D^\uparrow(\mathbf{r}_1, \dots, \mathbf{r}_{N_\uparrow}) D^\downarrow(\mathbf{r}_{N_\uparrow+1}, \dots, \mathbf{r}_N) = D^\uparrow(\mathbf{r}^\uparrow) D^\downarrow(\mathbf{r}^\downarrow), \quad (3.28)$$

equivalent to a block-diagonal  $N \times N$  determinant.

The use of spin-assigned wavefunctions is computationally convenient. The  $O(N^3)$  evaluation of the full Slater determinant is replaced with the cheaper  $O(N_\uparrow^3 + N_\downarrow^3)$  evaluation of the individual blocks. Additionally, we do not have to worry about sampling over the spin co-ordinates in the Monte Carlo evaluation of the energy expectation value eq. (3.4). All the calculations presented in the later chapters are carried out with spin-assigned wavefunctions.

The choice of single-particle orbitals appearing in the Slater determinants is extremely important, given that Hartree-Fock theory alone typically obtains  $> 90\%$  of the total variational energy. Furthermore, if these orbitals are not representative of the appropriate ground-state wavefunction, it is hopeless to imagine our wavefunction ansatz will yield a good approximation. Fortunately, in the case of molecular systems, we already have rather good mean-field approximations that can be used to produce appropriate single-particle orbitals. Typically, orbitals produced by an initial HF or DFT calculation are used for single-determinant wavefunctions. Multi-determinant wavefunctions can be constructed using the output of CI or CASSCF calculations [26, 40], with the advantage that the accuracy of the wavefunction can be systematically converged to the FCI limit. Regardless of the origin of the single-particle orbitals used to populate the Slater determinants, they will almost always be expressed as a linear combination of a set of basis functions.

The coefficients of the orbital expansion can be treated as variational parameters to be optimized in the course of a VMC calculation; however, it is also common for calculations with Slater-Jastrow wavefunctions to freeze these coefficients and optimize only the parameters of the Jastrow factor. Simple expressions can be obtained for the gradient of the variance with respect to specific forms of the Jastrow factor, offering the strategy of optimizing the parameters in the Jastrow factor by variance minimization, and subsequently optimizing the parameters of the Slater determinants by energy minimization [26, 42].

### 3.2.2 Jastrow factors and the Kato cusp conditions

The Jastrow factor [43],  $J(\mathbf{r})$ , is a permutation invariant function of the many-electron co-ordinates. An extremely generic representation of the Jastrow factor can be written as a body-order expansion of sums over  $M$ -particle functions,

$$J(\mathbf{r}) = \sum_{i=1}^N J_1(\mathbf{r}_i) + \sum_{i=1}^N \sum_{j=1}^N J_2(\mathbf{r}_i, \mathbf{r}_j) + \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N J_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \dots, \quad (3.29)$$

which is manifestly invariant under permutations of the particle co-ordinates due to the exhaustive summations. Evaluating the  $M^{\text{th}}$  order term incurs a computational expense  $\sim O(N^M)$ , which quickly becomes prohibitive for increasing  $M$ . Thus, Jastrow factors are frequently truncated to a maximum of  $M = 2$  or 3 body terms.

The construction of accurate Jastrow factors is a large topic that we will not delve into here. However, one important aspect of the Jastrow factor, which will be relevant to our later discussion of neural-network wavefunctions, is the imposition of the ‘Kato cusp conditions’ [44].

The molecular Hamiltonian contains a bare Coulomb interaction that diverges as  $1/r$  as the separation,  $r$ , between any two charged particles tends to zero. The Kato cusp conditions are analytical conditions on the derivatives of the wavefunction at the coalescence of charged particles: the wavefunction must include a cusp at the coalescence to ensure that the divergence of the Coulomb potential is exactly cancelled by an opposing divergence in the kinetic energy.<sup>9</sup> The one- and two-body terms in the Jastrow factor are chosen to enforce the electron-nuclear and electron-electron cusp conditions, respectively. Written explicitly, the electron-nuclear cusp condition becomes,

$$\left. \frac{d\tilde{\Psi}}{dr} \right|_{r \rightarrow 0} = -Z\tilde{\Psi}, \quad (3.30)$$

where  $\tilde{\Psi}$  represents a spherical average of the wavefunction,  $\Psi$ , about the coalescence

---

<sup>9</sup>We may deduce that the divergence must be exactly cancelled because  $E_L$  is constant and finite everywhere in the exact ground state.

point  $r = 0$ .<sup>10</sup> Fermionic antisymmetry complicates the treatment of the electron-electron cusp condition. The antiparallel spin electron-electron cusp can be written in a similar form to the electron-nuclear cusp,

$$\left. \frac{d\tilde{\Psi}}{dr} \right|_{r \rightarrow 0} = \frac{1}{2} \tilde{\Psi}. \quad (3.31)$$

However, the spherical average,  $\tilde{\Psi}$ , is zero in the limit  $r \rightarrow 0$  for particles with mutual fermionic statistics. Instead, the cusp condition for parallel spin electrons is,

$$\left. \frac{d\Psi_{\ell=1,m}}{dr} \right|_{r \rightarrow 0} = \frac{1}{4} \Psi_{\ell=1,m}, \quad (3.32)$$

where  $\Psi_{\ell=1,m}$  is the projection of the wavefunction onto the spherical harmonic,  $Y_{\ell=1,m}(\mathbf{r})$ , in the separation co-ordinate  $\mathbf{r}$  [44]. Later, we will see that imposing the cusp conditions yields important design considerations for neural network wavefunction ansätze.

The optimization of a Jastrow factor alone does not have any impact on the minimum attainable variational accuracy in an FN-DMC calculation, because the Jastrow factor does not modify the nodal surface of the trial wavefunction. However, the Jastrow factor can impact the nodal surface if the expansion coefficients of the single-particle orbitals populating the Slater determinant(s) are re-optimized together with, or after, the parameters of the Jastrow factor. Furthermore, satisfying the cusp conditions in the trial wavefunction through the Jastrow factor substantially reduces the variance in a DMC calculation, as an accurate Jastrow factor improves the importance sampling transformation imposed by the trial wavefunction.

### 3.2.3 The backflow transformation

The expressivity of the SJ ansatz can be improved by the ‘backflow transformation’, yielding the Slater-Jastrow-Backflow (SJB) ansatz. The backflow transformation is enacted by describing the system in terms of ‘quasiparticle’ co-ordinates obtained via a many-body transformation of the electronic co-ordinates,

$$\zeta_i = \mathbf{r}_i + \sum_{i \neq j} \boldsymbol{\eta}(\mathbf{r}_{ij}, \sigma_i, \sigma_j), \quad (3.33)$$

where  $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$  is the vector separation of electrons  $i$  and  $j$ , and  $\boldsymbol{\eta}$  is a parameterized vector-valued function to be optimized variationally. The explicit double

---

<sup>10</sup>Note that this cusp condition is only strictly valid given an infinite nuclear mass, as taken in the Born-Oppenheimer molecular Hamiltonian. When correcting for the nuclear motion, the right-hand side of eq. (3.30) gains a multiplicative factor of the reduced mass of the electron-nucleus system. For a complete discussion, see [45].

counting of electron pairs in the summation ensures that the backflow transformation is permutation invariant with respect to electron indices. In a fermionic system, this transformation is incorporated by evaluating the single-particle orbitals entering the Slater determinant with respect to the quasiparticle co-ordinates,

$$\begin{vmatrix} \phi_1(\mathbf{r}_1, \sigma_1) & \dots & \phi_1(\mathbf{r}_N, \sigma_N) \\ \vdots & \ddots & \vdots \\ \phi_N(\mathbf{r}_1, \sigma_1) & \dots & \phi_N(\mathbf{r}_N, \sigma_N) \end{vmatrix} \xrightarrow{\text{backflow}} \begin{vmatrix} \phi_1(\boldsymbol{\zeta}_1, \sigma_1) & \dots & \phi_1(\boldsymbol{\zeta}_N, \sigma_N) \\ \vdots & \ddots & \vdots \\ \phi_N(\boldsymbol{\zeta}_1, \sigma_1) & \dots & \phi_N(\boldsymbol{\zeta}_N, \sigma_N) \end{vmatrix}. \quad (3.34)$$

We can also consider higher order backflow transformations, produced via a body-order expansion similar to one use in the generic Jastrow factor in eq. (3.29).

This approach is inspired by the variational description of excitations in liquid Helium adopted by Feynman and Cohen [46, 47]. However, if one expands the definition of a backflow transformation to include any correction to the single-particle orbitals that depends upon the positions of all the electrons, then the first such construction was explored by Wigner and Seitz in 1934 [48]. Later we discuss neural network wavefunction ansätze which adopt this more general definition of the backflow transformation.

While backflow transformations improve the energy obtained in a VMC calculation, their use substantially increases the computational complexity of such calculations. In the absence of backflow, updates to the determinant part of the wavefunction after Monte Carlo moves of a single particle co-ordinate can be efficiently calculated through the co-factor decomposition of the determinant. Because the backflow transformation makes every column of the Slater determinant dependent on the co-ordinates of every particle, this is no longer possible when a backflow wavefunction is used. This drawback also impacts neural network wavefunction ansätze.

Unlike the Jastrow factor, optimization of the backflow transformation at the VMC level does improve the variational accuracy of a subsequent FN-DMC calculation. This is because the backflow transformation directly alters the nodal surface of the trial wavefunction.

### 3.3 Neural network variational Monte Carlo

The VMC method is attractive for two reasons in addition to its conceptual simplicity. First, evaluation of the energy by Monte Carlo integration affords substantial freedom in the choice of wavefunction ansatz. Second, the computational complexity of evaluating a typical wavefunction ansatz for fermions is  $\sim O(N^3)$ , a dramatic improvement over the complexity of accurate quantum chemistry methods, for ex-

ample the  $O(N^7)$  scaling of CCSD(T). However, this raises an obvious question, which highlights the main drawback of traditional VMC: if the FCI expansion of a general wavefunction grows exponentially with system size, and we cannot invoke a justified truncation as for approximate quantum chemistry methods, how can we obtain any guarantee that a particular wavefunction ansatz will produce a reasonable approximation of the ground-state wavefunction?

In this section we will describe the recent developments in combining neural networks with the VMC framework. In the idealized limit of an arbitrary number of parameters and a perfect optimization method, neural networks are universal function approximators, promising to provide the guarantee we seek. We might hope that with realistic neural network wavefunction ansätze we can approach this theoretical ideal in a computationally tractable manner. Indeed, it has been shown that neural-network wavefunction ansätze strike a good balance between expressivity and computational tractability, producing highly accurate approximations of the ground-state wavefunctions of a wide range of different atomic, molecular, and solid-state systems.

### 3.3.1 Deep neural networks

In the modern usage, the phrase ‘neural network’ refers to a rather broad class of schemes for parameterizing arbitrary functions. The adjective ‘neural’ refers to the original inspiration for neural networks as simple models of the human brain in the context of computational neuroscience. Typically, neural networks are used to represent arbitrary mappings; for example, mapping pictures of animals to the names of the animals depicted, or predicting house prices from properties like location and floor space. All neural network architectures can be summarized as a consecutive series of non-linear transformations, each consisting of a linear transformation of a (set of) input vectors followed by the element-wise application of a non-linear ‘activation’ function.

To guide our discussion of neural networks and demonstrate the main principles, we introduce a simple neural network architecture: the ‘multi-layer perceptron’ (MLP). A vector input to the neural network,  $\mathbf{h}_0 \in \mathbb{R}^{n_0}$ , is iteratively updated according to,

$$\mathbf{h}_i = \sigma(\mathbf{W}_i \mathbf{h}_{i-1} + \mathbf{b}_i), \quad (3.35)$$

where  $\sigma$  is the activation function applied element-wise to its vector-valued argument,  $i \in 1..L$  labels the ‘hidden layers’ of the network,  $\mathbf{W}_i \in \mathbb{R}^{n_i \times n_{i-1}}$  are the ‘weights’ in each hidden layer, and  $\mathbf{b}_i \in \mathbb{R}^{n_i}$  are the ‘biases’ in each hidden layer. The integers  $n_i$  correspond to the vector dimensions of the intermediate representations at each hidden layer,  $\mathbf{h}_i$ , and are known as the ‘layer widths’. The composition

of a linear transformation followed by a non-linear activation function is often referred to as a ‘dense’ layer – so-called because each element in the output vector is related to all of (i.e., is dense in) the elements of the input vector. After  $L$  such hidden layers (defining the ‘depth’ of the network), the representation in the final hidden layer is projected into the output space by a final linear transformation

$$\mathbf{h}_{\text{out}} = \mathbf{W}_{\text{out}}\mathbf{h}_L + \mathbf{b}_{\text{out}}. \quad (3.36)$$

This neural network architecture is so ubiquitous in modern deep learning that it is often written simply as a single function,

$$\mathbf{h}_{\text{out}} = \text{MLP}(\mathbf{h}_0). \quad (3.37)$$

Remarkably, it is possible to show that multi-layer perceptrons with bounded depth, an arbitrarily large width, and a non-polynomial activation function are ‘universal function approximators’ of continuous functions [49, 50]. In practice, a computationally tractable neural network cannot be arbitrarily wide. Nonetheless, realistically sized neural networks have been shown to be extremely effective function approximators.

Improving the accuracy of a neural-network-based machine learning algorithm by increasing the width of the hidden layers is now understood to be computationally inefficient. A much better trade-off between performance and computational cost is achieved by making the neural network deeper [51].<sup>11</sup> Developments in the efficient design, implementation, and optimization of such ‘deep neural network’ architectures are the driving force behind the modern ‘deep learning revolution’, which has upended the fields of computer vision and natural language processing [53].

Central to the success of the deep learning revolution is the synergy between neural-network-based algorithms and modern highly parallel computing hardware, primarily graphical processing units (GPUs). GPU hardware architectures were originally designed to efficiently execute the mathematical manipulations required for rendering three-dimensional computer graphics in a highly parallel manner. The main problem to solve in that case is calculating how to draw an object, represented as a large list of triangles (nowadays up to many millions), displaced and rotated relative to an observer and some arrangement of light sources, as pixels on a screen. In practice, this computation reduces to many matrix-matrix multiplications. Likewise, neural networks, and the machine learning algorithms that employ them, can be decomposed into a highly parallelizable series of matrix-algebraic manipulations. Thus, deep learning algorithms are extremely well suited to execution on GPUs [54].

---

<sup>11</sup>It is also possible to construct universal approximation theorems for multi-layer perceptrons of a bounded width and an arbitrary depth [52].

### 3.3.2 Neural quantum states

Many authors have been inspired to apply deep learning to the natural sciences, given the successes of the field in the computational sciences. Deep neural-network based force fields for molecular dynamics were an early, prominent, example of success with this approach [55, 56]. The first attempt to apply neural networks to the problem of representing the ground-state wavefunction of a many-body interacting Hamiltonian was made in the seminal work of Carleo and Troyer [57]. They applied a neural network architecture known as the ‘restricted Boltzmann machine’ (RBM) – an older neural-network architecture, inspired by statistical mechanics – to the problem of finding the ground state of the 1D transverse field Ising model and the 1D and 2D antiferromagnetic nearest-neighbor Heisenberg models. This was the first published instance of a neural-network wavefunction ansatz being used in a variational Monte Carlo calculation.<sup>12</sup> Carleo and Troyer showed that their approach offers a variational improvement in the ground-state energy compared to matrix product states [60, 61], the previous state-of-the-art method.

The promising results of Carleo and Troyer have spawned an entire sub-field focused on the construction of ‘neural quantum states’ – neural-network-based wavefunction ansätze for variational Monte Carlo calculations in lattice systems. This new field includes applications to lattice bosons [62–64] and fermions [65–67], complex spin Hamiltonians beyond the Ising model [68–70], nuclear matter [71], and molecules in a second quantized representation [72–75]. While these calculations *are* NNVMC calculations, we defer to the terminology adopted by researchers in the respective fields: we will refer to VMC calculations with neural-network ansätze as ‘neural quantum states’ (NQS) calculations when the Hilbert space is discrete, and as ‘neural-network variational Monte Carlo’ calculations when the Hilbert space is continuous. This distinction is semantic; however, the research themes and numerical obstacles in the ‘two’ fields are quite different.

### 3.3.3 Continuum neural network wavefunctions

After the pioneering work of Carleo and Troyer on NQS, several groups began developing neural network wavefunction ansätze for systems in the continuum. Deep-Wavefunction (DeepWF) was the first application of neural network wavefunction ansätze to electrons in continuous space [76]. This is structured similarly to the Slater-Jastrow ansatz, with a symmetric Jastrow-like term multiplying an antisymmetric function of the particle co-ordinates; however, the anti-symmetrization is

---

<sup>12</sup>The ‘iterative backflow’ ansatz of Taddei et al. [58] is precisely equivalent to a neural-network backflow based wavefunction ansatz [59]. Strictly, then, this was the first VMC calculation using a neural network ansatz, although at the time it was not recognized as such.

enforced using a ‘Vandermonde determinant’,

$$V(r) = \prod_{1 \leq i \leq j \leq N} (a(r_i, r_j) - a(r_j, r_i)), \quad (3.38)$$

where the functions,  $a(r_i, r_j)$ , are represented by neural networks. The computational complexity of evaluating a Vandermonde determinant is only  $O(N^2)$ , versus the  $O(N^3)$  complexity of evaluating a standard matrix determinant. However, this reduced computational complexity comes with a reduced expressivity: DeepWF obtains only 43.6% of the correlation energy of the beryllium atom and does not even reach Hartree-Fock accuracy for the boron atom.<sup>13</sup>

The first neural network wavefunction ansätze to rival the performance of state-of-the-art quantum chemistry methods were the FermiNet [78, 79] and PauliNet [80, 81]. PauliNet draws heavy inspiration from the SJB wavefunction. It retains the Slater determinants of single particle orbitals obtained from Hartree-Fock or post-Hartree-Fock (in the PauliNet paper, CASSCF) calculations, and parameterizes the Jastrow factor and backflow transformation using deep neural networks. The backflow transformations used in PauliNet are strictly more general than those employed in the SJB ansatz: in addition to introducing a set of quasiparticle co-ordinates, they expand the definition of the backflow transformation to include additive [80] and multiplicative [81] many-body terms that modify the single-particle orbitals. This expanded definition of the backflow transformation is best summarized as a general permutation invariant many-body transformation of a set of reference orbitals.

The FermiNet approaches the construction of the many-body wavefunction from a different philosophical tack, described by a recent review as a ‘machine-learning maximalist’ approach [82]. That is, rather than retaining the physical priors encoded by single-particle orbitals obtained from a quantum chemical calculation, FermiNet opts to represent the wavefunction solely using the output of a neural network. This is achieved by replacing the single particle orbitals entering the Slater determinant by many-body functions that are permutation invariant with respect to all-but-one of the particle co-ordinates. To clearly distinguish them from Slater determinants, we will refer to these objects as ‘Pfau-Spencer’ determinants. We stress that this is (at the time of writing) a non-standard nomenclature. In principle, a single Pfau-Spencer determinant is sufficient to represent *any* fermionic wavefunction [83].

---

<sup>13</sup>The inaccuracy of DeepWF relegates its importance to something of a historical footnote. However, there have been more recent attempts to use computationally cheaper anti-symmetrization schemes than Slater determinants, most prominently via the ‘sortlet’ ansatz [77], which improves upon the accuracy of DeepWF but still struggles to rival the accuracy of Slater-determinant based neural network ansätze. The complexity of evaluating the Slater determinant and its derivatives is the limiting factor in the asymptotic computational complexity of mainstream neural network wavefunction ansätze, making the search for more efficient antisymmetric parameterizations a valuable one despite the limited progress thus far.

This, combined with the universal approximation property of neural networks, would imply that the FermiNet is a universal approximator of fermionic wavefunctions. Unfortunately, however, the proof of the universality of Pfau-Spencer determinants relies upon the representation of highly discontinuous functions which cannot be represented by a realistic neural network. In practice, we find that small linear combination of Pfau-Spencer determinants work well. The FermiNet is one of two neural network wavefunction ansätze that will be employed in the later chapters of the present work, and we will provide a detailed mathematical exposition of the ansatz later in this chapter.

For systems of comparable sizes, the FermiNet ansatz requires more parameters than the PauliNet. This is perhaps not surprising given that the single-particle orbitals entering the PauliNet are already ‘most of the way there’. As a result, VMC calculations using the FermiNet are more expensive, measured in terms of processor-hours, than equivalent PauliNet calculations. However, FermiNet ultimately produces more accurate ground-state energies [82].

NNVMC calculations based on PauliNet or FermiNet produce ground-state energies that are significantly more accurate than traditional SJB-VMC calculations. The accuracy of these ansätze is frequently compared with the accuracy of CCSD(T) calculations extrapolated to the CBS limit (CCSD(T)/CBS). FermiNet typically exceeds the accuracy of finite basis CCSD(T) and SJB-DMC calculations for first-row atoms and small molecules, and exceeds the accuracy of CCSD(T)/CBS on first row atoms, but falls slightly short for small molecules [78]. It is to be appreciated that, until the development of these methods, VMC calculations were regarded as a fairly inaccurate initial step to be followed up by an FN-DMC calculation to obtain accurate ground-state energies. FermiNet and PauliNet alter this narrative, showing that a sufficiently expressive wavefunction ansatz (in this case, offered by a neural network parameterization) can bring VMC calculations to the same level of accuracy as SJB-DMC and sophisticated quantum chemistry methods.

A spiritual successor to the FermiNet, the PsiFormer [84], replaces the neural network architecture used to produce the many-body orbitals populating the Pfau-Spencer determinants. The FermiNet architecture uses a variant of message-passing networks to mix information between electronic orbitals in a permutation equivariant manner. Instead, the PsiFormer utilizes multi-head self-attention, the same mechanism used in the popular Transformer architecture [85], which underlies modern large language models. The PsiFormer produces more accurate ground-state energies for large molecules than the FermiNet with an equal number of parameters [84].

Numerous variants of these three neural network wavefunction ansätze have since been developed; however, they can all be broadly classified as ‘PauliNet-like’ (based

on modifying a set of reference orbitals) or ‘FermiNet-like’ (based on populating Pfau-Spencer determinants with neural network orbitals).

### 3.3.4 FermiNet and PsiFormer

The architectures of the FermiNet and PsiFormer follow the same basic framework. A set of ‘input features’ is derived from the electronic co-ordinates and molecular geometry. These input features are transformed by the neural network backbone – a message-passing neural network [86] in the case of FermiNet, and a Transformer network [85] in the case of the PsiFormer – to produce high-dimensional embeddings for each electron; these, we hope, encode the salient features of the many-electron configuration. Finally, orbitals are produced from these embeddings by a linear transformation and used to populate Pfau-Spencer determinants. To make this process concrete, we will describe the evaluation of the FermiNet and PsiFormer, and the numerical implementation of a NNVMC calculation with these ansätze, from start to finish.

#### FermiNet

Given electronic co-ordinates  $\{\mathbf{r}_i\}$  and nuclear co-ordinates  $\{\mathbf{R}_i\}$ , the input features to the FermiNet can be written

$$\mathbf{h}_{0,i}^\alpha = (\mathbf{r}_i^\alpha - \mathbf{R}_j, |\mathbf{r}_i^\alpha - \mathbf{R}_j| \quad \forall j) \quad (3.39)$$

and

$$\mathbf{h}_{0,ij}^{\alpha\beta} = \left( \mathbf{r}_i^\alpha - \mathbf{r}_j^\beta, |\mathbf{r}_i^\alpha - \mathbf{r}_j^\beta| \right), \quad (3.40)$$

where the brackets are understood to represent a single vector obtain by concatenation of all the enclosed elements,  $\alpha$  and  $\beta \in \{\uparrow, \downarrow\}$  label spin species (chosen for each electron according to the spin assignment of the wavefunction, see section 3.2.1), and the subscripted zeroes indicate that these are the features in the ‘zeroth’ layer of the neural network. The inputs carrying a single subscripted index,  $\mathbf{h}_{0,i}^\alpha$ , are the ‘single-particle’ features, and those carrying two subscripted indices  $\mathbf{h}_{0,ij}^{\alpha\beta}$ , are the ‘two-particle’ features. We can understand this featurization as a complete graph. The nodes of the graph correspond to electrons, where each node carries ‘single-particle’ features. Likewise, the edges of the graph connect pairs of electrons and carry ‘two-particle features’, as depicted in fig. 3.1.

In each layer of the network, nodes and edges are updated in parallel. For each electron, a ‘message’ is constructed by aggregating information between all the nodes in the network and all the edges connected to the node representing that electron.

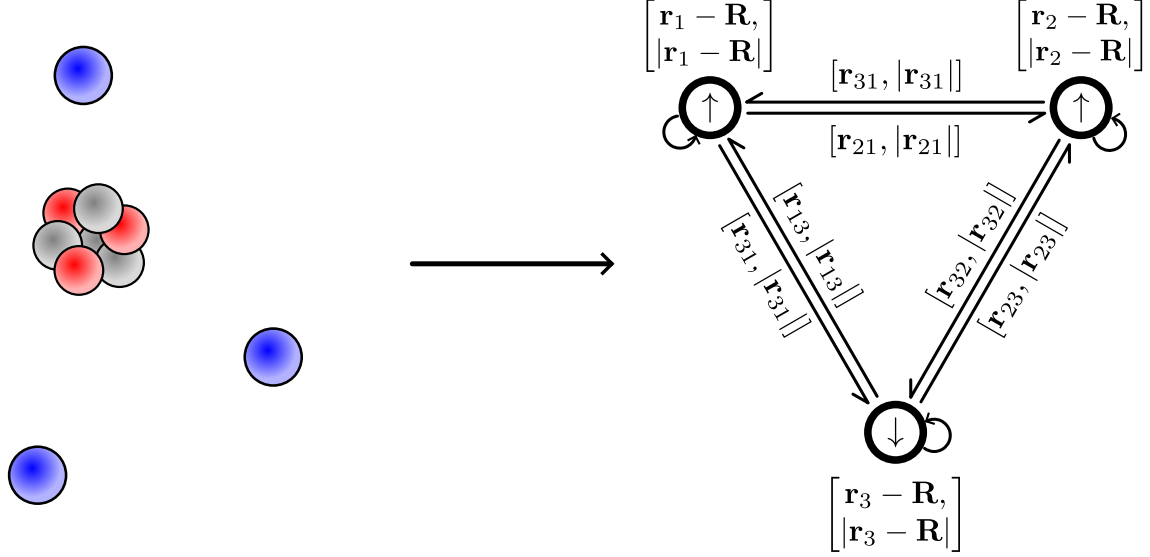


Figure 3.1: Schematic of the FermiNet input featurization of a lithium atom ( $n_{\uparrow} = 2$ ,  $n_{\downarrow} = 1$ ). Electrons are represented by nodes on a complete graph. One-particle features, associated with the nodes, are given by the vector separation and its magnitude between the electrons and the nucleus. Two-particle features, associated with the edges, are given by the vector separation and its magnitude between a pair of electrons.

This message can be written,

$$\mathbf{f}_{l,i}^{\alpha} = \text{concatenate}(\mathbf{h}_{l,i}^{\alpha}, \mathbf{g}_l^{\uparrow}, \mathbf{g}_l^{\downarrow}, \mathbf{g}_{l,i}^{\alpha\uparrow}, \mathbf{g}_{l,i}^{\alpha\downarrow}) \quad (3.41)$$

where the aggregated  $\mathbf{g}$  vectors are given,

$$\mathbf{g}_l^{\alpha} = \frac{1}{N_{\alpha}} \sum_i^{N_{\alpha}} \mathbf{h}_{l,i}^{\alpha} \quad (3.42)$$

for the nodal features, and

$$\mathbf{g}_{l,i}^{\alpha\beta} = \frac{1}{N_{\beta}} \sum_j^{N_{\beta}} \mathbf{h}_{l,ij}^{\alpha\beta} \quad (3.43)$$

for the edge features.

The aggregated message at each node is used to update its features as,

$$\mathbf{h}_{l+1,i}^{\alpha} = \tanh(\mathbf{W}_l \mathbf{f}_{l,i}^{\alpha} + \mathbf{b}_l) + \mathbf{h}_{l,i}^{\alpha}, \quad (3.44)$$

where  $\mathbf{W}_l$  and  $\mathbf{b}_l$  are matrices and vectors (respectively) of variational parameters, and the hyperbolic tangent function is understood to operate element-wise on its vector input. The typical width,  $w_l$ , of the single-particle hidden layers is 256 or 512 neurons [78, 79]. This message-passing and update procedure defines a single

‘FermiNet layer’. The edge features are updated without any message passing step (that is, information is not mixed between edges) according to,

$$\mathbf{h}_{l+1,ij}^{\alpha\beta} = \tanh(\mathbf{V}_l \mathbf{h}_{l,ij}^{\alpha\beta} + \mathbf{c}_l) + \mathbf{h}_{l,ij}^{\alpha\beta}, \quad (3.45)$$

where  $\mathbf{V}_l$  and  $\mathbf{c}_l$  are again matrices and vectors (respectively) of variational parameters. The typical width,  $v_l$ , of the two-particle hidden layers is 32 or 64 neurons [78, 79]. Note that the edge features are not updated at the  $L^{\text{th}}$  FermiNet layer, because they are not used in producing the output of the network beyond this point. A typical FermiNet trial wavefunction consists of  $L = 4$  hidden layers.

Equations (3.44) and (3.45) are equivalent to the dense layers used in the definition of the MLP in section 3.3.1, but with the addition of a ‘residual connection’ made by adding the layer input to its output. Of course, this residual connection requires that the linear dimension of the hidden layer vector is the same as that of its input. In layers where this is not the case (*i.e.*, going from the input features  $\mathbf{h}_0$  to the first hidden layer  $\mathbf{h}_1$ ), residual connections are omitted. Residual connections are used to mitigate the ‘vanishing gradients’ problem in training deep neural networks. Without residual connections, evaluating gradients with respect to the parameters of a neural network requires repeated multiplications by values  $< 1$ . This exponentially suppress gradients, meaning that weights in the early layers do not change value during gradient descent. Residual connections alleviate this problem by allowing gradients to propagate directly back to the earlier layers rather than passing, via the chain rule, through the layers in-between. Residual connections also aid optimization by re-framing the function that is represented by a hidden layer as a perturbation to the identity rather than learning the full mapping from input to output. Ultimately, the primary justification for residual layers is empirical: it is well established that the inclusion of residual connections eases the optimization of deep neural networks [87].

After  $L$  layers of message-passing updates, the nodal features are projected into orbitals via a spin-dependent linear transformation

$$\boldsymbol{\varphi}^{k\alpha}(\mathbf{r}_j^\alpha; \{\mathbf{r}_{/j}^\alpha\}, \{\mathbf{r}^{\bar{\alpha}}\}) = \mathbf{U}^{k\alpha} \cdot \mathbf{h}_{L,j}^\alpha + \mathbf{d}^{k\alpha}, \quad (3.46)$$

where  $\mathbf{U}^{k\alpha}$  and  $\mathbf{d}^{k\alpha}$  are variational parameters defined separately for each spin channel, and for each of the  $n_{\text{det}}$  (typically 16 or 32 [78, 79]) Pfau-Spencer determinants labelled by the index  $k$ . The arguments to  $\boldsymbol{\varphi}$  in curly braces indicate that the Pfau-Spencer orbitals are permutation invariant functions (*i.e.*, functions of the set) of the co-ordinates of the other electrons with the same spin (denoted  $\alpha$ ), and of all co-ordinates of other electrons with the opposite spin (denoted  $\bar{\alpha}$ ). However, they are *not* permutation invariant under the exchange of the *spatial* coordinates of

electrons with different spin assignments. These properties are guaranteed by the separation of sums over different spins in the message aggregation step in eq. (3.41), and the spin-dependent orbital projection in eq. (3.46). From here on, we will write this dependency in shorthand form as  $f(\mathbf{r}_j^\alpha; \{\mathbf{r}_{/j}\})$ , omitting the explicit separation of the other electrons by spin.

To produce the final entries of the Pfau-Spencer determinants for molecular systems, these orbitals are multiplied by an isotropic, exponentially decaying envelope function to enforce the boundary condition that the wavefunction vanishes as  $\mathbf{r}_i \rightarrow \infty$ ,

$$\phi_i^{k\alpha}(\mathbf{r}_j^\alpha; \{\mathbf{r}_{/j}\}) = \varphi_i^{k\alpha}(\mathbf{r}_j^\alpha; \{\mathbf{r}_{/j}\}) \sum_{m=1}^{n_{\text{atom}}} \pi_{im}^{k\alpha} \exp(-\sigma_{im}^{k\alpha} |\mathbf{r}_j^\alpha - \mathbf{R}_m|), \quad (3.47)$$

where  $\pi_{im}^{k\alpha}$  and  $\sigma_{im}^{k\alpha}$  are variational parameters associated with the  $n_{\text{atom}}$  atoms, labelled by the index  $m$ . The indices  $i$ ,  $j$ , and  $k$  index orbitals, electrons, and Pfau-Spencer determinants, and the superscript  $\alpha \in \{\uparrow, \downarrow\}$  distinguishes spin species. The FermiNet wavefunction is then evaluated as,

$$\Psi(\mathbf{r}) = \sum_{k=1}^{n_{\text{det}}} \det \left[ \phi_i^{k\uparrow}(\mathbf{r}_j^\uparrow; \{\mathbf{r}_{/j}\}) \right] \det \left[ \phi_i^{k\downarrow}(\mathbf{r}_j^\downarrow; \{\mathbf{r}_{/j}\}) \right], \quad (3.48)$$

where the expansion coefficients are omitted as they may be absorbed into the determinants. Because the correct exchange symmetry of the orbitals is explicitly enforced in the network architecture, it is possible to utilize ‘dense’ Pfau-Spencer determinants, which are not block diagonal,

$$\Psi(\mathbf{r}) = \sum_{k=1}^{n_{\text{det}}} \det \left[ \begin{array}{c|c} \begin{array}{c} | \\ \phi_i^{k\uparrow} \\ | \end{array} & \begin{array}{c} | \\ \phi_i^{k\downarrow} \\ | \end{array} \end{array} \right]. \quad (3.49)$$

Dense Pfau-Spencer determinants yield variationally improved ground-state energies over block-diagonal determinants [88].

The full FermiNet architecture, mapping the input features to the wavefunction, is depicted in fig. 3.2. All of the symbols used in defining the FermiNet architecture are summarized in table 3.1, including the number of variational parameters associated with each component of the network. Assuming typical values for the parameters of the network ( $w_{l>0} = 256$ ,  $v_{l>0} = 32$ ,  $n_{\text{det}} = 16$ ,  $L = 4$ ) and dense determinants, a FermiNet wavefunction for a bicyclobutane molecule ( $n_{\text{atom}} = 10$ ,  $N = 30$ ,  $w_0 = 4n_{\text{atom}}$ ,  $v_0 = 4$ ) contains 942016 variational parameters.

The message-passing procedure for updating the nodal features, eq. (3.41), is invariant under the permutation of like-spin electron indices: the value of a sum

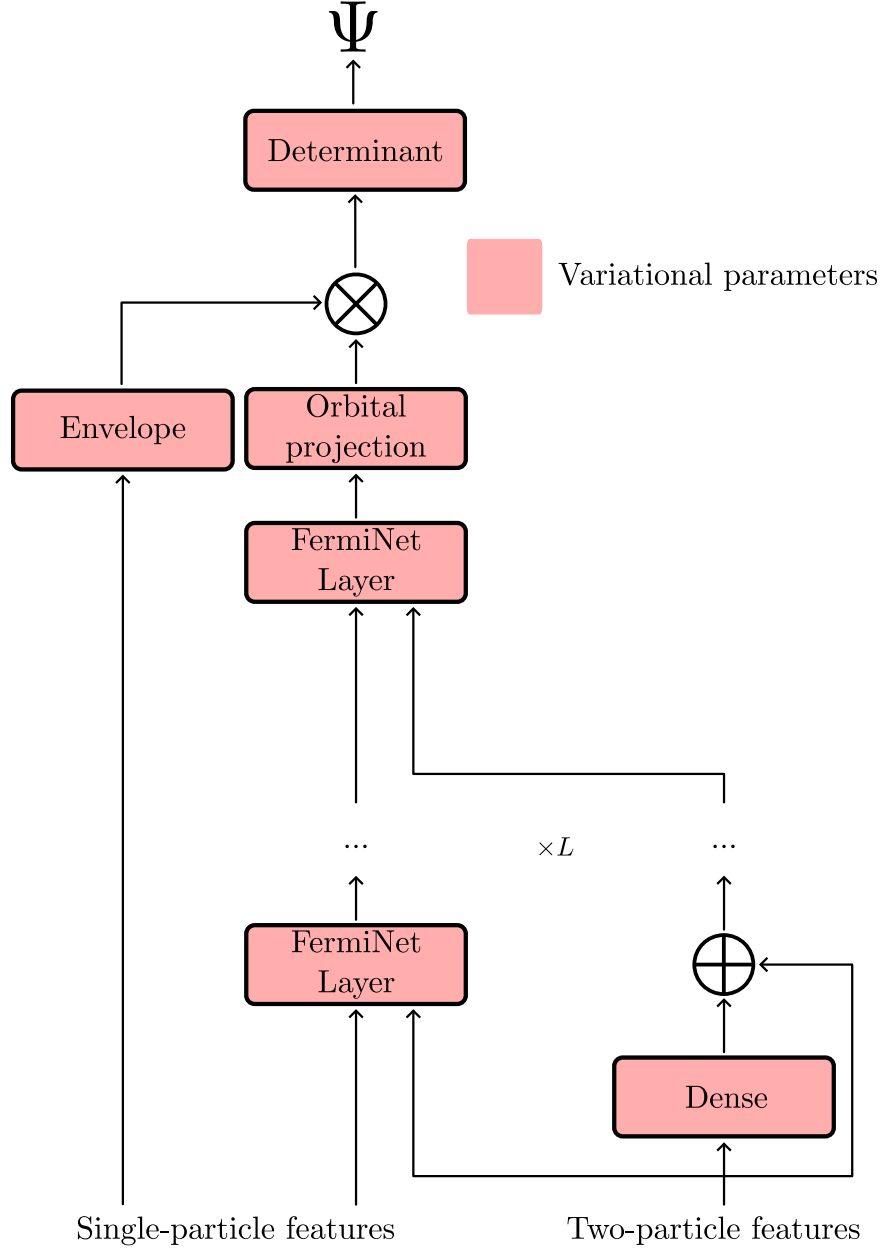


Figure 3.2: Schematic depiction of the FermiNet architecture, mapping input features derived from the electron co-ordinates to the many-body wavefunction. The single-particle and two-particle features are updated in parallel streams, merged via concatenation in the message-passing FermiNet layers. Figure adapted from [84].

| Symbol                  | # Parameters                                                                           | Description                                                      |
|-------------------------|----------------------------------------------------------------------------------------|------------------------------------------------------------------|
| $N$                     |                                                                                        | Number of electrons                                              |
| $n_{\text{det}}$        |                                                                                        | Number of Pfau-Spencer determinants                              |
| $n_{\text{atom}}$       |                                                                                        | Number of atomic nuclei                                          |
| $n_{\alpha}$            |                                                                                        | Number of $\alpha \in \{\uparrow, \downarrow\}$ electrons        |
| $L$                     |                                                                                        | Number of FermiNet layers (depth)                                |
| $w_l$                   |                                                                                        | Width of single-particle features $\mathbf{h}_{l,i}^{\alpha}$    |
| $v_l$                   |                                                                                        | Width of two-particle features $\mathbf{h}_{l,ij}^{\alpha\beta}$ |
| $\mathbf{W}_l$          | $\sum_{l=1}^L (3w_{l-1} + 2v_{l-1})w_l$                                                | Weights for single-particle dense layers                         |
| $\mathbf{b}_l$          | $\sum_{l=1}^L w_l$                                                                     | Biases for single-particle dense layers                          |
| $\mathbf{V}_l$          | $\sum_{l=1}^{L-1} v_{l-1}v_l$                                                          | Weights for two-particle dense layer                             |
| $\mathbf{c}_l$          | $\sum_{l=1}^{L-1} v_l$                                                                 | Biases for two-particle dense layer                              |
| Block determinants      |                                                                                        |                                                                  |
| $\mathbf{U}^{k\alpha}$  | $\sum_{\alpha \in \{\uparrow, \downarrow\}} n_{\text{det}} n_{\alpha} w_L$             | Weights for orbital projection                                   |
| $\mathbf{d}^{k\alpha}$  | $\sum_{\alpha \in \{\uparrow, \downarrow\}} n_{\text{det}} n_{\alpha}$                 | Biases for orbital projection                                    |
| $\pi_{im}^{k\alpha}$    | $\sum_{\alpha \in \{\uparrow, \downarrow\}} n_{\text{det}} n_{\text{atom}} n_{\alpha}$ | Atomic envelope pre-factors                                      |
| $\sigma_{im}^{k\alpha}$ | $\sum_{\alpha \in \{\uparrow, \downarrow\}} n_{\text{det}} n_{\text{atom}} n_{\alpha}$ | Atomic envelope decay coefficients                               |
| Dense determinants      |                                                                                        |                                                                  |
| $\mathbf{U}^{k\alpha}$  | $2n_{\text{det}} N w_L$                                                                | Weights for orbital projection                                   |
| $\mathbf{d}^{k\alpha}$  | $2n_{\text{det}} N$                                                                    | Biases for orbital projection                                    |
| $\pi_{im}^{k\alpha}$    | $2n_{\text{det}} n_{\text{atom}} N$                                                    | Atomic envelope pre-factors                                      |
| $\sigma_{im}^{k\alpha}$ | $2n_{\text{det}} n_{\text{atom}} N$                                                    | Atomic envelope decay coefficients                               |

Table 3.1: Summary of symbols and parameters used in defining the FermiNet architecture.

does not depend upon the ordering of the operands. Phrased alternately, the neural network architecture is agnostic to the labelling of the nodes in the graph and relabelling the inputs (exchanging electron co-ordinates) simply relabels the corresponding outputs. The ordering of columns in the Pfau-Spencer determinants, however, *does* depend upon the labelling, and exchanging the ordering of columns implements the required fermionic exchange symmetry.

By including the magnitude of the vector separations in the input features, eqs. (3.39) and (3.40), we enable the FermiNet to *learn* the Kato cusp conditions. With smooth input features and a continuous activation function, the output of the neural network will also be smooth and continuous, and cannot represent the required cusps. However, the vector magnitude is non-smooth as the vector separation of the particles crosses zero, and the associated discontinuity in derivatives of the wavefunction cannot be removed by the neural network. Thus, the presence of a cusp is incorporated by design (although, the appropriate limiting value of the gradients around the cusp must still be learned by the network). In practice, it is found that the spatial gradient of the optimized wavefunction near the cusps closely obeys the Kato cusp conditions [78].

## PsiFormer

The input features to the PsiFormer are very similar to those used in the FermiNet, with three key differences [84]. First, the PsiFormer discards the notion of two-particle features as inputs to the neural network. Instead, the cusp conditions are implemented using a simple Jastrow factor of the kind typically employed in SJ wavefunctions. Second, the correct spin-assigned exchange symmetry is enforced by including the spin of the electron in the input features, instead of through the design of the network architecture. Finally, the PsiFormer is found to exhibit improved numerical stability when the input features are re-scaled according to the logarithm of the vector separations. These input features then read,

$$\mathbf{h}_{0,i} = (\sigma_i, s_{ij}(\mathbf{r}_i - \mathbf{R}_j), s_{ij}|\mathbf{r}_i - \mathbf{R}_j| \quad \forall j), \quad s_{ij} = \log(1 + |\mathbf{r}_i - \mathbf{R}_j|)/|\mathbf{r}_i - \mathbf{R}_j|. \quad (3.50)$$

In the PsiFormer, the electron features are updated by successive applications of a ‘multi-headed self-attention’ mechanism in place of the message-passing mechanism in the FermiNet. Consider first a single ‘attention head’. The self-attention mechanism constructs intermediate feature vectors,

$$\mathbf{F}_{lh} = \text{softmax} \left( \frac{(\mathbf{H}_l \mathbf{Q}_{lh}) \cdot (\mathbf{H}_l \mathbf{K}_{lh})^T}{\sqrt{d_l}} \right) \cdot \mathbf{H}_l \mathbf{V}_{lh}, \quad (3.51)$$

where  $l$  and  $h$  index layers and heads, respectively,  $\mathbf{H}_l$  is a matrix whose rows are the electron feature vectors  $\mathbf{h}_{l,i}$ ,  $d_l$  is the dimension of the vectors  $\mathbf{h}_{l,i}$ , and  $\mathbf{Q}_{lh}/\mathbf{K}_{lh}/\mathbf{V}_{lh}$  are matrices of variational parameters known as the ‘query’/‘key’/‘value’ weights. The ‘softmax’ function, operates on a matrix  $\mathbf{A}$  as follows,

$$\text{softmax}(\mathbf{A})_{ij} = \frac{\exp(-A_{ij})}{\sum_j \exp(-A_{ij})}, \quad (3.52)$$

replacing each element of  $\mathbf{A}$  by a row-wise normalized Boltzmann factor. This operation maps values on the real number line to the interval  $[0, 1]$ , such that relatively small (large) entries in each row are exponentially suppressed (amplified).

In *multi-headed* self-attention, the self-attention mechanism is executed in parallel for several attention heads with different query, key, and value weights. The output from each attention head is concatenated along the feature axis (i.e., the rows of  $\mathbf{F}$ ) and projected to the desired output dimension (typically, matching the input dimension to facilitate residual connections) by a final linear transformation. Writing the rows of  $\mathbf{F}_{lh}$  (correspond to individual electrons) as vectors  $\mathbf{f}_{lh,i}$ , the

electronic features output from a multi-headed self-attention layer can be written

$$\begin{aligned}\tilde{\mathbf{f}}_{l,i} &= \mathbf{W}_l \mathbf{f}_{l,i} + \mathbf{h}_{l,i}, \quad \mathbf{f}_{l,i} = \text{concatenate}_h(\mathbf{f}_{lh,i}), \\ &\triangleq \text{MHA}(\mathbf{h}_{l,i}) + \mathbf{h}_{l,i},\end{aligned}\tag{3.53}$$

where we have included a residual connection.

The self-attention mechanism offers the following interpretation. Features corresponding to electrons, (i.e., rows of  $\mathbf{H}$ ) are projected into an embedding space by the query/key/value weights. The similarity between the queries and keys corresponding to electrons  $i$  and  $j$  is obtained by taking inner products between the query and key vectors in this space. These similarities, after the application of the softmax, give the entries of an ‘attention matrix’. Updated electron features are then obtained by taking the matrix product between the attention matrix and the ‘value matrix’  $\mathbf{H}_l \mathbf{V}_l$ . Entries  $F_{l,ij}$  in the resulting matrix, where  $i$  indexes electrons and  $j$  indexes the feature dimension, are thus obtained by a weighted sum (over electrons) of each feature, with the sum weights used by electron  $i$  given by the  $i^{\text{th}}$  row of the attention matrix. In this way, each electron is *attending* to all of the other electrons in a way which depends upon the similarity of their features in a (learned) latent space. Loosely, self-attention is nothing but a linear transformation of the inputs where the parameters of the transformation are themselves functions of those inputs. In multi-head attention, each head represents a different latent space in which the similarities are measured, so electrons will attend to different features in each head. Self-attention, like the message-passing mechanism in the FermiNet, is agnostic to the labelling of electrons. Exchanging any two rows of  $\mathbf{H}_l$  (i.e., exchanging the spin and spatial co-ordinates of any two electrons) simply exchanges the resulting rows of  $\mathbf{F}_l$ .

The outputs from the multi-headed self-attention layer are subject to a ‘layer normalization’ operation,

$$\text{LayerNorm}(\mathbf{x}) = \frac{\kappa(\mathbf{x} - \bar{x})}{\sqrt{\text{Var}(\mathbf{x}) + \epsilon}} + \delta,\tag{3.54}$$

where  $\epsilon$  is a small, positive constant that prevents the inverse square root from diverging when  $\text{Var}(x) \rightarrow 0$ , and  $\kappa$  and  $\delta$  are variational parameters. This operation is understood to act element-wise on the  $n_x$  entries of  $\mathbf{x}$ , with

$$\bar{x} = \frac{1}{n_x} \sum_{i=1}^{n_x} x_i\tag{3.55}$$

| Symbol                                            | # Parameters                                                                           | Description                                                   |
|---------------------------------------------------|----------------------------------------------------------------------------------------|---------------------------------------------------------------|
| $N$                                               |                                                                                        | Number of electrons                                           |
| $n_{\text{det}}$                                  |                                                                                        | Number of Pfau-Spencer determinants                           |
| $n_{\text{atom}}$                                 |                                                                                        | Number of atomic nuclei                                       |
| $n_{\alpha}$                                      |                                                                                        | Number of $\alpha \in \{\uparrow, \downarrow\}$ electrons     |
| $L$                                               |                                                                                        | Number of self-attention layers (depth)                       |
| $w_l$                                             |                                                                                        | Width of single-particle features $\mathbf{h}_{l,i}^{\alpha}$ |
| $w_{h,l}$                                         |                                                                                        | MLP hidden layer width                                        |
| $\mathbf{Q}_{lh}/\mathbf{K}_{lh}/\mathbf{V}_{lh}$ | $3 \sum_{l=1}^L w_{l-1} w_l$                                                           | Query/key/value weight matrices                               |
| $\mathbf{W}_l$                                    | $\sum_{l=1}^L w_l^2$                                                                   | Weights for head projection                                   |
| $\text{MLP}_l$                                    | $\sum_{l=1}^L 2w_l w_{h,l} + w_{h,l} + w_l$                                            | MLP parameters                                                |
| $\kappa_l$                                        | $2L$                                                                                   | LayerNorm multipliers                                         |
| $\delta_l$                                        | $2L$                                                                                   | LayerNorm offsets                                             |
| $\beta$                                           | 1                                                                                      | Jastrow factor decay parameter                                |
| Block determinants                                |                                                                                        |                                                               |
| $\mathbf{U}^{k\alpha}$                            | $\sum_{\alpha \in \{\uparrow, \downarrow\}} n_{\text{det}} n_{\alpha} w_L$             | Weights for orbital projection                                |
| $\mathbf{d}^{k\alpha}$                            | $\sum_{\alpha \in \{\uparrow, \downarrow\}} n_{\text{det}} n_{\alpha}$                 | Biases for orbital projection                                 |
| $\pi_{im}^{k\alpha}$                              | $\sum_{\alpha \in \{\uparrow, \downarrow\}} n_{\text{det}} n_{\text{atom}} n_{\alpha}$ | Atomic envelope pre-factors                                   |
| $\sigma_{im}^{k\alpha}$                           | $\sum_{\alpha \in \{\uparrow, \downarrow\}} n_{\text{det}} n_{\text{atom}} n_{\alpha}$ | Atomic envelope decay coefficients                            |
| Dense determinants                                |                                                                                        |                                                               |
| $\mathbf{U}^{k\alpha}$                            | $2n_{\text{det}} N w_L$                                                                | Weights for orbital projection                                |
| $\mathbf{d}^{k\alpha}$                            | $2n_{\text{det}} N$                                                                    | Biases for orbital projection                                 |
| $\pi_{im}^{k\alpha}$                              | $2n_{\text{det}} n_{\text{atom}} N$                                                    | Atomic envelope pre-factors                                   |
| $\sigma_{im}^{k\alpha}$                           | $2n_{\text{det}} n_{\text{atom}} N$                                                    | Atomic envelope decay coefficients                            |

Table 3.2: Summary of symbols and parameters used in defining the PsiFormer architecture.

and

$$\text{Var}(\mathbf{x}) = \frac{1}{n_x} \sum_{i=1}^{n_x} (x_i - \bar{x})^2, \quad (3.56)$$

is the variance of the entries of  $\mathbf{x}$ . Layer normalization keeps fixed the mean and variance of the values at each layer, preventing gradients with respect to particular weights from becoming very large or very small. This is particularly important in later layers of the network [89].

The normalized features are updated by an MLP with a single hidden layer, combined with another residual connection and a final layer normalization. This series of transformations defines the ‘Transformer’ architecture [85] – now famous for its role in large language models. The full update rule for the electron features  $\mathbf{h}_{l,i}$  can be written,

$$\begin{aligned} \mathbf{g}_{l,i} &= \text{LayerNorm}_{l,1}(\text{MHA}_l(\mathbf{h}_{l,i}) + \mathbf{h}_{l,i}), \\ \mathbf{h}_{l+1,i} &= \text{LayerNorm}_{l,2}(\text{MLP}_l(\mathbf{g}_{l,i}) + \mathbf{g}_{l,i}). \end{aligned} \quad (3.57)$$

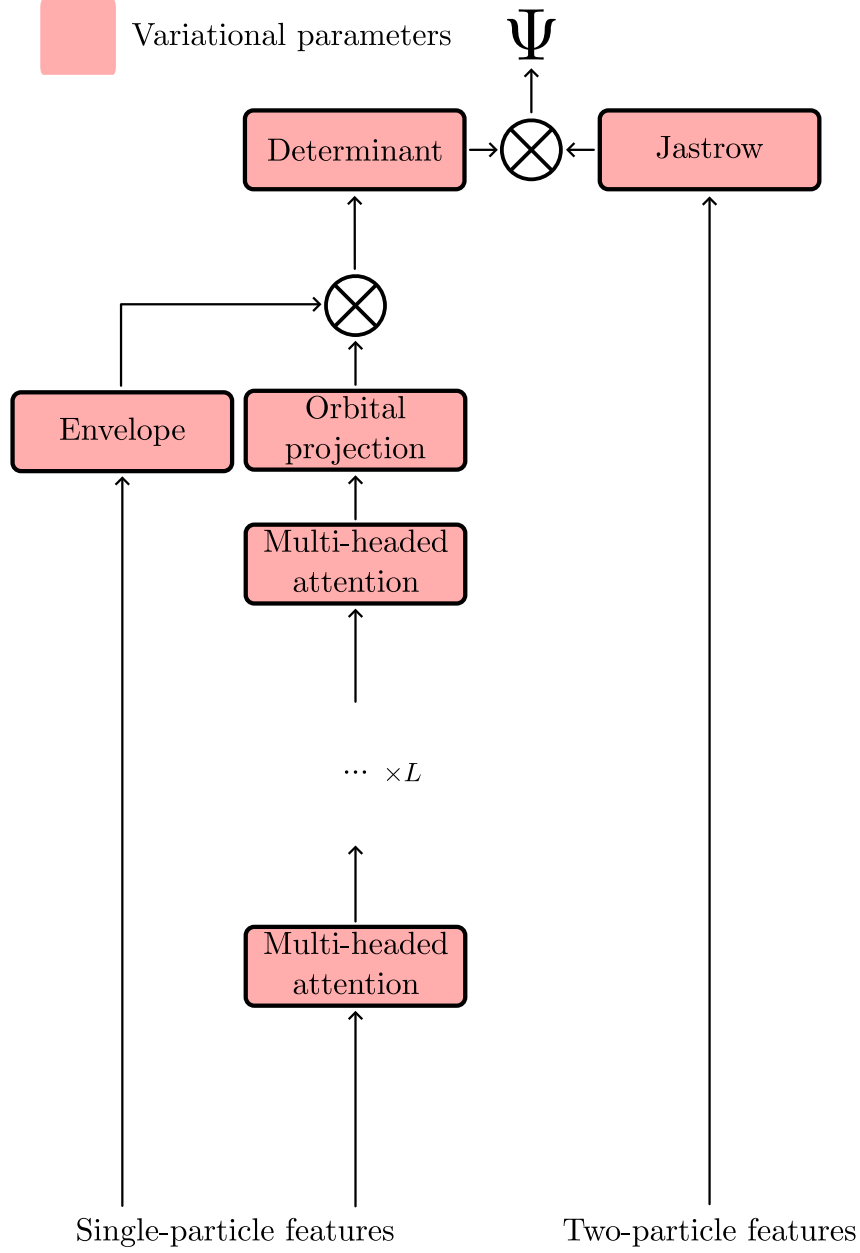


Figure 3.3: Schematic depiction of the PsiFormer architecture, mapping input features derived from the electron co-ordinates to the many-body wavefunction. The two-particle features are not included in the self-attention mechanism, which mixes information between particles. Instead, the cusp conditions are satisfied using a simple Jastrow factor. Figure adapted from [84].

This architecture is depicted schematically in fig. 3.3. After  $L$  Transformer layers, the electronic features are projected into orbitals and shaped into Pfau-Spencer determinants by exactly the same procedure as the FermiNet, see eqs. (3.46) to (3.48). To satisfy the Kato cusp conditions – given that the PsiFormer architecture thus described does not incorporate the magnitudes of the separations between electrons – we also introduce an explicit Jastrow factor. The final wavefunction is then written,

$$\Psi(\mathbf{r}) = \exp(J(\mathbf{r})) \sum_k \det \left[ \phi_i^{k\uparrow}(\mathbf{r}_j^\uparrow; \{\mathbf{r}_{/j}\}) \right] \det \left[ \phi_i^{k\downarrow}(\mathbf{r}_j^\downarrow; \{\mathbf{r}_{/j}\}) \right], \quad (3.58)$$

for block diagonal determinants, or

$$\Psi(\mathbf{r}) = \exp(J(\mathbf{r})) \sum_k \det \left[ \phi_i^{k\alpha}(\mathbf{r}_j^\alpha; \{\mathbf{r}_{/j}\}) \right] \quad (3.59)$$

for dense determinants. A simple Jastrow factor,  $J(\mathbf{r})$ , is chosen to explicitly satisfy the Kato cusp conditions,

$$J(\mathbf{r}) = \sum_{i \neq j} \frac{-\gamma\beta^2}{\beta + r_{ij}}, \quad r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|, \quad (3.60)$$

where  $\gamma = 0.25$  for pairs of electrons with the same spin,  $\gamma = 0.5$  for pairs of electrons with opposing spin, and  $\beta$  is a variational parameter.

The full PsiFormer architecture, mapping the input features to the wavefunction, is depicted in fig. 3.3. The symbols used in defining the PsiFormer architecture are summarized in table 3.2, including the numbers of variational parameters associated with each component in the network. Assuming typical values for the parameters of the network (four heads of width 64,  $w_{l>0} = 256$ ,  $w_{h,l} = 256$ ,  $L = 2$ ,  $n_{\text{det}} = 16$ ) and dense determinants, a PsiFormer wavefunction for a bicyclobutane molecule ( $n_{\text{atom}} = 10$ ,  $N = 30$ ,  $w_0 = 4n_{\text{atom}} + 1$ ) contains 895497 variational parameters.

### 3.3.5 Implementation of NNVMC

In addition to the theoretical framework of the VMC method, and the architectures described in the previous section, there are a number of numerical considerations to be made in the implementation of the NNVMC method. In this section we describe the particulars of the numerical implementation of NNVMC used in the later chapters of this thesis. An open source version of this implementation is available online.<sup>14</sup>

As we discussed in section 3.3.1, the mathematical operations required to evaluate neural networks are extremely well suited to execution on GPUs. In the present

<sup>14</sup><https://github.com/google-deepmind/ferminet>

work, we implement NNVMC using the JAX deep learning framework for the Python programming language [90]. JAX offers three important features: just-in-time compilation of code written in standard Python to machine code executable on GPUs via the XLA compiler, automatic batching of calculations on and between GPU devices, and automatic differentiation.

Automatic differentiation allows the evaluation of the derivatives of functions in a computer program. This is achieved by composing, via the chain rule, the derivatives of the elementary operations that constitute the function. Unlike finite-difference approaches, derivatives computed by automatic differentiation are numerically exact up to computer precision.

MCMC sampling of the wavefunction and the energy gradients is trivially parallelizable across multiple GPUs; however, this only affords the ability to reduce statistical noise by accumulating a greater number of MCMC samples. Training by gradient descent is fully sequential, and cannot be further accelerated by increased parallelization after sufficient statistics are already being acquired. Parallelizing across multiple devices becomes necessary for larger systems, or wavefunction ansätze containing more parameters, as limitations on device memory restrict the number of samples that can be taken in parallel per device.

Where possible, all calculations involving the wavefunction are carried out in the logarithmic domain, because the value of the wavefunction varies by many orders of magnitude throughout the  $3N$ -dimensional configuration space. The determinants appearing in eqs. (3.48) and (3.58) are evaluated in the logarithmic domain, and the kinetic energy term in the local energy is written,

$$K_L(\mathbf{r}) = -\frac{1}{2} \sum_{i=1}^N \sum_{d \in \{x,y,z\}} \left[ \frac{\partial^2 \log |\Psi(\mathbf{r})|}{\partial r_{id}^2} + \left( \frac{\partial \log |\Psi(\mathbf{r})|}{\partial r_{id}} \right)^2 \right], \quad (3.61)$$

where  $r_{id}$  indicates the spatial co-ordinate of the  $i^{\text{th}}$  electron along the Cartesian axis,  $d$ . The gradient terms in the kinetic energy are evaluated via automatic differentiation. Note that the gradients of the logarithm of the wavefunction diverge where  $|\Psi| = 0$ . This is a simple consequence of the definition of the local energy. However, these divergences do not make infinite contributions to the energy expectation value because  $|\Psi|$ , and therefore  $p$ , vanish at the nodes by definition.

Numerical evaluation of the Coulomb potential term in the local energy is also challenging. Here, the local energy diverges in regions of configuration space where  $|\Psi|$  is non-zero. As we have discussed, this is mitigated by satisfying the Kato cusp conditions on the wavefunction. However, because the FermiNet and PsiFormer rely on *learning* the Kato cusp conditions, they may not be exactly satisfied during the course of optimization. Furthermore, due to fixed-precision floating point errors, we

cannot rely on the exact cancellation of the kinetic and potential contributions to the energy at the cusps. Instead, we mitigate these numerical errors by ‘clipping’ the local energy. We compute the median, and the mean absolute deviation from the median, of the local energy over a batch of MCMC samples. Samples of the local energy which lie more than some number of standard deviations (usually chosen as five standard deviations) away from the median are clipped to this threshold. Clipping with respect to the median is preferable to clipping with respect to the mean as the former is more robust to large outliers. The mean of the energy is then calculated with respect to these clipped values. This procedure biases evaluations of the energy expectation value; the divergence of the Coulomb interaction is physical reality. In the course of optimization (especially in the presence of noise resulting from small batches of MCMC walkers) the impact of this bias is negligible; however, the unclipped local energy evaluations must be used when computing the final expectation value of the ground-state energy.

The parameters of neural networks are usually initialized to random values drawn from a uniform or normal distribution [91, 92]. With such a random initialization, the neural network trial wavefunction is a very poor approximation to the ground-state wavefunction.<sup>15</sup> This results in a large sample variance for the local energy and its gradients. Under these conditions, optimization with a few MCMC samples can be numerically unstable. This increases the likelihood of sampling regions of parameter space where the Pfau-Spencer determinants become singular.

A better starting point for optimization is obtained by ‘pre-training’ the Pfau-Spencer orbitals to match Hartree-Fock orbitals. The optimization objective in the pre-training phase is not local energy minimization, but the minimization of the sum of squared differences between the Pfau-Spencer orbitals and Hartree-Fock orbitals,

$$\mathcal{L}_{\text{pre}}(\theta) = \int d\mathbf{r} p_{\theta}(\mathbf{r}) \left[ \sum_{\alpha \in \{\uparrow, \downarrow\}} \sum_{ijk} (\phi_{i,\theta}^{k\alpha}(\mathbf{r}_j^{\alpha}; \{\mathbf{r}_{/k}\}) - \phi_{\text{HF},i}^{\alpha}(\mathbf{r}_j^{\alpha}))^2 \right], \quad (3.62)$$

where  $\phi_{\text{HF},i}^{\alpha}$  are the Hartree-Fock orbitals. The Hartree-Fock orbitals are obtained in a minimal basis set, as we are only trying to achieve an approximate initialization for the parameters  $\theta$ . Note that the original FermiNet paper deviates from this scheme, sampling the pre-training loss from a mixed distribution incorporating the density of the Hartree-Fock wavefunction [78]. The pre-training objective is minimized using ADAM [93], a variant of stochastic gradient descent, eq. (3.18), which incorporates momentum.

In the course of (pre-)training, samples from  $p_{\theta}$  are generated via MCMC sam-

---

<sup>15</sup>In molecular systems, the exponentially decaying isotropic envelope alleviates this problem somewhat: the randomly initialized orbitals are still approximately hydrogenic.

pling, as discussed in section 3.1.1. We utilize the Metropolis algorithm [27] (see appendix B), with a  $3N$ -dimensional isotropic Gaussian proposal distribution. The width of the proposal distribution is tuned periodically during training to maintain a  $\sim 50\%$  acceptance rate: every  $t_m$  parameter update steps, the proposal width is increased by 5% if the acceptance rate is above 55%, and decreased by 5% if the acceptance rate is below 50%. Multiple MCMC moves, typically  $\sim 10$ , are made between parameter updates to reduce autocorrelation between successive energy and gradient samples. The initial MCMC walker positions are drawn from isotropic Gaussian distributions around the atomic nuclei, with electrons assigned to atomic nuclei according to their spin according to Hund's rules for the electronic structure of the constituent atoms.

We optimize the wavefunction using the Kronecker-Factored Approximate Curvature (KFAC) algorithm [94], an approximation to natural gradient descent. The natural gradient descent update rule is,

$$\text{update}(\theta_{\text{old}}, \nabla_{\theta} \langle E_L \rangle_{p_{\theta}}) = -\eta \mathcal{F}_{\theta}^{-1} \nabla_{\theta} \langle E_L \rangle_{p_{\theta}} \quad (3.63)$$

where  $\eta$  is the learning rate and  $\mathcal{F}_{\theta}^{-1}$  is the inverse of the Fisher information matrix,

$$\mathcal{F}_{\theta} \propto \left\langle \frac{\partial \log |\Psi_{\theta}|}{\partial \theta_i} \frac{\partial \log |\Psi_{\theta}|}{\partial \theta_j} \right\rangle_{p_{\theta}}. \quad (3.64)$$

Inverting this matrix for the possibly millions of parameters in a neural-network wavefunction ansatz is impossible. The KFAC algorithm makes two simplifying assumptions to produce a tractable approximation to the inverse of the Fisher matrix. First, the covariance between parameters in different layers of the network is assumed to be zero, bringing the Fisher matrix into block diagonal form. Second, these diagonal blocks are approximated by recognizing that the blocks corresponding to a linear transformation can be written as,

$$\left\langle \frac{\partial \log |\Psi_{\theta}|}{\partial \mathbf{W}_l} \frac{\partial \log |\Psi_{\theta}|}{\partial \mathbf{W}_l}^T \right\rangle_{p_{\theta}} = \langle (\mathbf{a}_l \otimes \mathbf{e}_l)(\mathbf{a}_l \otimes \mathbf{e}_l)^T \rangle_{p_{\theta}}, \quad (3.65)$$

where  $\otimes$  indicates the Kronecker product,  $\mathbf{W}_l$  is the matrix of weights in the linear layer, understood to be flattened into a vector,  $\mathbf{a}_l$  are the inputs to the linear transformation, and  $\mathbf{e}_l$  are the gradients of  $\log |\Psi_{\theta}|$  with respect to the entries of the flattened weight matrix,

$$\mathbf{e}_l = \frac{\partial \log |\Psi_{\theta}|}{\partial W_{l,i}}. \quad (3.66)$$

The blocks of the inverse Fisher matrix corresponding to the linear layers are then

approximated as,

$$\langle (\mathbf{a}_l \otimes \mathbf{e}_l)(\mathbf{a}_l \otimes \mathbf{e}_l)^T \rangle_{p_\theta}^{-1} = \langle \mathbf{a}_l \mathbf{a}_l^T \rangle_{p_\theta}^{-1} \otimes \langle \mathbf{e}_l \mathbf{e}_l^T \rangle_{p_\theta}^{-1}. \quad (3.67)$$

The resulting Kronecker product of small, inverted matrices is cheaper to evaluate than the inverse of a single much larger Kronecker product matrix. To justify this approximation, we defer to the inventors of KFAC [94],

The expectation of a Kronecker product is, in general, not equal to the Kronecker product of expectations, and so this is indeed a major approximation to make, and one which likely won't become exact under any realistic set of assumptions, or as a limiting case in some kind of asymptotic analysis. Nevertheless, it seems to be fairly accurate in practice[...]

KFAC incorporates a number of additional numerical tricks to stabilize optimization: an exponential moving average of the covariances on the right-hand side of eq. (3.67) is accumulated between parameter updates, and utilized in place of the estimate from a single batch of MCMC samples; a damping term,  $\lambda$ , is added to the diagonal entries of the Fisher information matrix to prevent singularity,

$$\mathcal{F} \rightarrow \mathcal{F} + \lambda \mathbf{I}; \quad (3.68)$$

and the norm of the final gradient update is constrained by multiplying eq. (3.63) by a scalar such that,

$$(\eta \mathcal{F}_\theta^{-1} \nabla_\theta \langle E_L \rangle_{p_\theta})^T (\eta \nabla_\theta \langle E_L \rangle_{p_\theta}) \leq C, \quad (3.69)$$

where  $C$  is a parameter known as the ‘norm constraint’.

The KFAC optimizer has been shown to yield much faster convergence to the ground state for neural network ansätze [78], compared to more commonly used first-order stochastic optimizers for neural networks like ADAM [93]. We utilize an open-source JAX implementation of the KFAC optimizer.<sup>16</sup>

The learning rate,  $\eta$ , is adjusted according to a learning-rate schedule during training. In early training iterations, the learning rate is large to help avoid local minima and saddle points, and rapidly progress toward regions of parameter space where the wavefunction is sensible; we do not care much in the early optimization steps about carefully navigating the energy landscape. Later in training, we require a much smaller learning rate to ensure stable convergence to the minimum. We

---

<sup>16</sup><https://github.com/google-deepmind/kfac-jax>

utilize the simple learning rate schedule,

$$\eta(t) = \left( \frac{\eta_0}{1 + \frac{t}{t_0}} \right)^\gamma, \quad (3.70)$$

where  $t$  labels the training step,  $\eta_0$  is the initial learning rate,  $t_0$  is a ‘delay’ parameter, and  $\gamma$  is a ‘decay’ parameter.

Training is carried out for a fixed number,  $T$ , of gradient-descent update steps. After training, inference is performed by accumulating many batches of MCMC evaluations of the local energy – typically  $\sim 10^4$  batches. With a batch size of  $\sim 10^3$  we obtain a statistical estimate of the final energy of the trial wavefunction from  $\sim 10^7$  point estimates of the integrand. Of course, more evaluations can be obtained to reduce the statistical error in the final energy if required.

Sequential evaluations of the local energy are autocorrelated due to our choice of a local proposal distribution in the Metropolis algorithm. These correlations reduce the effective sample size for our final energy evaluation, such that a naive evaluation of the standard error in the final energy will produce too-small a value (fundamentally, due to the fact that our samples are not *i.i.d.*).

The standard error can be evaluated correctly from a set of temporally correlated samples can be evaluating via the ‘re-blocking’ method [95]. The original time series (where time is measured in steps on the Markov chain) is ‘coarse grained’ by being histogrammed into temporal bins. The standard error evaluated using the binned values will be larger than the standard error evaluated over the original data for an autocorrelated time series. The binning procedure is repeated for an exponentially increasing sequence of bin widths. The standard error evaluated with respect to the coarse-grained time series converges to the correct value (according to the true variance of the underlying distribution) when the bin width exceeds the autocorrelation time of the Markov chains. We apply the re-blocking method to the local energy samples obtained during inference to estimate the standard error in our estimate of the final energy.

In the preceeding discussion, we have introduced (or implicitly defined) a large number of ‘hyperparameters’ – parameters of the learning process, such as the number of parameters in the network, or the learning rate. The ultimate accuracy of a NNVMC calculation is a function of the hyperparameters. Choosing the best set of hyperparameters is itself a highly complex optimization problem. However, solving this optimization problem programatically is currently too expensive in the context of NNVMC. Instead, we rely on intuition (and luck) to choose appropriate hyperparameters. The hyperparameters of our implementation of NNVMC are listed in table 3.3, alongside typical values according to Pfau [78] and Spencer [79].

While the asymptotic computational complexity of NNVMC is lower than that

| Hyperparameter                      | <i>Symbol</i> | Typical value      |
|-------------------------------------|---------------|--------------------|
| Batch size (Number of MCMC walkers) | —             | 4096               |
| Training iterations                 | $T$           | $3 \times 10^6$    |
| Pre-training iterations             | —             | $1 \times 10^3$    |
| Initial learning rate               | $\eta_0$      | $5 \times 10^{-2}$ |
| Learning rate delay                 | $t_0$         | $1 \times 10^4$    |
| Learning rate decay                 | $\gamma$      | 1                  |
| KFAC norm constraint                | $C$           | $1 \times 10^{-3}$ |
| KFAC damping                        | $\lambda$     | $1 \times 10^{-3}$ |
| MCMC moves per parameter update     | —             | 10                 |
| MCMC width update period            | $t_m$         | 100                |

Table 3.3: Summary of hyperparameters used in our implementation of the NNVMC method. We have omitted the architectural hyperparameters of the ansätze from this list; these are summarized separately in tables 3.1 and 3.2.

of the accurate quantum chemistry methods with which it competes, there is a large pre-factor due to the expense of evaluating the neural network and its derivatives, and the lengthy optimization required to converge ansätze with so many parameters. At the time of writing, obtaining the energy of a large molecule ( $N \simeq 50$ ) takes around a week utilizing 16 A100 GPUs, and the largest system studied with NNVMC contains  $N = 108$  electrons [96]. The current state of affairs is cleanly summarized in [82]:

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

Improvements in the computational efficiency of NNVMC are likely to be focused in the following areas: cheaper wavefunction ansätze (either through network design or improved implementation) that do not sacrifice much, if any, accuracy or retain qualitative accuracy in some area of interest; approaches that greatly simplify or do not require the evaluation of determinants (in the spirit of Vandermonde determinants or sortlets) or simplify the evaluation of the Laplacian; improved optimization methods that reduce the number of gradient evaluations required to reach the ground state; and sampling methods that improve the efficiency of MCMC sampling.

Dramatic improvements in the efficiency of evaluating the Laplacian can be obtained by altering how gradients are propagated through the network during automatic differentiation and exploiting the sparsity of intermediate calculations. This technique, known as the ‘forward Laplacian’, and first proposed by Li *et al.* [97], alleviates the need to evaluate the full Hessian (whose trace yields the Laplacian) of the wavefunction with respect to the electron co-ordinates. Easy-to-use, open-source implementations of the forward Laplacian are now available<sup>17,18</sup>, and the technique is in common usage. Adopting the forward Laplacian commonly reduces the wall time per gradient step in a NNVMC calculation by a factor of  $\sim 2$ . Li *et al.* also modify the PsiFormer architecture to further exploit the efficiency of sparse Laplacian calculations. They introduce additional sparsity into the structure of the intermediate gradients in the neural network layers. The resulting ansatz, LapNet, reduces the wall time per gradient step by a factor of  $\sim 5$  ( $\sim 23$ ) for a polyethylene chain containing 10 monomers versus a PsiFormer calculation with (without) the forward Laplacian.

Stochastic reconfiguration/natural gradient descent is more effective than stochastic gradient descent (and variants incorporating momentum, such as ADAM) for optimizing neural network wavefunctions [78]. However, straightforward application of stochastic reconfiguration requires inverting a matrix with linear dimension equal to the number of parameters in the neural network. Typical parameter counts in neural network ansätze number in the millions, making such a straightforward application a computational impossibility. The most commonly used approximation to stochastic reconfiguration is the KFAC algorithm. However, it has recently been demonstrated that a numerically exact formulation of stochastic reconfiguration over a limited number of samples can be obtained via some straightforward algebraic manipulations, where the preconditioner has a linear dimension equal to the number of samples taken [98, 99]. The resulting algorithm, MinSR, is highly effective in NQS calculations. Unfortunately, this formulation is not well suited for NNVMC, where the configuration space is infinite and the variance in the local energy is typically much larger due to the divergences in the Coulomb potential. The main difficulty is that the estimate of the preconditioner obtained over a limited number of samples is not sufficiently accurate to ensure numerically stable optimization. A proposed extension to MinSR, SPRING [100], seeks to alleviate this issue by introducing a scheme that maintains an exponential moving average of the preconditioner (similar in spirit to the covariance averaging in KFAC) over a number of gradient update steps. The authors claim that SPRING yields faster convergence and asymptotically more accurate ground-state energies than KFAC [100].

---

<sup>17</sup><https://github.com/microsoft/folx>

<sup>18</sup><https://github.com/YWolfeee/lapjax>

A distinct approach is proposed by Drissi *et al.* [101], who reformulate the VMC method in a game theoretic framework and propose a novel second-order optimizer based on decision geometry. In their scheme, the energy expectation value becomes analogous to a generalized entropy for a probability distribution defined by a given wavefunction, and the geometry in parameter space is subsequently described in terms of cross-entropies between distributions described by different parameters. The resulting optimization algorithm, ‘decisional gradient descent’, yields rapid and numerically stable convergence to the ground-state energy for a simple model Hamiltonian; however, this method and framework are yet to be extended to the treatment of molecular systems.

The pseudopotential method is another common strategy for reducing the computational cost of quantum chemistry methods applied to molecules or solids containing heavy atoms. The most tightly bound atomic orbitals, occupied by the ‘core electrons’, are removed from the calculation. The Coulomb interaction with the atomic nucleus and the core electrons is replaced by a pseudopotential designed to mimic its effect on the ‘valence electrons’ (those occupying the least tightly bound atomic orbitals). This approach has been demonstrated in NNVMC calculations of the ground-state energy of 3d transition metal atoms [102], yielding dramatic speedups for atoms where most electrons can be removed. For example, the time per training iteration is reduced by around two orders of magnitude for the gallium atom.

### 3.3.6 Applications of NNVMC

The publication of the FermiNet / PauliNet, and subsequently the PsiFormer papers was followed by a flurry of activity aiming to expand the scope of application of NNVMC to a variety of common problems in quantum chemistry and electronic structure theory.

#### NNDMC

Given the traditional perspective from SJB-VMC calculations, it is natural to ask if the accuracy of NNVMC can be improved by using optimized neural network wavefunctions as trial wavefunctions in DMC calculations. This question was explored by Wilson *et al.* [103], using a FermiNet with smaller layers than used in the original work of Pfau *et al.* [78], for VMC calculations of the ground-state energy of first-row atoms. Wilson demonstrated that, while NNDMC does slightly improve the variational energy for first-row atoms, the improved ground-state energies were not better than those obtained by Pfau *et al.* with larger FermiNet layers at the VMC level. The inability of NNDMC to obtain more accurate ground-state ener-

gies than NNVMC suggests that neural network wavefunction ansätze produce very accurate approximations to the ground-state wavefunction away from the nodal surface, and that the remaining error for first-row atoms is dominated by inaccuracies in the nodal surface. This seems reasonable, as the trial wavefunction is infrequently sampled near the nodal surface in the course of NNVMC training. The nodal error cannot be reduced by a subsequent FN-DMC calculation, which is not, therefore, worth the additional computational effort in small systems.

A similar work by Ren *et al.* [104] suggests, however, that NNDMC calculations do offer some value in refining the ground-state energy of larger molecules. Like Wilson *et al.*, they see that FN-DMC can improve the accuracy of a FermiNet calculation with fewer parameters for first-row atoms, but do not obtain energies better than the NNVMC results of Pfau *et al.* However, they go on to study the dissociation of the nitrogen dimer with FN-DMC with a NNVMC trial wavefunction, and find that, in this case, FN-DMC offers reductions in the ground state energy on the order of  $\sim 1\text{--}2$  milliHartrees in the dissociative region of the bond length of the dimer. At this geometry, the ground state of the nitrogen dimer is known to be strongly multi-reference. These results suggest that FN-DMC with a NNVMC trial wavefunction can offer a small improvement over NNVMC alone when describing the ground-state wavefunction is more challenging.

## Periodic boundary conditions

Several extensions of NNVMC to calculations of the ground-state properties of materials, utilizing small simulation cells subject to periodic boundary conditions (see section 4.1), were published shortly after the publication of the PauliNet and FermiNet papers. Cassella *et al.* [105] extend the FermiNet to periodic boundary conditions and study the homogeneous electron gas, improving upon SJB-DMC benchmark calculations for a simple cubic simulation cell containing 14 electrons. Further, they demonstrate the ability of neural network wavefunction ansätze to converge spontaneously to the appropriate ground state either side of the Wigner transition for a body-centered-cubic simulation cell containing 27 electrons (see Chapter 3). Wilson *et al.* [106] adopt a PauliNet-like architecture with heavily modified input features, explicit Jastrow factors, and backflow transformations applied to Hartree-Fock orbitals in the determinant. They study the homogeneous electron gas, obtaining similar accuracy to Cassella *et al.* for a simple cubic simulation cell containing 14 electrons. Li *et al.* [96] apply the FermiNet to the study of the homogeneous electron gas and a number of atomic solids, providing a proof-of-concept of the application of NNVMC to real solids, yielding ground-state energies competitive with SJB-DMC calculations – although, their ground-state energies for the homogeneous electron gas fall short of their SJB-DMC benchmarks by  $\sim 2$  milliHartrees. The same au-

thors applied their periodic neural network wavefunction parameterization to the calculation of electric polarization [107], forces and stresses in solids [108, 109], demonstrating that optimized neural network trial wavefunctions produce accurate estimates of observables other than the ground-state energy.

A later work on the homogeneous electron gas by Pescia *et al.* [110] utilizes a PauliNet-like ansatz with message-passing neural network backflow and Jastrow terms applied to a Slater determinant of plane wave orbitals and body-centered-cubic site-centered Gaussian orbitals. Slater determinants in these bases are already excellent approximations to the ground-state wavefunction in the dense (plane-wave orbitals) and diffuse (Gaussian orbitals) limits, allowing Pescia *et al.* to exceed the accuracy of SJB-DMC calculations with around two orders of magnitude fewer parameters than related works.

Pescia *et al.* have also applied NNVMC to the study of bosonic systems in periodic boundary conditions [111]. Their bosonic ansatz, based on the ‘Deep Sets’ neural network architecture [112], is equivalent to a neural network Jastrow factor, maintaining the necessary bosonic permutation symmetry. They study  $^4\text{He}$  atoms confined to a single spatial dimension interacting via a model potential, and obtain close agreement with pure DMC calculations. Additionally, they study bosons in one- and two-dimensions interacting via a Gaussian potential. In one dimension, the ground-state energies obtained are in extremely close agreement with pure DMC results; however, in two dimensions, their ground-state energies fall short of the pure DMC benchmarks by 1–2 milliHartrees.

Lou *et al.* [113] extend NNVMC to the study of superfluid ground states in the unitary Fermi gas model. They show that Pfau-Spencer determinants are poor approximators for strongly paired wavefunctions. They then extend the Pfau-Spencer determinant to produce an ‘antisymmetrized geminal product of singlets’, or AGPs, wavefunction, inspired by the mean-field ansatz in the BCS theory of superconductivity. This wavefunction parameterization still adopts the matrix determinant as its antisymmetrizer, but the entries in the matrix are ‘geminals’: pairing orbitals indexed by pairs of electron indices. In their construction of the geminals, Lou *et al.* show that this parameterization is equivalent to an expansion in a factorially large (with respect to particle number) linear combination of Pfau-Spencer determinants, and that it remedies the shortcoming of Pfau-Spencer determinants in describing paired wavefunctions. They obtain clear signatures of superfluid pairing in the unitary Fermi gas, and more accurate ground-state energies than SJB-DMC for the  $N = 66$  particle system, which is known to closely approximate the properties of the thermodynamic limit ( $N \rightarrow \infty$ ) for this model.

## Excited states and finite temperature calculations

While the majority of NNVMC calculations have been applied to finding ground-state wavefunctions, the technique has also been extended to the computation of electronic excited states. Entwistle *et al.* [114] demonstrate that a penalty method can be effectively applied to find the electronic excited states of first row atoms and small to medium-sized molecules by simultaneously optimizing and orthogonalizing an ensemble of PauliNet-based wavefunctions. They find that their method offers good agreement for low-lying excited state energies and transition dipole moments for first row atoms and small molecules, compared to accurate reference methods. Further, they model a conical intersection between the ground state and first excited state electronic potential energy surfaces in the ethylene molecule as the C=C torsion angle is varied. They obtain a variational improvement of  $\sim 10$  milliHartrees over the full potential energy surface compared to reference MR-CI calculations.

The work of Entwistle *et al.* highlights a drawback of NNVMC compared to traditional quantum chemistry methods. In calculating relative energies – dissociation energies, electronic excitation energies, etc. – there is no guarantee in a NNVMC calculation that the energies of the individual systems being subtracted to obtain the relative energy are equally well converged. In systems where it is possible to optimize the wavefunction of either state to an accuracy (relative to the ‘true’ energy of the state) that far exceeds chemical accuracy, this effect can be ignored. Otherwise, the different levels of accuracy to which each state is optimized will introduce a source of uncontrolled, non-variational, error in the relative energy. In traditional quantum chemistry, the same problem can often be neglected due to cancellations of errors that can be carefully manipulated by appropriate choices of basis set. Entwistle *et al.* propose to mitigate the uncontrolled error in NNVMC through a variance matching scheme: here, relative energies are calculated by comparing the energies of the different states at points during the VMC optimization where the variance of the local energy is equal for both wavefunctions. They show that this results in more accurate estimates of relative energies.

A very different approach to electronic excited state calculations was proposed later by Pfau *et al.* [115] The problem is reformulated as finding the ground state of an extended system. This is achieved by defining a separate set of electron coordinates per excited state, parameterizing each state as a neural network wavefunction. Orthogonality of the states is maintained by evaluating a ‘meta’-wavefunction in the form of a many-system Slater determinant, the columns of which consist of each excited state many-body wavefunction evaluated on one of the distinct sets of electron co-ordinates. This is a very unlike penalty methods, where one has to tune a Lagrange multiplier for the penalty term. Additionally, using the same neural

network to model the wavefunctions of all the states simultaneously may mitigate the source of uncontrolled error in the work of Entwistle *et al.* This is corroborated by the excellent agreement between the excitation energies obtained by Pfau *et al.* and experimental values, without the need for variance matching.

Treating quantum systems at finite temperature is a challenging problem in the framework of NNVMC. At constant temperature, the natural approach to consider is variational minimization of the Helmholtz free energy, formulated in terms of the density matrix  $\rho$ ,

$$F[\rho] = \text{tr}(\rho\mathcal{H}) + \frac{1}{\beta}\text{tr}(\rho \log \rho), \quad (3.71)$$

where  $\beta$  is the inverse temperature. Unfortunately, the second term in the above expression, corresponding to the entropy, is very difficult to compute without reference to a set of orthonormalized many-body basis states. Within the typical NNVMC framework we have presented thus far, even given access to a robust method for the computation of excited states, it is difficult to obtain more than a few orthogonal states, and harder still to handle their normalization (which can usually be ignored in ground-state NNVMC calculations).

Xie *et al.* propose to tackle this problem by applying an orthonormality-preserving backflow transformation to the particle co-ordinates of a set of reference states that are explicitly orthonormal – typically obtained as solutions of the non-interacting Hamiltonian (plane-wave orbitals for the electron gas; quantum harmonic oscillator orbitals in quantum dots) [116, 117]. So long as the transformation enacts a unitary transformation on the set of wavefunctions, orthonormality is preserved. Xie *et al.* parameterize the backflow transformation using a continuous normalizing flow, a class of neural-network-based generative models that sample complex probability distributions by modifying a simple base distribution through the change-of-variables formula [118]. A variational ansatz for the many-body density matrix is then expanded in terms of these orthonormal basis states, combined with another probabilistic model that parameterizes the Boltzmann distribution. They apply this approach to calculating the quasiparticle effective mass in the two-dimensional homogeneous electron gas at finite temperature. It is not possible to make variational comparisons for this quantity, so the results must be interpreted as a prediction to be verified by experiment. However, they obtain ground-state energies exceeding the accuracy of SJ-VMC benchmarks.

In systems containing atomic nuclei, one could imagine obtaining the reference states from a HF or DFT calculation to form the orthonormal basis. This is much more closely aligned with the PauliNet philosophy than the FermiNet philosophy. It remains an open question how one might realize a model that produces orthonormalized states without a set of reference states. Excited states methods do provide

a basis (once appropriately normalized) on which such a finite temperature scheme could be realized, but not the only one.

Xie *et al.* extend their variational free energy approach to the equation of state of warm dense hydrogen under pressure [119]. The many-body density matrix over the protonic degrees of freedom is taken to be approximately diagonal by neglecting the nuclear kinetic energy in the Hamiltonian (the effects of which are estimated classically), reducing the density matrix to a  $3N$ -dimensional Boltzmann distribution over the positions of the  $N$  protons. The proton distribution is parameterized by a probability density modelled via a normalizing flow. Further, it is assumed that the electronic degrees of freedom remain in the ground state (*i.e.*, the energy of the lowest lying electronic excitation is larger than the thermal energy scale). Under these approximations, the problem is reduced to minimizing the variational free energy with respect to the parameters of the protonic distribution and a ground-state wavefunction conditioned on the co-ordinates of protons sampled from the protonic distribution. Xie *et al.* obtain internal energies, entropies, and pressures comparable with other methods (empirical models, *ab initio* molecular dynamics, and other quantum Monte Carlo approaches), but predict a denser equation of state. However, systematic comparison of the accuracy between methods in this case is challenging due to the non-variational nature of the quantities being calculated and the absence of accurate experimental data.

## Transferable wavefunction models

Given the high expense of NNVMC calculations, and the flexibility of neural networks, it is natural to ask if one might amortize the cost of each individual ground state calculation by training a single neural network model to represent the ground-state wavefunctions of multiple molecules. This is strongly reminiscent of the recent trend of training large generative neural network ‘foundation models’: a single model is trained on an extremely large corpus of data to capture features common to all examples in the data space, and then fine-tuned on a small sub-set of the space to produce a more specific generative model.

The foundation model approach is very effective in improving the data efficiency of generative models. If one lacks the necessary quantity of data to train an accurate model for the specific application of interest, much better results can be obtained by fine-tuning a foundation model using the small amount of task-specific data available. The analogy for NNVMC would avoid expending the computational resources necessary to train wavefunctions from scratch for every molecular geometry of interest, obtaining optimized wavefunctions much more efficiently by fine-tuning a foundational model.

There are substantial practical challenges involved in realizing such a model.

Foundation models are typically trained using vast datasets, such as all the text or images on the public internet. Should one expect to have to train on a similarly large number of examples in a chemical or material space to produce an effective foundation model? This would be quite an undertaking. Similarly, how does one decide which molecules or materials to include in the training set? Is it important to be careful about selecting and balancing the kinds of systems used for training? One would hope that given a sufficient breadth and diversity of examples, something about the general nature of electronic wavefunctions could be captured, but it is unclear what constitutes sufficiency in this context. Data problems aside, it is almost equally difficult to devise an appropriate model architecture for representing the wavefunctions of arbitrary systems, containing differing numbers of electrons and nuclei.

Despite the challenges, some authors have made brave initial forays into the world of transferable wavefunction models. Gao and Günnemann proposed PESNet, a FermiNet-like ansatz where the variational parameters are themselves the output of another neural network model that operates on molecular geometries. In principle, this allows the representation of the wavefunctions of arbitrary molecular geometries [120]. They demonstrate that PESNet, trained simultaneously over the full potential energy surface, produces ground-state energies for a variety of systems with accuracy comparable to individual FermiNet calculations carried out at each individual point sampled over the surface. This is equivalent to a  $\sim 40$  times reduction in calculation time over the entire surface. The PESNet approach was later extended by including a surrogate model to fit the potential energy surface during training of the neural network wavefunction model, obviating the need to perform additional inference after training [121]. This decreases the wall time taken to obtain point evaluations of the potential energy surface from hours (performing inference with an optimized trial wavefunction) to milliseconds (evaluating the surrogate model for the potential energy surface). While this approach works well for one- and two-dimensional potential energy surfaces, scaling to larger dimensional potential energy surfaces is challenging due to the necessarily sparse sampling of the nuclear configuration space during training.

In later work, Gao and Günnemann extended their approach to include the possibility of representing the wavefunctions of molecules that differ not only by geometry but also by the number and type of the atomic nuclei [122]. Here, they adopt a FermiNet-like ansatz, Moon, with modifications chosen to enforce size consistency, constructed such that the number of parameters in the ansatz is independent of the number of electrons and nuclei in the system. They generalize Moon across multiple systems simultaneously by defining a separate model, Globe, that outputs the parameters of a Moon-based wavefunction ansatz given a molecular geometry.

PESNet and Globe+Moon exhibit promising results for molecular geometries in the interpolated region of the training set, but struggle to reach even Hartree-Fock level accuracy when extrapolating beyond the training set.

Concurrently, Scherbela *et al.* proposed two novel approaches to transferable wavefunction models. The first, DeepErwin, is inspired by transfer learning methods in the machine learning literature. A PauliNet-like ansatz is trained for several molecules in tandem, with the constraint that  $\sim 75\%$  of the variational parameters are chosen to be the same for all molecules [123]. They find that this ‘weight sharing’ approach accelerates optimization, and achieves equivalent ground-state energies (compared to a non-weight-sharing ansatz for each individual system) after the optimization has converged. Further, they demonstrate that the same rapid convergence can be achieved by re-using the shared weights on unseen molecules. They leverage this approach to explore the potential energy surfaces of a range of systems, obtaining results in close agreement with accurate quantum chemistry methods. In the second approach, they define a transferable wavefunction model that combines embeddings of the many-electron co-ordinates with a featurization of the maximally localized Hartree-Fock molecular orbital coefficients<sup>19</sup> on atoms in the molecule [124]. This approach is distinguished from that of Gao and Günnemann in two main ways: first, the parameters of the neural network backbone are the same for all molecular geometries; and second, the ansatz explicitly incorporates physical priors by depending upon molecular descriptors obtained from a Hartree-Fock calculation. This approach is still distinguished from a backflow approach, in that the Hartree-Fock *orbitals* are not used in the construction of the final wavefunction ansatz. Additionally, the authors explore the improvement in ground-state energies obtained by applying a small additional number of VMC optimization steps on a per-system basis as a ‘fine-tuning’ method, and demonstrate that this can greatly improve the out-of-distribution performance of the generalized model alone. In contrast to PESNet and Globe+Moon, this approach offers better ground-state energies than Hartree-Fock when extrapolating beyond systems in the training set, suggesting that incorporating the physical priors encoded in Hartree-Fock-based orbital descriptors aids generalization.

These initial attempts at transferable wavefunction models show promise as proofs of concept; however, realizing the scale of model required to make NNVMC as efficient as, say, workhorse DFT methods is still a distant dream. Such a project will likely have to wait until substantial computational advances are made in the efficiency of single-system wavefunction models. Furthermore, this kind of undertaking

---

<sup>19</sup>These are the molecular orbital coefficients obtained under a unitary transformation that is chosen to maximize some heuristic measure of the localization of the electronic charge of each molecular orbital.

would require a large academic collaboration to obtain the necessary resources and expertise. Despite the challenges, there is no *a priori* reason to expect that such an approach might not one day yield the kind of deep-learning revolution in quantum chemistry that we are today seeing in the computational sciences.

# Chapter 4

## The homogeneous electron gas

The homogeneous electron gas (HEG) is the prototypical model of the behavior of electrons in metallic solids. The cations of the crystal lattice are imagined to be smeared out into a homogeneous, positively charged background. Thus, we only consider a gas of interacting electrons. This is a simple setting for studying the physics of interacting electrons. However, despite its simplicity, the HEG captures many properties of real metals at low temperatures.

The HEG also exhibits an interaction-driven quantum phase transition. First predicted by Wigner [125, 126], the appropriately named ‘Wigner transition’ occurs due to the scaling, with respect to average electron density, of the ratio between kinetic and potential energy contributions to the ground-state energy. At very high density, the kinetic energy term dominates and the HEG resembles a non-interacting electron gas. At very low density, the Coulomb term dominates, and the electrons assemble into a crystal lattice that maximizes their mutual separation. This ‘Wigner crystal’ has been observed experimentally in two-dimensional electron gases on semiconductor monolayers [127], but is yet to be observed experimentally in three dimensions.

The zero-temperature phase diagram of the HEG can be understood perturbatively in the low- and high-density regimes [128, 129]. However, understanding the phase diagram in the intermediate density regime<sup>1</sup> requires the application of *ab initio* electronic structure methods. Of these, QMC methods offer the most successful description of the zero-temperature phase diagram of the HEG. In a now-famous work, Ceperley and Alder used the FN-DMC method<sup>2</sup> to study the relative stabil-

---

<sup>1</sup>Intermediate being loosely defined here as somewhere between infinite density and the transition density to the Wigner crystal.

<sup>2</sup>In fact, Ceperley and Alder extend the FN-DMC method by ‘releasing’ the nodal surface. In this ‘release-node’ DMC scheme, walkers are allowed to cross the nodal surface of the trial wavefunction, so long as the solution is only propagated for a small imaginary time period after releasing the nodes. This slightly improves the variational bound of FN-DMC, while limiting the exponential accumulation of noise due to the fermion sign problem.

ity of paramagnetic and ferromagnetic Fermi liquid phases, and the Wigner crystal phase, as a function of density [21]. In addition to offering an estimate of the Wigner transition density, they predicted a region of stability for a ferromagnetic Fermi liquid phase – an ‘itinerant ferromagnet’ [130]. Since the pioneering work of Ceperley and Alder, many QMC studies of the ground-state phase diagram of the electron gas have been published, refining the original results. We provide an overview of the evolution of these studies at the end of this chapter.

More recently, the ‘initiator full configuration interaction quantum Monte Carlo’ (i-FCIQMC) method has emerged as a promising approach for describing the ground state of the HEG [131]. i-FCIQMC stochastically samples the exact ground-state wave function via imaginary time evolution, analogous to the principle of pure DMC. This sampling is carried out in a basis of Slater determinants (as opposed to the real-space basis of pure DMC). The accuracy of an i-FCIQMC calculation is limited by the set of one-particle basis functions used to construct many-particle Slater determinant basis; however, this can be systematically converged (or, extrapolated) to the complete basis set limit. This eliminates the variational error in FN-DMC calculations due to the imposition of a fixed nodal surface.<sup>3</sup> Computational expense is the main drawback of the i-FCIQMC method, limiting its application to small numbers of electrons at high densities.

These features – conceptual simplicity, rich correlated physics, and a long history of being studied with wavefunction-based methods – make the HEG an ideal testing ground for applying NNVMC to solid systems. Given the favorable comparisons between NNVMC and SJB-DMC in the study of first- and second-row atoms and small molecules, it is natural to ask if NNVMC can match the accuracy of SJB-DMC applied to the HEG. In this chapter, we will describe how the HEG, and solid-state systems in general, can be treated within the NNVMC framework, outline the implementation of solid-state calculations with FermiNet – including the treatment of periodic boundary conditions – and demonstrate our method by studying the Wigner transition in the HEG. Our results for the Wigner transition demonstrate a novel advantage for neural network wavefunction based methods, namely that the lack of dependence on a basis set allows for a true *ab initio* discovery of quantum phase transitions.

## 4.1 Periodic boundary conditions

When you hold a piece of matter in the palm of your hand – imagine, perhaps, a 3 gram cube of aluminium – you are holding  $\sim 10^{23}$  electrons. Describing the

---

<sup>3</sup>Of course, the pain is conserved by the difficulty of choosing and converging appropriate basis sets.

full many-electron wavefunction of this piece of matter is a computational impossibility. Instead, we argue that the properties of the bulk of the material<sup>4</sup> can be described by considering a much smaller number of electrons, subject to ‘periodic boundary conditions’ (PBCs). So long as the spatial extent of the system subject to PBCs is large compared to the typical correlation length between electrons in the real bulk material, this is a good approximation. Discrepancies between experimental quantities, and quantities obtained under this approximation, are known as ‘finite-size errors’. Finite-size errors are often treated by extrapolation to an *infinite* number of electrons, the so-called ‘thermodynamic limit’, which we consider to be representative of real bulk materials.<sup>5</sup>

We consider a parallelepiped ‘simulation cell’ described by a set of primitive lattice vectors  $\mathbf{a}_1^S, \mathbf{a}_2^S, \mathbf{a}_3^S$ . This simulation cell is immersed in a periodic array of identical replicas of itself. Therefore, particles in the simulation cell interact with images of themselves, separated by ‘simulation cell lattice vectors’,

$$\mathbf{d}_S = l\mathbf{a}_1^S + m\mathbf{a}_2^S + n\mathbf{a}_3^S, \quad l, m, n \in \mathbb{Z}. \quad (4.1)$$

These are associated with a set of simulation cell *reciprocal* lattice vectors defined

$$\mathbf{g}_S = l\mathbf{b}_1^S + m\mathbf{b}_2^S + n\mathbf{b}_3^S, \quad l, m, n \in \mathbb{Z}, \quad (4.2)$$

where

$$\begin{aligned} \mathbf{b}_1^S &= \frac{2\pi}{V} \mathbf{a}_2^S \times \mathbf{a}_3^S, \\ \mathbf{b}_2^S &= \frac{2\pi}{V} \mathbf{a}_3^S \times \mathbf{a}_1^S, \\ \mathbf{b}_3^S &= \frac{2\pi}{V} \mathbf{a}_1^S \times \mathbf{a}_2^S, \end{aligned} \quad (4.3)$$

and

$$V = \mathbf{a}_1^S \cdot (\mathbf{a}_2^S \times \mathbf{a}_3^S) \quad (4.4)$$

is the volume of the simulation cell.

The reciprocal lattice vectors have the useful property that

$$\mathbf{g}_S \cdot \mathbf{d}_S = 2\pi N, \quad N \in \mathbb{Z}. \quad (4.5)$$

Functions that are periodic with respect to the simulation cell lattice can be decomposed as Fourier series in a basis of plane waves with wavevectors corresponding to

---

<sup>4</sup>Leaving the description of the surface for some other unfortunate soul.

<sup>5</sup>Of course, our aluminium cube does not contain a truly infinite number of electrons, but  $10^{23}$  is close enough.

the reciprocal lattice vectors.

Within the simulation cell, atomic nuclei are arranged in a crystalline lattice. The primitive ‘unit cell’ lattice vectors –  $\mathbf{a}_1^P, \mathbf{a}_2^P, \mathbf{a}_3^P$  – relate equivalent atoms. We define unit cell lattice vectors  $\mathbf{d}_P$ , and a unit cell reciprocal lattice spanned by the vectors  $\mathbf{b}_1^P, \mathbf{b}_2^P, \mathbf{b}_3^P$ , in analogy to eqs. (4.1) and (4.3). To preserve crystalline symmetries when working in PBCs, we require that the primitive simulation cell lattice vectors are integer multiples of the primitive unit cell lattice vectors:

$$\mathbf{a}_i^S = N_i \mathbf{a}_i^P, \quad N_i \in \mathbb{N}. \quad (4.6)$$

The molecular Hamiltonian in PBCs, with a simulation cell containing  $N$  electrons and  $M$  nuclei, can be written in a real space basis as

$$\mathcal{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 + U_{\text{e-e}}(\mathbf{r}) + U_{\text{e-n}}(\mathbf{r}, \mathbf{R}) + U_{\text{n-n}}(\mathbf{R}), \quad (4.7)$$

where

$$U_{\text{e-e}}(\mathbf{r}) = \frac{1}{2} \sum_{\{\mathbf{d}_S\}} \sum_{i \neq j}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j - \mathbf{d}_S|}, \quad (4.8)$$

$$U_{\text{e-n}}(\mathbf{r}, \mathbf{R}) = - \sum_{\{\mathbf{d}_S\}} \sum_{i=1}^N \sum_{j=1}^M \frac{Z_j}{|\mathbf{r}_i - \mathbf{R}_j - \mathbf{d}_S|}, \quad (4.9)$$

and

$$U_{\text{n-n}}(\mathbf{R}) = \frac{1}{2} \sum_{\{\mathbf{d}_S\}} \sum_{i \neq j}^M \frac{Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j - \mathbf{d}_S|}. \quad (4.10)$$

The Coulomb sums, eqs. (4.8) to (4.10), are conditionally convergent<sup>6</sup> and slowly decaying. Naive evaluation demands the evaluation of many terms in the series to obtain accurate results. Instead, a much more efficient evaluation is performed via Ewald summation [132–134]: the Coulomb sum is decomposed into short-ranged contributions, given by a rapidly converging sum over lattice vectors, and long-ranged contributions, given by a rapidly converging sum in reciprocal lattice vectors. This is defined by the unique (to within an additive constant) solution of Poisson’s equation for a source in periodic boundary conditions. The potential produced by

---

<sup>6</sup>A sum that is conditionally convergent has a finite limit obtains a different value by rearranging the order in which terms appear in any finite truncation of the sum.

a unit charge at  $\mathbf{r}_i$  in the simulation cell is given by,

$$v_E(\mathbf{r}; \mathbf{r}_i) = \sum_{\{\mathbf{d}_S\}} \frac{\text{erfc}(\gamma^{1/2}|\mathbf{r} - \mathbf{r}_i - \mathbf{d}_S|)}{|\mathbf{r} - \mathbf{r}_i - \mathbf{d}_S|} + \frac{4\pi}{V} \sum_{\{\mathbf{g}_S \neq \mathbf{0}\}} \frac{\exp(-g_S^2/4\gamma)}{g_S^2} \exp(i\mathbf{g}_S \cdot (\mathbf{r} - \mathbf{r}_i)) - \frac{\pi}{V\gamma}. \quad (4.11)$$

Here,  $\gamma$  is a free parameter that balances the rate at which the real- and reciprocal-space sums converge, but does not affect the value of the potential. In addition to inter-particle interactions, we include the energy due to self-interaction between a charge in the simulation cell and its images through the ‘Madelung constant’,

$$v_M = \sum_{\{\mathbf{d}_S \neq \mathbf{0}\}} \frac{\text{erfc}(\gamma^{1/2}d_S)}{d_S} - \frac{2\gamma^{1/2}}{\pi^{1/2}} - \frac{\pi}{V\gamma} + \frac{4\pi}{V} \sum_{\{\mathbf{g}_S \neq \mathbf{0}\}} \frac{\exp(-g_S^2/4\gamma)}{g_S^2}. \quad (4.12)$$

Using the Ewald expressions, eqs. (4.8) to (4.10) become,

$$U_{e-e}(\mathbf{r}) = \frac{1}{2} \sum_{i \neq j}^N (v_E(\mathbf{r}_i; \mathbf{r}_j) - v_M), \quad (4.13)$$

$$U_{e-n}(\mathbf{r}) = - \sum_{i=1}^N \sum_{j=1}^M Z_j v_E(\mathbf{r}_i; \mathbf{R}_j), \quad (4.14)$$

$$U_{n-n}(\mathbf{r}) = \frac{1}{2} \sum_{i \neq j}^M Z_i Z_j (v_E(\mathbf{R}_i; \mathbf{R}_j) - v_M), \quad (4.15)$$

where we have simplified the inclusion of the Madelung constant by invoking the charge neutrality of the simulation cell, which implies that  $q_i = -\sum_{j \neq i} q_j$ .

We choose the commonly adopted value of  $\gamma = (2.8/V^{1/3})^2$  [134, 135], and evaluate the Ewald sums over a  $6 \times 6 \times 6$  grid of lattice points around the origin in real and reciprocal space. We have found that, under these choices, inclusion of additional terms in the sum changes the value by less than the limit of computational precision in our calculations.

#### 4.1.1 Many-body eigenstates under PBCs

The Hamiltonian, eq. (4.7), possesses two important symmetries. It is invariant under the translation of a single electronic co-ordinate by a simulation cell lattice vector, and it is invariant under the simultaneous application of an element of the space group of the crystal lattice to all the electronic co-ordinates. The former

symmetry is an artifact of imposing periodic boundary conditions: in a real solid, electronic interactions break all single-electron symmetries. These symmetries impose some restrictions on the many-body eigenstates of the Hamiltonian.

First we consider the translation of a single electron by a supercell lattice vector. We define a set of translation operators,  $\{\hat{T}^{(i)}(\mathbf{d}_S)\}$ , such that

$$\hat{T}^{(i)}(\mathbf{d}_S)\Psi(\dots, \mathbf{r}_i, \sigma_i, \dots) = \Psi(\dots, \mathbf{r}_i + \mathbf{d}_S, \sigma_i, \dots). \quad (4.16)$$

These operators commute with the Hamiltonian, permitting a set of simultaneous eigenstates,

$$\hat{T}^{(i)}(\mathbf{d}_S)\Psi_\lambda(\mathbf{x}) = \lambda^{(i)}(\mathbf{d}_S)\Psi_\lambda(\mathbf{x}). \quad (4.17)$$

Considering the antisymmetry requirement for fermionic wavefunctions, we see that

$$\begin{aligned} \hat{T}^{(i)}(\mathbf{d}_S)\Psi(\dots, \mathbf{x}_i, \dots, \mathbf{x}_j, \dots) &= \lambda^{(i)}(\mathbf{d}_S)\Psi(\dots, \mathbf{x}_i, \dots, \mathbf{x}_j, \dots) \\ &= -\lambda^{(i)}(\mathbf{d}_S)\Psi(\dots, \mathbf{x}_j, \dots, \mathbf{x}_i, \dots) \\ &= -\hat{T}^{(j)}(\mathbf{d}_S)\Psi(\dots, \mathbf{x}_j, \dots, \mathbf{x}_i, \dots). \end{aligned} \quad (4.18)$$

Thus

$$\hat{T}^{(j)}(\mathbf{d}_S)\Psi(\dots, \mathbf{x}_j, \dots, \mathbf{x}_i, \dots) = \lambda^{(i)}(\mathbf{d}_S)\Psi(\dots, \mathbf{x}_j, \dots, \mathbf{x}_i, \dots). \quad (4.19)$$

We see that the eigenvalue,  $\lambda^{(i)} = \lambda^{(j)} = \lambda$ , is independent of which electron is being translated, as one might expect for identical particles. Thus, we can omit the electron index from the translation operators and their eigenvalues.

Given two simulation cell lattice vectors,  $\mathbf{d}_1$  and  $\mathbf{d}_2$ , we see that,

$$\hat{T}(\mathbf{d}_1)\hat{T}(\mathbf{d}_2) = \hat{T}(\mathbf{d}_1 + \mathbf{d}_2). \quad (4.20)$$

Thus

$$\lambda(\mathbf{d}_1)\lambda(\mathbf{d}_2) = \lambda(\mathbf{d}_1 + \mathbf{d}_2) \quad (4.21)$$

and we find, by inspection, that

$$\lambda(\mathbf{d}_S) = \exp(\mathbf{v} \cdot \mathbf{d}_S), \quad (4.22)$$

for some vector  $\mathbf{v}$  to be determined.

To further refine the form of  $\lambda$ , we note that the translation operators must be

unitary. Given two states,  $|\Psi\rangle$  and  $|\Phi\rangle$ ,

$$\begin{aligned}\langle\Psi|\hat{T}^\dagger(\mathbf{d}_S)\hat{T}(\mathbf{d}_S)|\Phi\rangle &= \int d\mathbf{x}\Psi^*(\dots,\mathbf{r}_i+\mathbf{d}_S,\dots)\Phi(\dots,\mathbf{r}_i+\mathbf{d}_S,\dots), \\ &= \int d\mathbf{x}'\Psi^*(\mathbf{x}')\Phi(\mathbf{x}') \\ &= \langle\Psi|\Phi\rangle.\end{aligned}\tag{4.23}$$

Thus,

$$\lambda(\mathbf{d}_S) = \exp(i\mathbf{k} \cdot \mathbf{d}_S).\tag{4.24}$$

Substituting this eigenvalue into eq. (4.16) we obtain the familiar relationship

$$\Psi(\dots,\mathbf{r}_i+\mathbf{d}_S,\dots) = \exp(i\mathbf{k} \cdot \mathbf{d}_S)\Psi(\mathbf{x}).\tag{4.25}$$

That is, we see that eigenfunctions of eq. (4.7) can be written in the Bloch form

$$\Psi_{\mathbf{k}}(\mathbf{x}) = \exp\left(i\mathbf{k} \cdot \sum_{i=1}^N \mathbf{r}_i\right) u_{\mathbf{k}}(\mathbf{x}),\tag{4.26}$$

where  $u_{\mathbf{k}}(\mathbf{r})$  is strictly invariant under the translation of any single electron by any simulation cell lattice vector. The vector,  $\mathbf{k}$ , is known as the ‘twist vector’. A zero twist vector corresponds to strictly periodic boundary conditions, while a non-zero twist vector corresponds to ‘twisted boundary conditions’. For a non-interacting Hamiltonian, quantities evaluated by averaging expectation values over the ground-state wavefunction for all possible twist vectors in the supercell Brillouin zone are exactly equal to their value in the thermodynamic limit.<sup>7</sup> By analogy, twist averaging is often used in QMC to treat finite size effects for interacting many-body Hamiltonians. This is *not* exact; however, in practice, twist averaging reduces finite-size errors [136].

Applying the same treatment with respect to the simultaneous translation of all electrons by a unit cell lattice vector  $\mathbf{d}_P$  yields the relation,

$$\Psi(\mathbf{r}_1+\mathbf{d}_P,\sigma_1,\dots,\mathbf{r}_N+\mathbf{d}_P,\sigma_N) = \exp(i\mathbf{k}' \cdot \mathbf{d}_P)\Psi_{\mathbf{k}'}(\mathbf{x}).\tag{4.27}$$

Thus,

$$\Psi_{\mathbf{k}'}(\mathbf{x}) = \exp\left(i\mathbf{k}' \cdot \frac{1}{N} \sum_{i=1}^N \mathbf{r}_i\right) v_{\mathbf{k}'}(\mathbf{x}),\tag{4.28}$$

---

<sup>7</sup>This can be seen by considering the equivalent expectation values of operators obtained by: (i) averaging over twist vectors for the expectation value for a single Slater determinant ansatz containing a finite number of Bloch-function orbitals; and (ii) evaluating the expectation value for the (phenomenally large) Slater determinant containing a number of Bloch-function orbitals approaching the thermodynamic limit.

where  $v_{\mathbf{k}}(\mathbf{x})$  is strictly invariant under the translation of all electrons by  $\mathbf{d}_P$ .

Implementing periodic boundary conditions requires that we constrain the wavefunction ansatz to represent functions of the form in eq. (4.26). This symmetry must be enforced explicitly: we cannot hope that the trial wavefunction will spontaneously converge to a solution that is symmetric over the infinite space covered by the simulation cell translation operators given a local, finite sampling. We do not explicitly enforce the all-electron symmetry, eq. (4.28); instead, we leave this to be learned during optimization.

Unlike simulation cell translation, the unit cell translation operators can only traverse a finite space (the simulation cell) once the simulation cell translation symmetry has been enforced. In theory, there are additional symmetry constraints on the many-body wavefunction due to the non-translational elements of the space group of the crystalline lattice. To our knowledge, the imposition of space group symmetries on the many-body wavefunction has not yet been explored in the NNVMC context.

#### 4.1.2 Implementing PBCs in NNVMC

We now consider how to construct strictly simulation-cell periodic (i.e.,  $\mathbf{k} = 0$ ) wavefunction ansätze. We choose input features that are explicitly periodic under the translation of an electronic co-ordinate by a simulation cell lattice vector. This constrains the neural network, forcing it to represent wavefunctions that obey the periodic boundary conditions.<sup>8</sup>

Our approach is guided by the common physicists’ intuition that a parallelepiped cell subject to periodic boundary conditions is topologically equivalent to a 3-torus. This is construed as the Cartesian product of three circles, corresponding to the three primitive simulation-cell lattice vectors. Under this interpretation, constructing a simulation-cell periodic function for any parallelepiped cell is equivalent to constructing a function on a set of co-ordinates uniquely labelling the points on the 3-torus.

To generalize to any Bravais lattice, we transform electronic co-ordinates from the real-space basis (Cartesian co-ordinates) to the lattice vector basis (‘fractional co-ordinates’). This is achieved by recognizing that any vector,  $\mathbf{r}$ , in real space can be decomposed in terms of the lattice vectors as

$$\mathbf{r} = \mathbf{A}\mathbf{s}, \quad \mathbf{A} = \begin{pmatrix} | & | & | \\ \mathbf{a}_1^S & \mathbf{a}_2^S & \mathbf{a}_3^S \\ | & | & | \end{pmatrix}, \quad (4.29)$$

---

<sup>8</sup>This is not the only way to construct a periodic ansatz. One could take the convolution of an aperiodic ansatz, that decays to zero as  $r \rightarrow \infty$ , and a Dirac comb – represented in practice by a truncated sum, or Ewald summation.

where  $\mathbf{s}$  is the vector in fractional lattice-vector co-ordinates. Simulation-cell periodic functions in Cartesian co-ordinates are given by functions that are periodic under translations by Euclidean unit vectors in fractional co-ordinates.

Due to the Cartesian product structure, we can treat each direction in the fractional co-ordinate space separately. To map the fractional co-ordinate space onto the 3-torus we define a mapping  $\mathbf{f}(s_i)$  (where we understand  $s_i$  to represent one of the three components of a vector in fractional co-ordinate space) that is:

1. Periodic, such that  $\mathbf{f}(s_i) = \mathbf{f}(s_i + n)$  for  $n \in \mathbb{Z}$
2. One-to-one in the unit cube, such that points on the 3-torus are uniquely labelled
3. Differentiable up to at least second-order

Requirements 1 and 2 enforce the wrapping of the fractional co-ordinate space into a unit 3-torus, and requirement 3 is informed by the physical problem we are trying to solve: discontinuities in the wavefunction, and its derivatives up to second order, add spurious contributions to the kinetic energy.

These requirements are satisfied by defining any smooth (up to second derivatives), closed, and non-self-intersecting parametric curve on the interval  $[0, 1]$  (where fractional co-ordinates outside this interval are understood to wrap back around). Parametric curves that are non-self-intersecting are known as ‘simple’ curves. Note that it is only possible to define a closed, simple curve in a space with two or more dimensions. This is easily understood by tracing out a continuous and periodic scalar function on a one-dimensional number line in the interval  $[0, 1]$ , beginning at the origin. Such a function *must* turn back on itself to return to its value at the origin, violating requirement 2.

The most obvious choice satisfying our criteria is the unit circle embedded in  $\mathbb{R}^2$ ,

$$\mathbf{f}(s_i) = \begin{bmatrix} \sin(2\pi s_i) \\ \cos(2\pi s_i) \end{bmatrix}. \quad (4.30)$$

In addition to the properties above, this choice of mapping is ‘flat’: motion along the interval  $s_i \in [0, 1]$  at a constant speed results in motion along the unit circle,  $\mathbf{f}(s_i)$ , at constant speed. Thus, we write the mapping between a vector in the fractional co-ordinate basis and our parametric manifold equivalent to the 3-torus as,

$$\mathbf{f}(\mathbf{s}) = \begin{bmatrix} \sin(2\pi \mathbf{s}) \\ \cos(2\pi \mathbf{s}) \end{bmatrix}, \quad (4.31)$$

where the sin and cos functions are understood to act elementwise on the vector  $\mathbf{s}$ .<sup>9</sup>

---

<sup>9</sup>This treatment is analogous to the ‘Clifford torus’ in geometric topology, a flat embedding of

We also desire a definition of distances between points on the 3-torus. To realize this, it is convenient to offset the parametric manifold so that it contains the origin at  $\mathbf{s} = 0$ ,

$$\mathbf{g}(\mathbf{s}) = \begin{bmatrix} \sin(2\pi\mathbf{s}) \\ 1 - \cos(2\pi\mathbf{s}) \end{bmatrix}. \quad (4.32)$$

One approach to measuring distances on  $\mathbf{g}$  is to use the Euclidean distance in the embedding space,  $\mathbb{R}^6$ ,

$$\begin{aligned} \|\mathbf{r}\|_{\text{per}}^2 &= \mathbf{g}^T \mathbf{g} \\ &= \sum_{n=1}^3 [\sin^2(2\pi s_n) + (1 - \cos(2\pi s_n))^2] \\ &= \sum_{n=1}^3 4 \sin^2(\pi s_n). \end{aligned} \quad (4.33)$$

However, if we inspect the behavior of this distance measure as  $\mathbf{r} \rightarrow 0$ ,

$$\begin{aligned} \|\mathbf{r}\|_{\text{per}}^2 \Big|_{\mathbf{r} \rightarrow 0} &= \sum_{n=1}^3 4\pi^2 s_n^2 + O(s_n^4), \\ &\simeq \sum_{m=1}^3 \sum_{n=1}^3 4\pi^2 [(\mathbf{A}^{-1})_{mn} r_n]^2, \end{aligned} \quad (4.34)$$

we see that the projection from the simulation cell to fractional co-ordinates has distorted the Euclidean distance near the origin.

To ease the representation of the Kato cusp conditions, we would like to restore the Euclidean limit for our distance metric near the origin in real space. Thus, we make a slight modification of eq. (4.33):

$$\begin{aligned} 4\pi^2 \|\mathbf{r}\|_{\text{per}}^2 &= \mathbf{g}^T (\mathbf{S} \otimes \mathbb{I}) \mathbf{g}, \quad \mathbf{S} = \mathbf{A}^T \mathbf{A}, \\ &= \sum_{m=1}^3 \sum_{n=1}^3 [\sin(2\pi s_m) S_{mn} \sin(2\pi s_n) \\ &\quad + (1 - \cos(2\pi s_m)) S_{mn} (1 - \cos(2\pi s_n))], \end{aligned} \quad (4.35)$$

where  $\mathbb{I}$  is the  $2 \times 2$  identity matrix, such that

$$\|\mathbf{r}\|_{\text{per}}^2 \Big|_{\mathbf{r} \rightarrow 0} = (\mathbf{s}^T \mathbf{A}^T) (\mathbf{A} \mathbf{s}) + O(s^4) \simeq \mathbf{r}^T \mathbf{r}, \quad (4.36)$$

as required. This form of the distance metric is motivated by noticing that as  $\|\mathbf{s}\|^2 \rightarrow 0$ ,  $\|\mathbf{r}\|_{\text{per}}^2 \rightarrow \mathbf{s}^T \mathbf{S} \mathbf{s} = \mathbf{r}^T \mathbf{r} = \|\mathbf{r}\|^2$ , *i.e.* the usual Euclidean metric is reproduced at small distances. Note that our choice to use a ‘flat’ parametric curve

---

the 2-torus.

has become important here: Euclidean geometry at small distances in fractional co-ordinate space is preserved by our choice of mapping to the 3-torus.

We are now in a position to extend neural network ansätze to periodic boundary conditions. It is sufficient to embed all vector-valued input features on the 3-torus using eq. (4.31), and embed all the scalar-valued (distance) features using the distance measure eq. (4.35). The resulting input features uniquely label all possible electronic configurations subject to periodic boundary conditions, such that the neural network can only represent functions that obey the boundary conditions. The input features to the FermiNet, modified to incorporate periodic boundary conditions, are written

$$\mathbf{h}_{0,i}^\alpha = \left( \sin \left( 2\pi \mathbf{A}^{-1} [\mathbf{r}_i^\alpha - \mathbf{R}_j] \right), \cos \left( 2\pi \mathbf{A}^{-1} [\mathbf{r}_i^\alpha - \mathbf{R}_j] \right), \|\mathbf{r}_i^\alpha - \mathbf{R}_j\|_{\text{per}} \quad \forall j \right), \quad (4.37)$$

for the single-particle features, and

$$\mathbf{h}_{0,ij}^{\alpha\beta} = \left( \sin \left( 2\pi \mathbf{A}^{-1} [\mathbf{r}_i^\alpha - \mathbf{r}_j^\beta] \right), \cos \left( 2\pi \mathbf{A}^{-1} [\mathbf{r}_i^\alpha - \mathbf{r}_j^\beta] \right), \left\| \mathbf{r}_i^\alpha - \mathbf{r}_j^\beta \right\|_{\text{per}} \right), \quad (4.38)$$

for the two-particle features. We will use an analogous treatment to extend the PsiFormer to periodic boundary conditions, discussed in chapter 6.

In addition to the input features, we must modify the envelope, eq. (3.47), to satisfy periodic boundary conditions. Omitting the envelope entirely is the simplest solution. The envelope is included in open boundary conditions to enforce the decay of the wavefunction as  $r \rightarrow \infty$ , but this is not required in periodic boundary conditions. The only other option is to adopt a form of the envelope that is strictly periodic under the translation of an electron by a simulation cell lattice vector. Any such envelope can be expressed as a Fourier series using the simulation cell reciprocal lattice vectors, written for real wavefunctions as,

$$f_i^{k\alpha}(\mathbf{r}_j^\alpha) = \sum_{\mathbf{g}_S n} \left[ \nu_{i\mathbf{g}_S}^{k\alpha} \cos(\mathbf{g}_S \cdot [\mathbf{r}_j^\alpha - \mathbf{R}_n]) + \mu_{i\mathbf{g}_S}^{k\alpha} \sin(\mathbf{g}_S \cdot [\mathbf{r}_j^\alpha - \mathbf{R}_n]) \right], \quad (4.39)$$

and for complex wavefunctions as,

$$f_i^{k\alpha}(\mathbf{r}_j^\alpha) = \sum_{\mathbf{g}_S n} \nu_{i\mathbf{g}_S}^{k\alpha} \exp(\mathbf{g}_S \cdot [\mathbf{r}_j^\alpha - \mathbf{R}_n]), \quad (4.40)$$

where  $\nu$  and  $\mu$  are variational parameters indexed by determinants ( $k$ ), spin ( $\alpha$ ), orbital ( $i$ ), atom ( $n$ ), and reciprocal lattice vector ( $\mathbf{g}_S$ ). In practice, this envelope must be truncated at a finite number of terms. Typically, we choose to truncate the sum at the Fermi wavevector of a non-interacting electron gas in a given simulation cell. We have found that omitting the envelope or using the Fourier envelope both

work well in practice; however, using the Fourier envelope leads to faster initial convergence during optimization. We suspect this is due to the Fourier envelope being trivially capable of representing the ground-state of the non-interacting electron gas.

Finally, we make a small modification to the propagation of MCMC walkers. Allowing MCMC walkers to move far away from the (arbitrary) origin of our coordinate system introduces two sources of numerical error: first, the Ewald summation used to evaluate the Coulomb potential is only convergent near the origin; and second, extremely large values for the electronic co-ordinates introduce round-off errors due to finite precision floating point arithmetic. Both of these issues are mitigated simply by moving the MCMC walkers back into the primitive simulation cell (through translation by an appropriate simulation-cell lattice vector) after every Metropolis-Hastings update step.

Let us now generalize the strictly periodic ( $\mathbf{k} = 0$ ) treatment we have developed to twisted boundary conditions ( $\mathbf{k} \neq 0$ ). Fortunately, we have already solved most of the problem: the  $u_{\mathbf{k}}$  component in eq. (4.26) is a strictly periodic function of the kind we have developed in the  $\mathbf{k} = 0$  case. It is tempting, then, to take the periodic neural network ansatz and simply multiply it by the appropriate Bloch factor. Unfortunately, this is not sufficient. We have not yet described how to represent a *complex-valued* wavefunction using a neural network, and, as we will now show, the function  $u_{\mathbf{k}}$  is complex-valued in the general case. Consider the energy functional at a non-zero  $\mathbf{k}$ ,

$$\begin{aligned} E[\Psi_{\mathbf{k}}] &= \frac{\int d\mathbf{x} \Psi_{\mathbf{k}}^* \mathcal{H} \Psi_{\mathbf{k}}}{\int d\mathbf{x} \Psi_{\mathbf{k}}^* \Psi_{\mathbf{k}}}, \\ &= \frac{\int d\mathbf{x} \exp(-i\mathbf{k} \cdot \sum_{i=1}^N \mathbf{r}_i) u_{\mathbf{k}}^* \mathcal{H} \exp(i\mathbf{k} \cdot \sum_{i=1}^N \mathbf{r}_i) u_{\mathbf{k}}}{\int d\mathbf{x} |u_{\mathbf{k}}|^2}, \\ &= \frac{\langle u_{\mathbf{k}} | \mathcal{H}_{\mathbf{k}} | u_{\mathbf{k}} \rangle}{\langle u_{\mathbf{k}} | u_{\mathbf{k}} \rangle}. \end{aligned} \tag{4.41}$$

We see that this takes the form of the energy expectation value with respect to the strictly periodic function  $u_{\mathbf{k}}$  for a twist-dependent Hamiltonian,  $\mathcal{H}_{\mathbf{k}}$ . The twist-dependent Hamiltonian is written in a real-space basis as

$$\begin{aligned} \mathcal{H}_{\mathbf{k}} &= \exp(-i\mathbf{k} \cdot \sum_{i=1}^N \mathbf{r}_i) \mathcal{H} \exp(i\mathbf{k} \cdot \sum_{i=1}^N \mathbf{r}_i), \\ &= -\frac{1}{2} \sum_{i=1}^N (\nabla_i + i\mathbf{k})^2 + V(\mathbf{r}), \end{aligned} \tag{4.42}$$

where all the potential energy terms are bundled into  $V(\mathbf{r})$ . Twisted boundary conditions have altered the momentum operator, resulting in a term akin to the

canonical momentum obtained under the application of an external magnetic field. The problem of finding the ground-state wavefunction of the original Hamiltonian subject to twisted boundary conditions is thus reformulated as the problem of finding the ground-state wavefunction,  $u_{\mathbf{k}}$ , of the twist-dependent Hamiltonian,  $\mathcal{H}_{\mathbf{k}}$ . This Hamiltonian is *not* time-reversal invariant as it contains imaginary-valued terms, and thus does not, in general, permit real-valued eigenfunctions.

To represent the ground-state wavefunction subject to twisted boundary conditions, we must extend the neural network wavefunction ansatz to represent complex-valued functions. This is achieved by doubling the number of orbitals output from the neural network backbone, eq. (3.46), and treating these alternately as the real and imaginary components of the orbitals entering the Pfau-Spencer determinants. One might also consider using the two outputs as the magnitude and phase of the orbitals entering the Pfau-Spencer determinants. Given a complex-valued, strictly periodic, wavefunction, we can solve the system in twisted boundary conditions by multiplying the wavefunction by the simulation cell Bloch phase and minimizing the expectation value of the original Hamiltonian, or by omitting the Bloch phase and minimizing the expectation value of the twist-dependent Hamiltonian,  $\mathcal{H}_{\mathbf{k}}$ . In theory, both treatments are identical.

In the present work, we do not utilize twist averaging, as it is computationally prohibitive for the present scope, and discuss its treatment only to motivate future work. Certainly, our calculations will inherit a non-trivial finite-size effect due to the neglect of twist averaging and utilisation of small unit cells. However, we will compare against calculations carried out in strictly periodic boundary conditions,  $\mathbf{k} = 0$ , where all calculations will inherit an equivalent error due to finite size effects, enabling fair comparison.

## 4.2 Results and discussion

Under periodic boundary conditions, the HEG Hamiltonian for a simulation cell containing  $N$  electrons, which we will refer to by the shorthand  $N$ -HEG, is written

$$\mathcal{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 + \frac{1}{2} \sum_{\{\mathbf{d}_S\}} \sum_{i \neq j}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j - \mathbf{d}_S|} + \xi(\{\mathbf{a}_i\}), \quad (4.43)$$

where  $\xi(\{\mathbf{a}_i\})$  is a cell-geometry dependent constant representing the interaction of each electron with the positively charged background, exactly equal to the sum of self-interaction energies (given by the Madelung constant) for each electron. The Coulomb sum is evaluated via Ewald summation.

In the thermodynamic limit, the physics of the HEG should be independent of

the exact simulation cell geometry. The properties of the ground state depend only upon the density,  $\rho = N/V$ . The density is typically characterized in terms of the ‘Wigner-Seitz radius’,  $r_s$  – the radius of a sphere that, on average, encloses a single electron. The Wigner-Seitz radius and simulation cell volume are related as,

$$V = \frac{4}{3}\pi N r_s^3. \quad (4.44)$$

For small  $N$ , there are substantial finite-size effects that *do* depend upon the simulation cell geometry. Therefore, it is not sensible to compare different simulation methods for finite systems unless all simulations used the same simulation cell geometries.

The periodic single-particle input features derived in the previous section are taken relative to nuclear co-ordinates. In the HEG, there are no nuclei, so we take the single-particle input features with reference to an arbitrary origin,

$$\mathbf{h}_{0,i}^\alpha = \left( \sin(2\pi \mathbf{A}^{-1} \mathbf{r}_i^\alpha), \cos(2\pi \mathbf{A}^{-1} \mathbf{r}_i^\alpha), \|\mathbf{r}_i^\alpha\|_{\text{per}} \right). \quad (4.45)$$

For all calculations in this chapter, we utilize a FermiNet wavefunction with a Fourier series envelope. We initialize all the parameters of the envelope to zero, except  $\nu_{i0} = 1$ , *i.e.*, the envelope is equal to unity at the beginning of training. We do not employ any pre-training in our calculations for the HEG. The parameters of the wavefunction are randomly initialized using the Kaiming initialization method [92]. The hyperparameters for the majority of our calculations are chosen equal to the ‘sensible defaults’ in table 3.3. Near the transition density (as determined by benchmark calculations), we find difficulty converging the NNVMC calculations with these defaults. Convergence is improved in these cases by increasing the KFAC norm constraint and damping hyperparameters by two orders of magnitude.

### 4.2.1 Unpolarized simple cubic 14-HEG

The first system we consider is the unpolarized ( $S_z = 0$ , fixed by spin assignment) 14-HEG in a simple cubic simulation cell. This system was chosen due to the availability of highly accurate i-FCIQMC benchmarks. We utilize a FermiNet wavefunction ansatz, modified to account for PBCs as described in the previous section. The NNVMC calculations are implemented as described in section 3.3.5, and we calculate the ground-state energy of the unpolarized 14-HEG for the same simulation cells as the i-FCIQMC benchmarks. The results of these calculations, given in terms of the correlation energy (the difference between the Hartree-Fock energy and the ground-state energy obtained by a correlated method), are shown in table 4.1. Table 4.1 also contains benchmark energies calculated via the SJB-VMC and SJB-DMC methods,

| Method                       | Correlation energy [Hartrees] |                   |                   |                  |
|------------------------------|-------------------------------|-------------------|-------------------|------------------|
|                              | $r_s = 0.5$                   | $r_s = 1.0$       | $r_s = 2.0$       | $r_s = 5.0$      |
| SJB ( $n_{\text{det}} = 1$ ) |                               |                   |                   |                  |
| VMC                          | −0.58624(1)                   | −0.5254(1)        | −0.437(3)         | −0.30339(2)      |
| DMC                          | −0.58778(1)                   | −0.5254(1)        | −0.4385(3)        | −0.30474(8)      |
| FermiNet                     |                               |                   |                   |                  |
|                              | Block determinants            |                   |                   |                  |
| $n_{\text{det}} = 1$         | −0.58831(7)                   | −0.52510(3)       | −0.43842(1)       | −0.30433(1)      |
| $n_{\text{det}} = 16$        | −0.58962(6)                   | −0.52558(3)       | −0.43876(1)       | −0.30466(1)      |
|                              | Dense determinants            |                   |                   |                  |
| $n_{\text{det}} = 1$         | −0.58895(6)                   | −0.52568(3)       | −0.43881(1)       | −0.30468(1)      |
| $n_{\text{det}} = 16$        | −0.59094(6)                   | −0.52682(3)       | −0.44053(1)       | −0.30495(1)      |
| i-FCIQMC[131]                |                               |                   |                   |                  |
| finite basis                 | −0.5939(4)                    | −0.5305(5)        | −0.4430(7)        | −0.304(1)        |
| basis set limit              | <b>−0.5969(3)</b>             | <b>−0.5325(4)</b> | <b>−0.4447(4)</b> | <b>−0.306(1)</b> |

Table 4.1: Correlation energy of the spin-polarized  $N = 14$  HEG subject to simple cubic boundary conditions, obtained via i-FCIQMC, SJB-VMC/DMC, and FermiNet-NNVMC. Table reproduced from [105].

using the `CASINO` program [26, 39]. The SJB ansatz consists of a single Slater determinant of plane waves – *i.e.*, the Hartree-Fock orbitals of this system – and Jastrow and backflow factors as described in ref [137].

The ground-state correlation energy has been obtained using a FermiNet ansatz with 1 and 16 determinants, with both block and dense determinants. An improvement in the variational lower bound is obtained for all densities when using more determinants, and when using dense determinants. The most accurate ground-state correlation energies obtained with NNVMC are within a few milliHartrees of the finite basis i-FCIQMC results. Despite falling a few milliHartrees short of the i-FCIQMC energies, this benchmark validates the applicability of the NNVMC approach for studying the HEG. Importantly, we note that this near i-FCIQMC accuracy is being obtained by a method which can scale to larger systems. We expect that the performance of NNVMC could be improved with further computational effort, but do not think this is worthwhile for a simple benchmark system.

The discrepancy between the best correlation energies obtained by NNVMC and those obtained with i-FCIQMC decreases with increasing  $r_s$ . This is because the fermion sign problem in i-FCIQMC is exacerbated at low densities, which inhibits the use of basis sets large enough to produce accurate results [131]. In contrast, the accuracy of NNVMC does not seem to suffer as the density is decreased. In fact, the variance in the FermiNet energies decreases with increasing  $r_s$ , likely because the magnitude of the total energy also decreases sharply with  $r_s$ . There is

a systematic discrepancy of  $\sim 3$  milliHartrees between the FermiNet energies and finite-basis i-FCIQMC energies. This vanishes when comparing to the i-FCIQMC energies extrapolated to the complete basis set limit. However, it has been shown that there is reason to doubt the accuracy of the complete basis set extrapolations, which may be 1–2 milliHartrees too low [138].

These results demonstrate that our implementation of NNVMC in periodic boundary conditions produces very accurate ground-state wavefunctions. Despite being a VMC method, NNVMC obtains a lower variational bound on the correlation energy at all densities than SJB-DMC. Of course, with a multi-determinant expansion, or a more complicated form for the Jastrow and backflow functions, there is no reason to believe that SJB-DMC cannot also meet this level of accuracy. The point here is rather to demonstrate that NNVMC does yield accurate correlation energies for this system, providing a proof of concept for our implementation of PBCs.

#### 4.2.2 Polarized body-centered cubic 27-HEG

To highlight the capabilities of the NNVMC method in representing quantum phase transitions, it is important to first understand how the Wigner transition is treated in typical SJB-based calculations. One defines two ansätze, one for the Fermi liquid (high density) phase, and one for the Wigner crystal (low density) phase. These ansätze are distinguished by the choice of single-particle orbitals in the Slater determinants. The Fermi liquid ansatz uses the Hartree-Fock orbitals, simple plane-waves, as the single-particle orbitals; the Wigner crystal ansatz utilizes site-centered Gaussian orbitals arranged in the body-centered cubic lattice of the Wigner crystal. Neither of these ansätze is capable of adequately representing the other phase without taking an intractably large linear combination of Slater determinants. In order to locate the density at which the Wigner transition occurs, these two ansätze are used as trial wavefunctions in VMC/DMC calculations of the ground-state energy, and the transition point is taken as the density at which the corresponding ground-state energies cross. This is an effective approach in systems, like the HEG, where the nature of the phase diagram is well understood *a priori*. However, in systems where the nature of the appropriate ground-state wavefunction is not well understood, it is not possible to discover quantum phase transitions with such an approach.

Having established the quantitative accuracy of NNVMC in benchmark calculations for the HEG, we wish to assess the capabilities of the method in treating the Wigner transition. We study the  $N=27$  electron spin-polarized ( $S_z = N/2$ , fixed by spin assignment) HEG in a body-centered cubic simulation cell. Note that choosing a cubic number of electrons in a bcc simulation cell biases the ground-state

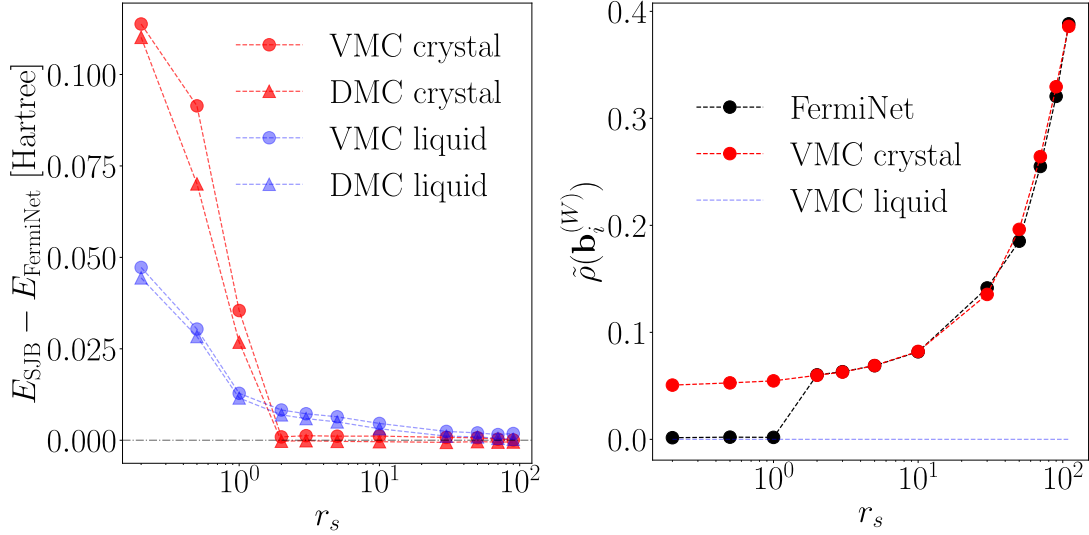


Figure 4.1: **Left:** Total ground-state energy per electron obtained via SJB-VMC and DMC, relative to the energy obtained via NNVMC. The ‘liquid’ and ‘crystal’ results were obtained using Slater determinants of plane waves and Gaussian orbitals, respectively. Error bars are smaller than the markers. **Right:** Fourier component of the one-electron density corresponding to the primitive reciprocal lattice vectors of the emergent Wigner crystal, averaged over the three primitive lattice vectors. This acts as an appropriate order parameter for the emergence of the Wigner crystal phase. Error bars are smaller than the markers. Figure reproduced, with modifications, from [105].

wavefunction toward a  $3 \times 3 \times 3$  Wigner crystal. This system does not resemble the thermodynamic limit; rather, it serves as a toy model for studying the behavior of NNVMC at phase transitions. This choice dramatically reduces the value of  $r_s$  at which the Wigner transition occurs, compared to the thermodynamic limit. Choosing a spin-polarized system is an arbitrary decision which we do not expect to affect our qualitative conclusions, although the study of magnetic effects, and the possible consequences for the phase diagram, in mixed-spin systems would carry some interest.

Ground-state energies obtained via NNVMC with the FermiNet ansatz are compared with SJB-VMC and SJB-DMC, using wavefunctions targeted at the liquid and crystal states, in fig. 4.1 (left). The SJB ansätze are implemented as described in refs. [137, 139], and the FermiNet ansatz is implemented as described in section 3.3.5. We find that NNVMC wavefunctions yield tighter variational lower bounds than SJB wavefunctions at all densities at the VMC level. Furthermore, NNVMC outperforms the liquid SJB wavefunction over the entire density range, even below the Wigner-Seitz radius ( $r_s \simeq 2$ ) at which SJB-DMC predicts the Wigner transition to occur in this small system. At the lowest densities, the crystal wavefunction SJB-DMC ener-

gies are slightly lower than those obtained with NNVMC, although the total energy scale of the system is very small at these densities. These results suggest that the nodal surface of the Slater determinant of site-centered Gaussians is almost exact as the zero-density limit is approached (given the accuracy of the crystal SJB-DMC ground-state), but the curvature of the wavefunction is better represented by the FermiNet away from the nodal surface (given the superior accuracy of NNVMC compared to SJB-VMC).

For  $r_s \leq 1$ , we find that the energy of the real-valued FermiNet trial wavefunction often plateaus during training, yielding energies  $\sim 0.1\%$  higher than the SJB-DMC benchmarks. However, these plateaus do not occur when using a complex-valued FermiNet trial wavefunction. The implementation of complex-valued wavefunctions does not meaningfully increase the representational capacity of the wavefunction ansatz. Thus, we hypothesize that using a complex wavefunction somehow eases optimization toward the ground state in the Fermi liquid phase.<sup>10</sup>

The onset of the Wigner transition can be seen by calculating the Fourier coefficients of the one-electron density at the reciprocal lattice vectors of the emergent bcc crystal. Writing the reciprocal lattice vectors of the Wigner crystal as  $\mathbf{b}_i^W$ , this order parameter is

$$\tilde{\rho}(\mathbf{b}_i^W) = \frac{\langle \Psi | \frac{1}{N} \sum_{j=1}^N \exp(i\mathbf{b}_i^W \cdot \mathbf{r}_j) | \Psi \rangle}{\langle \Psi | \Psi \rangle}. \quad (4.46)$$

For the chosen number of electrons and simulation cell, the emergent crystal packs  $3 \times 3 \times 3$  unit cells in the simulation cell. Thus, the appropriate primitive reciprocal lattice vectors are

$$\mathbf{b}_i^W = 3\mathbf{b}_i^S, \quad (4.47)$$

for  $\mathbf{b}_i^S$  as defined in eq. (4.3). The values of this order parameter, averaged over the three reciprocal lattice directions, are shown in fig. 4.1 (right), calculated using the SJB gas, SJB crystal, and FermiNet wavefunctions, as a function of  $r_s$ .

The order parameter,  $\tilde{\rho}$ , is zero at all  $r_s$  by construction when calculated with respect to the SJB liquid trial wavefunction. Likewise, the order parameter is finite at all  $r_s$  when calculated with respect to the SJB crystal trial wavefunction. At the Wigner transition density – defined by the density at which the SJB liquid and SJB crystal trial wavefunctions cross over in energy – we see that the order parameter calculated with respect to the FermiNet trial wavefunctions jumps from zero to the same finite value predicted by the SJB crystal trial wavefunction. This demonstrates that the FermiNet trial wavefunction is converging to a Wigner crystal state

---

<sup>10</sup>Through private communications with other researchers exploring NQS and the NNVMC method, we have been informed that this behavior has also been observed in other systems.

below the transition density, and to a Fermi liquid state above the transition density. In addition to the close agreement between the ground-state energy calculated via NNVMC and SJB-VMC/DMC, these results show that FermiNet is capable of converging to the appropriate ground-state wave function with good accuracy across the full density range. This is achieved with a single wavefunction ansatz: we do not supply the NNVMC calculation with any hand-crafted features to indicate the appropriate ground-state symmetry, specific designation of crystal sites, or any other information that a quantum phase transition should occur.

There is, however, an elephant in the room. The HEG Hamiltonian, eq. (4.43), commutes with the group of operators that simultaneously translate all the electronic co-ordinates by *any* vector. This is, in effect, the primitive unit cell Bloch symmetry discussed in section 4.1.1, in the limit where the single-particle potential is constant everywhere (rather than being periodic, as in a crystalline lattice of ions). The ground-state wavefunction of the finite  $N$ -HEG should be a simultaneous eigenstate of these operators (*i.e.* a total momentum eigenstate).<sup>11</sup> For such an eigenstate, single-particle observables like the one-electron density and order parameter in fig. 4.1 should be completely uniform in the simulation cell. In the Wigner crystal regime, such a state, known as a ‘floating crystal’, possesses a uniform one-electron density and a crystalline structure in the pair-correlation function. In refs [137, 139], it is shown (using a SJB crystal-type wavefunction ansatz) that the energy difference between the fixed and floating crystal is approximately  $\Delta E = 0.055r_s^{-3/2}$ . At low  $r_s$ ,  $\Delta E$  is large (20 milliHartrees at  $r_s = 2$ ). With a splitting this large, we should expect a FermiNet trial wavefunction to converge to the floating crystal state. Clearly, however, this is not the case. It is possible that this estimate of  $\Delta E$  is too large for the particular simulation cell geometry and number of electrons considered here. Alternatively, it is possible that it is necessary to surmount an entropic barrier to reach the floating crystal state during optimization, and that the trial wavefunction always becomes stuck in a region of parameter space corresponding to a particular fixed crystal configuration.

### 4.3 Related work and future directions

The zero-temperature phase diagram of the homogeneous electron gas has been thoroughly explored with QMC methods. To understand how NNVMC might offer new lines of inquiry, it is helpful to lay out a brief timeline. We begin around 1930, with the predictions of Bloch [140] and Wigner [125, 126]. Bloch predicted that the homogeneous electron gas should exhibit itinerant ferromagnetism (a ferromagnetic

---

<sup>11</sup>In the thermodynamic limit, the fixed and floating wavefunctions are degenerate and spontaneous symmetry breaking allows for a fixed Wigner crystal ground state.

Fermi liquid phase) at intermediate density, and Wigner predicted that at sufficiently low density the homogeneous electron gas would form the Wigner crystal. Estimates of the transition densities between these states were made using a variety of perturbative and variational methods over the ensuing decades [141–143], with widely varying results.

Substantial progress arrived in the 1980s due to the pioneering work of Ceperley and Alder, applying the SJ-VMC [144] and SJ-DMC [21] methods to the zero-temperature phase diagram of the homogeneous electron gas. They predicted phase transitions from the paramagnetic Fermi liquid phase to a ferromagnetic Fermi liquid phase at  $r_s = 75 \pm 5$ , and from the ferromagnetic Fermi liquid phase to a ferromagnetic bcc Wigner crystal phase at  $r_s = 100 \pm 20$ . These calculations were later refined to predict the onset of a partially polarized fluid at  $r_s \simeq 20$ , with a polarization between 50% and 100% in the vicinity of the transition to the Wigner crystal – although, the 100% polarized itinerant ferromagnet was found to be unstable at all densities [145].

The story continues in 1999 with the work of Ortiz, Harris, and Ballone [146], who attempted to mitigate the impact of finite size errors by performing SJ-DMC calculations with extremely large numbers of electrons ( $725 \leq N \leq 1450$ ) and extrapolating correlation energies to the thermodynamic limit. They found that the polarization undergoes a continuous transition over the density range  $20 \pm 5 \leq r_s \leq 40 \pm 5$ , with a ferromagnetic phase stabilizing at  $r_s \geq 40$ . Additionally, their calculations suggested that the onset of Wigner crystallization occurs at a much higher density, observing a first-order transition to the bcc Wigner crystal at  $r_s = 65 \pm 10$ .

In 2002, Zong, Lin, and Ceperley [147] built upon the predictions of Ortiz. They incorporated backflow transformations to study the Fermi liquid phase and utilized twist averaging and extrapolation to the thermodynamic limit to mitigate finite size errors. The inclusion of the backflow transformation limits these calculations to fewer electrons than the work of Ortiz, but Zong *et al.* argue that greater profit is obtained from a more accurate trial function alongside a careful treatment of finite-size effects. They predicted a continuous transition to a partially polarized Fermi liquid phase at  $r_s = 50 \pm 2$ . Zong *et al.* highlighted that it is crucial to understand the biases introduced by the choice of wavefunction ansatz, pointing out, for instance, that Hartree-Fock treatments of the HEG at intermediate density bias the ground state towards a polarized fluid because the exchange interaction is the only way to correlate electrons with the HF ansatz. Further, they offered some criticism of their own results:

Our QMC results [...] indicate that there is still an instability for spin polarization at a very low density. Although examination of the vari-

ance indicates that the polarized and unpolarized [...] trial functions are equally inaccurate, it is still not clear if our finding of a polarization transition is an artifact of the assumed trial wave function.

Drummond *et al.* [139] sought to supplement Zong *et al.*'s, results on the polarization of the Fermi liquid phase with equally high quality results in the Wigner crystal regime. This, they hoped, would resolve the discrepancy between the Wigner transition densities predicted by Ceperley and Alder [21], and Ortiz, Harris and Ballone [146]. They did not incorporate backflow transformations, or twist averaging; however, they carried out a careful examination of the sources of numerical and finite-size errors in their results. Their results were extrapolated to the thermodynamic limit based on calculations with  $N = 64$ , 216, and 512 electrons, and they predicted the Wigner transition to occur at  $r_s = 106 \pm 1$ , in closer agreement with Ceperley and Alder than with Ortiz, Harris and Ballone. They offered three possible explanations:

1. Ortiz *et al.* used a trial wavefunction with a plane-wave Slater determinant nodal surface – in contrast to the backflow transformed nodal surface of Zong *et al.* or the release-node method of Ceperley and Alder – reducing the stability of the Fermi liquid phase.
2. The error bars on the results of Ortiz *et al.* are substantially larger than those of Zong *et al.* or Drummond *et al.*
3. Differences in the magnitude of the fixed-node error between the trial wavefunctions used to model the Fermi liquid and Wigner crystal phases.

For some time, the debate was considered broadly settled: the zero-temperature phase diagram of the homogeneous electron gas consists of a high density paramagnetic Fermi liquid, a partially polarized itinerant ferromagnet at intermediate density, and a ferromagnetic bcc Wigner crystal at low density. Bloch and Wigner rest easy. However, in 2022 Azadi and Drummond [137] decided to revisit the problem, citing that:

Because of the fundamental role of the 3DHEG in condensed matter physics, the determination of its zero-temperature phase diagram and ground-state energy is a problem that should be revisited from time-to-time using state-of-the-art computational methods.

Backflow transformations were used for the fluid phases, but not the crystal phases, where little change in the DMC energy is obtained by incorporating backflow transformations. Finite-size effects were mitigated by the application of twist averaged

boundary conditions and extrapolations of the correlation energies to the thermodynamic limit. Their work refines the zero-temperature phase diagram of the HEG substantially. They predict there is *no* density regime in which the itinerant ferromagnetic phase is stable. They find only the first-order transition from the paramagnetic Fermi liquid phase to the ferromagnetic bcc Wigner crystal phase at  $r_s = 86.6 \pm 0.7$ . After a careful investigation of the differences between their new calculations and the SJB-DMC results of Zong *et al.* [147], they conclude that the disagreement over the stability of the itinerant ferromagnet is due to errors in the finite-size extrapolation of Zong *et al.*

Thus, we arrive at the current state of the art. The story of the zero-temperature phase diagram of the HEG in two dimensions proceeds similarly. There, the state-of-the-art phase diagram is qualitatively similar to the three-dimensional case: a paramagnetic Fermi liquid at high density undergoes a first-order phase transition to a triangular *anti*-ferromagnetic Wigner crystal at  $r_s = 35 \pm 1$ , that subsequently undergoes a magnetic transition to a triangular ferromagnetic Wigner crystal at even lower density,  $r_s = 38 \pm 5$  [148, 149], with no region of stability for an itinerant ferromagnetic phase.<sup>12</sup> The existence of an itinerant ferromagnetic phase had been suggested by experimental results [150]; however, more recent experiments do not corroborate this evidence, observing a direct transition from the paramagnetic Fermi liquid to the Wigner crystal state [151].

Given the demonstration by Drummond and Azadi that backflow transformations do not much alter the ground-state energy (and likewise, the nodal surface) of the Wigner crystal phase in a DMC calculation, and the extremely close agreement between our NNVMC and SJB-DMC calculations of the ground-state energy in the Wigner crystal phase of the 27-HEG, it is unlikely that NNVMC will much refine the Wigner crystal ground-state energy or the very low density phase-diagram of the homogeneous electron gas. However, there is some reason to believe that the NNVMC method may be able to provide useful information about the stability of the itinerant ferromagnetic phase at intermediate densities. The nodal surface of the Fermi liquid SJB wavefunction is restricted to the Hartree-Fock topology due to the use of Hartree-Fock orbitals in the Slater determinant and a continuous backflow transformation that, while being able to shift the nodal surface, cannot alter the nodal topology. Indeed, in the 14-HEG and 27-HEG results in table 4.1 and fig. 4.1, we see that NNVMC produces lower energies than SJB-DMC in the Fermi liquid phase. Furthermore, the i-FCIQMC calculations carried out in a basis of plane-waves show substantial occupation of plane-wave orbitals beyond those included in

---

<sup>12</sup>The splitting between the antiferromagnetic and ferromagnetic Wigner crystal states in three-dimensions is tiny due to the large separations between electrons at the relevant densities. The question of which is the true ground state has not been fully resolved. In the work of Azadi and Drummond [137], the splitting is below the statistical precision of their DMC calculations.

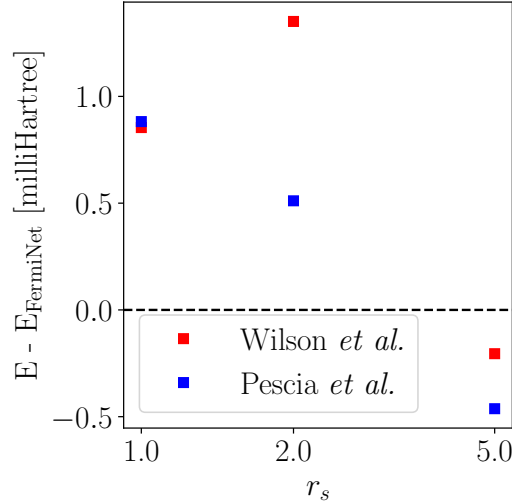


Figure 4.2: Correlation energy of the paramagnetic simple cubic 14-HEG obtained via NNVMC calculations using the WAPNet ansatz (Wilson *et al.* [106]) and MP-NQS ansatz (Pescia *et al.* [110]), relative to that obtained using FermiNet. Error bars are omitted as they are smaller than the markers.

the Hartree-Fock solution. This suggests that the Hartree-Fock nodal topology may not be appropriate for describing the true ground-state wavefunction. Therefore, we propose that an exploration of the intermediate density phase diagram of the three-dimensional homogeneous electron gas using NNVMC may help to resolve the question of the possibility of an itinerant ferromagnetic phase, as the nodal topology is not constrained. Indeed, a recent NNVMC study of the two-dimensional HEG did find non-trivial spin structures in the intermediate density Fermi liquid regime [152], but the impact of finite-size effects was not carefully explored, and these results cannot be considered conclusive. Recent experimental observations of mixed ‘micro-emulsion’ phases with anomalous fluctuations in spin susceptibility in two-dimensional electron gases confined to MoSe<sub>2</sub> monolayers [153] may be consistent with the recent NNVMC simulation results.

Several groups concurrently explored the application of NNVMC to the HEG while the work described in this chapter was being undertaken. This was briefly discussed in chapter 3; we would like to offer here a more detailed comparison of the results. We will not draw a direct comparison with the results of Li *et al.* [96], as they do not exceed the accuracy of their SJB-DMC benchmarks, and we believe those calculations should be regarded as a qualitative proof-of-concept rather than an attempt at quantitatively accurate calculations. Mainly, then, we will compare with the work of Pescia *et al.* [110], using the ‘MP-NQS’ ansatz, and Wilson *et al.* [106], using the ‘WAPNet’ ansatz. These works, and our own, benchmark their implementations of the NNVMC method against i-FCIQMC results for the simple

cubic paramagnetic 14-HEG. The correlation energies obtained are compared in fig. 4.2. The discrepancies are minimal and are not clearly systematic with  $r_s$ . Such sub-milliHartree differences in correlation energies could easily be removed by the tuning of hyperparameters, and should not be considered as quantitative evidence for the superiority of any of the three approaches.

Wilson *et al.* do not study the phase diagram of the homogeneous electron gas, but offer extensive benchmarks against SJB-DMC ground-state energies at a range of densities, finding favorable comparison. The Wigner crystal ground state is not observed at any of the values of  $N$  or  $r_s$  studied. Pescia *et al.* do study the Wigner transition, but their approach is more akin to traditional SJB calculations, where two separate ansätze (one using plane-wave Hartree-Fock orbitals, and another using site-centered Gaussian orbitals) are defined, and the transition density found by locating the ground-state energy crossing between the two. Both WAPNet and MP-NQS are PauliNet-like, depending on explicit (usually, Hartree-Fock) orbitals in the Slater determinant. As in SJB-DMC, the backflow transformations in these ansätze are strictly continuous and cannot alter the nodal topology of the reference determinant(s). Therefore, the FermiNet (and derived ansätze) are strictly more general than the WAPNet and MP-NQS. We do not believe that the *ab initio* discovery of quantum phase transitions we have demonstrated in this chapter would be possible using these ansätze.

## 4.4 Conclusion

In this chapter, we have introduced the HEG, a simple model of correlated electrons in metallic solids. This model is well suited as a test bed for the extension of NNVMC to PBCs, having been extensively studied using QMC methods for several decades. We have derived a Bloch-like condition on the many-body wavefunction subject to PBCs, and described how to enforce these conditions for neural network wavefunction ansätze, using an approach equivalent to representing functions on the 3-torus. We then studied the performance of NNVMC on small electron gases, finding favorable comparison with highly accurate i-FCIQMC and SJB-DMC benchmarks. Furthermore, we have demonstrated that a NNVMC trial wavefunction that does not employ a set of fixed reference orbitals can spontaneously converge to the appropriate ground-state symmetry on either side of the Wigner transition. This suggests that NNVMC is a promising method for discovering quantum phase transitions in condensed matter systems, particularly those where the natures of the ground states involved are not yet understood. Even in the well-studied case of the HEG, NNVMC may be able to resolve the long-standing debate over the existence of an itinerant ferromagnetic phase at intermediate densities.

# Chapter 5

## Positronic chemistry

In the previous chapter, we saw that the flexibility of neural networks wavefunctions facilitates the *ab initio* discovery of quantum phase transitions. Now, we will consider a very different problem: the formation of bound states between a positron – the positively charged anti-particle of the electron – and ordinary molecular matter. Here, again, the flexibility of neural network wavefunctions will prove essential.

The positron was first postulated by Dirac in 1930 as a consequence of his relativistic theory of quantum mechanics [154]. A mere two years hence, the existence of Dirac’s anti-particle was confirmed through the observation of cosmic ray particles with the mass of an electron and a unit positive charge [155]. Today, the positron is no longer a curiosity for theoretical physicists, but an important tool in practical fields as varied as medicine [156], astrophysics [157], and materials science [158, 159]; the gamma rays emitted via annihilation due to electron-positron collisions offer a sensitive probe of matter at the atomic length scales.

The discovery of the positron was followed by many predictions regarding the formation of bound states between positrons and ordinary matter. The simplest such bound state, positronium (Ps), consists of a positron and electron orbiting their mutual center-of-mass, analogous to the hydrogen atom [160]. Like the hydrogen atom, it is straightforward to solve the TISE for this bound state, yielding a binding energy of 0.25 Hartrees. In 1946, Wheeler further postulated the stability of a hierarchy of ‘polyelectrons’ – successively larger clusters of electrons and positrons [161]. The positronic analogue of the hydrogen molecule, dipositronium (Ps<sub>2</sub>), has now been observed experimentally through the bombardment of porous silica with a burst of positrons [162].

Until the turn of the 21<sup>st</sup> century, the enumeration of positronic bound states was focussed on stable complexes of Ps and a neutral atom; alternatively, these can be viewed as a positron forming a bound state with an atomic anion. This myopia is perhaps unsurprising, considering that the interaction potential between an anion and a positively charged particle is attractive everywhere. Strongly dipolar

molecules are another class of systems where it is straightforward to demonstrate the existence of positronic bound states by considering the electrostatic potential between the molecular dipole field and a positron [163].<sup>1</sup>

The long-standing question of the existence of bound states between a positron and neutral atoms – given that such bound states were known to exist for electrons – was first answered in the affirmative in 1997 using *ab initio* calculations [164]. At the level of electrostatics, the interaction potential between a positron and a neutral atom is repulsive everywhere, so bound states can only be understood through a correlated many-body treatment. The same holds for positronic bound states of non-polar molecules. However, nature is ignorant of our theoretical struggles: it is now a well-established experimental fact that positrons *readily* form bound states with many non-polar molecules and neutral atoms. In positron annihilation spectroscopy experiments, the formation of these bound states is enabled by an incident positron exciting a vibrational Feshbach resonance, which subsequently decays to the vibrational ground state of the positronic bound state. This process introduces peaks in the energy-dependent annihilation rate, and enables sensitive measurements of the positron binding energy [165–172].

Of course, positronic bound states are not *truly* stable. It is well known that annihilation occurs upon collision between a positron and an electron, producing  $\gamma$ -rays. The proximity of positrons and electrons in positronic bound states increases the likelihood of a chance meeting, enhancing the annihilation rate compared with that of a free positron passing through a gas of atoms or molecules. This enhancement of the annihilation rate is central to the technological application of positrons. Thus, calculating the positron annihilation rate of positronic atoms and molecules is a central goal of *ab initio* treatments of such systems, in addition to the calculation of binding energies. Typically, the lifetime with respect to annihilation of a positronic bound state is on the order of nanoseconds.

Despite the relativistic origins of the positron, positronic bound states can be understood with a non-relativistic treatment, considering the positron as naught but a positively charged electron. Equipped with this perspective, we can study the interactions between positrons and ordinary matter by solving the TISE with the usual tools of computational quantum chemistry, including NNVMC. However, solving for the ground-state wavefunction of positronic bound states is challenging for traditional quantum chemistry methods.

The success of modern quantum chemistry can be largely attributed to the development of carefully designed basis sets for describing electronic wavefunctions.

---

<sup>1</sup>This problem was first treated in the context of the formation of bound states between dipolar molecules and an electron. However, due to the symmetric nature of the dipole field, these results hold identically for positronic bound states.

Unfortunately, positronic orbitals do not much resemble electronic orbitals: due to the repulsion between positrons and atomic nuclei, the positronic orbitals are peaked away from the atomic nuclei, and are typically highly diffuse. Moreover, the bonding is often dominated by strong correlations between the positron and electrons through the process of ‘virtual’ Ps formation. Accurately describing the wavefunction around the coalescence of electrons and positrons is thus extremely important.

In this chapter we will discuss the successes and failures of traditional quantum chemistry in describing bound states between positrons and ordinary molecules, outline a simple approach to extending the FermiNet ansatz to incorporate positronic components of the wavefunction, and show that NNVMC calculations with this ansatz yield highly accurate, in some case state-of-the-art, ground-state energies and positron annihilation rates for positronic molecules. Furthermore, we will demonstrate that NNVMC works well for a wide range of molecules with qualitatively distinct positron binding characteristics.

## 5.1 Traditional approaches to positron binding

There are three mechanisms that accommodate the formation of bound states between positrons and ordinary matter. The first is the interaction between the positron and the static dipole moment of a molecule. Even at the level of simple electrostatics, the electric field produced by a sufficiently dipolar molecule (greater than  $\sim 1.625$  Debye) will accommodate a positronic bound state [163]. The second mechanism is through the spontaneous polarization of the electron cloud of the host atom or molecule, and the attractive interaction between the resulting dipole moment and the positron (reminiscent of van der Waals interactions). The final mechanism is due to the formation of ‘virtual’ Ps clusters, effectively the formation of a bound complex between Ps and the host atom or molecule due to strong electron-positron correlations (reminiscent, although fundamentally different from, covalent bonding). A special case of the final mechanism is found in atoms where the first ionization energy is less than 0.25 Hartrees (the binding energy of Ps): a lone Ps cluster will form which may then form a bound complex with the resulting cation.<sup>2</sup> In the absence of a large static dipole moment, positron binding typically occurs through a mix of the second and third mechanisms, with the precise balance dictated by the electronic structure of the host system.

The first method to successfully describe the formation of positronic bound states

---

<sup>2</sup>It is possible that the resulting cation and Ps atom will not form a bound complex, and dissociate into separated fragments. In these cases, we do not consider the atom to be capable of forming a positronic bound state.

with neutral atoms [164] was the ‘stochastic variational method with explicitly correlated Gaussians’ (ECG-SVM). Here, the wavefunction ansatz is expanded in a basis of Gaussian functions on the relative co-ordinates between the quantum particles. The efficiency with which Hamiltonian matrix elements between these basis elements can be computed allows for extremely large basis set expansions (typically comprising thousands of basis functions), and the ansatz yields highly accurate (for our purposes, nearly exact) results for few-body problems [173]. The explicit expansion of the wavefunction in the separations between pairs of particles also makes this ansatz well suited to the positron binding problem, where the description of the wavefunction around the coalescence between electrons and positrons is particularly important. The ECG ansatz is optimized via the SVM, where a guided random walk is enacted in parameter space (which may also include the addition or removal of basis functions), and moves that lower the energy expectation value are accepted. ECG-SVM has been used to obtain accurate positron binding energies for clusters of positrons and electrons (e.g. Ps, Ps<sub>2</sub>), dipolar alkali metal hydrides (e.g. LiH) [174], and group I/II atoms (H<sup>-</sup>, Li, Be, Na, Mg, Ca) [173]. Calculations involving the larger group I/II atoms are enabled by the frozen-core approximation, where the core electrons (those occupying the non-outermost electronic shells) are removed from the calculation and substituted by a model Hamiltonian. Such calculations are not totally *ab initio*, and the accuracy of the model Hamiltonian introduces a source of non-variational uncertainty in the ground-state energies obtained. The applicability of the ECG-SVM method is limited to systems containing only a handful of quantum particles, as the large basis sets required do not scale efficiently to larger systems.

The CI method has also been used to describe positron binding to several neutral atoms and dipolar molecules. As in the ECG method, this is typically limited to atoms where the number of electrons being treated explicitly can be reduced to one or two by the frozen-core approximation (group II atoms, and certain transition metal atoms, *e.g.* zinc) [175–180], and the strongly dipolar molecules LiH [181] and BeO [182, 183]. The main difficulty faced by the CI method in describing positron binding is the formation of virtual Ps atoms away from the atomic nuclei. Describing such a wavefunction with atom-centered basis functions yields extremely slow convergence of the energy with respect to the largest angular momentum represented in the basis set. This is starkly demonstrated by the difficulty of treating positron binding to group I atoms in the CI method, where the bleak outlook is well summarized by Bromley [184],

[The] CI method cannot be regarded as the solution to all positron binding problems. The mathematical basis simply does not correspond to the underlying physical representation of a system containing a positro-

nium cluster. In the case of  $e^+Li$  it is likely that the single particle basis would have to be increased to around 1000 electron and 1000 positron orbitals for the CI wavefunction to be reasonably close to converged. [...] Positronic sodium, with the positron and electron cluster located even further from the nucleus, would represent an even more arduous calculation for the CI method, probably requiring the inclusion of orbitals with  $\ell = 100$ .

QMC methods promise to alleviate the problems with the ECG-SVM (offering better scaling with particle number) and CI methods (offering greater flexibility in treating electron-positron correlations) in describing positronic bound states. Furthermore, FN-DMC should offer a very accurate description of the wavefunction around the coalescence of electrons and a positron: the lack of mutual exchange between the different species of particle means that these regions of configuration space do not, generally, contain nodes, and are not impacted by the fixed node approximation. Despite this promise, most DMC results for positron binding energies have been limited to the same settings as ECG-SVM and CI calculations: neutral atoms and strongly dipolar molecules [185–191]. Furthermore, the calculation of annihilation rates in the DMC method is complicated by the fact that, when employing a trial wavefunction, the method does not provide access to the true ground-state distribution [192].

The state of affairs as described is quite damning for the *ab initio* description of positronic bound states to molecules. Experiment has shown that positrons readily form bound states with many organic molecules [193–200]. Most of these molecules have not been explored with *ab initio* calculations, except for CI calculations for carbon disulfide  $CS_2$ , acetaldehyde  $C_2H_4O$ , propanal  $C_2H_5CHO$ , acetone  $(CH_3)_2CO$ , acetonitrile  $CH_3CN$ , and propionitrile  $C_2H_5CN$  [201–204]. All of these molecules, besides carbon disulfide, have large static dipole moments, and the CI predictions of their positron binding energies deviate from experiment by at least  $\sim 25\%$ . Furthermore, the calculations fail to predict positron binding for the carbon disulfide molecule. Clearly, there is a need for methods that offer an accurate description of the bound states between positrons and non-polar organic molecules where binding proceeds via a strong positronium forming character.

Recently, this shortcoming has been addressed by the work of Hofierka *et al.* [205–207], who develop a perturbative many-body theory of positron binding to molecules. This method produces positron binding energies in close agreement with experimental measurement for several non-polar organic species. In their work, Hofierka *et al.* highlight the limitations of QMC methods applied to positron binding that we have discussed. In this chapter, we address these limitations by demonstrating that the flexibility of NNVMC offers an accurate description of positronic bound

states with non-polar organic molecules, and represents a promising alternative to the many-body theory of Hofierka *et al.*

## 5.2 Multicomponent calculations in NNVMC

NNVMC offers a new approach to calculating the ground-state properties of positronic bound states. To treat multicomponent wavefunctions of electrons and positrons within the NNVMC method, we extend the architecture of the FermiNet ansatz.

Positrons are treated in a non-relativistic context by modifying the Born-Oppenheimer molecular Hamiltonian, eq. (2.3), to explicitly include the charge,  $q_i$ , of the quantum particles,

$$\mathcal{H} = -\frac{1}{2} \sum_{i=1}^N \nabla^2 + \sum_{i=1}^N \sum_{j>i}^N \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i=1}^N \sum_{j=1}^M \frac{q_i Z_j}{|\mathbf{r}_i - \mathbf{R}_j|} + \sum_{i=1}^M \sum_{j>i}^M \frac{Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j|}, \quad (5.1)$$

where  $q_i$  is the charge of the  $i^{\text{th}}$  quantum particle,  $\pm 1$  for positrons and electrons, respectively.

To incorporate positrons into the FermiNet ansatz, we draw inspiration from the treatment of the distinct spin species of electrons in a spin-assigned wavefunction. We treat positrons as nothing more than an additional species of distinguishable particle, and extend the spin indexing of the parameters appearing in the definition of the FermiNet (section 3.3.4), to also index the charge of the particle. The message constructed in the FermiNet layers is thus written,

$$\mathbf{f}_{l,i}^{\alpha q} = \left( \mathbf{h}_{l,i}^{\alpha q}, \mathbf{g}_l^{\beta p}, \mathbf{g}_{l,i}^{\alpha \beta q p} \quad \forall \beta, p \right), \quad (5.2)$$

where  $\alpha/\beta \in \{\uparrow, \downarrow\}$  label spin,  $p/q \in \{+, -\}$  label charge, and we have retained the bracketed notation to indicate the concatenation of vectors (see section 3.3.4). As in eqs. (3.42) and (3.43), the aggregated  $\mathbf{g}$  vectors are obtained by summing over the appropriate ranges of electron indices. For multi-positron systems, this construction also enforces fermionic exchange symmetry for positrons of the same spin assignment, but does not enforce any exchange symmetry under the exchange of the spatial co-ordinates of positrons and electrons, or positrons with different spin assignments.

In the original formulation of the FermiNet, the parameters of the neural network layers are shared between the two spin species of electrons, e.g. the weight matrices appearing in eq. (3.44) do not carry a spin index. In closed shell molecules, this seems like a well-justified choice: the physics of the electronic orbitals in that case should be identical for both spin species. For positrons, however, the nature of the positronic orbitals is quite different from that of the electronic orbitals. To

account for this, we modify the FermiNet such that a distinct set of neural network parameters are defined for each distinguishable species of particle. The nodal update rule, eq. (3.44), becomes

$$\mathbf{h}_{l+1,i}^{\alpha q} = \tanh(\mathbf{W}_l^{\alpha q} \mathbf{f}_{l,i}^{\alpha q} + \mathbf{b}_l^{\alpha q}) + \mathbf{h}_{l,i}^{\alpha q}, \quad (5.3)$$

and the edge update rule, eq. (3.45), becomes

$$\mathbf{h}_{l+1,ij}^{\alpha\beta qp} = \tanh(\mathbf{V}_l^{\alpha\beta qp} \mathbf{h}_{l,ij}^{\alpha\beta qp} + \mathbf{c}_l^{\alpha\beta qp}) + \mathbf{h}_{l,ij}^{\alpha\beta qp}, \quad (5.4)$$

where the weights and biases now carry labels that index the particle species. This multiplies the number of variational parameters used in the nodal updates by the number of distinguishable species, and in the edge updates by the square of the number of distinguishable species. The spin indices for the parameters of the orbital projection, eq. (3.46), and envelope, eq. (3.47), are modified analogously.

The Pfau-Spencer determinants produced by the modified FermiNet are again block diagonal, with the blocks being distinguished by the charge and spin of the particles to which they correspond. To extend the dense determinant scheme, we could choose to make the entire Pfau-Spencer determinant dense, in a sense allowing electrons to ‘occupy’ positronic orbitals. Alternatively, we could use dense determinants which retain the block diagonal structure with respect to the charge of the particles. We adopt the latter convention for dense determinants in the present work, but we have not explored the difference in results between the two choices.

To implement pre-training for positronic wavefunctions, we pre-train the electronic orbitals to the Hartree-Fock solution of the bare molecules. We do not pre-train the positronic orbitals, which are randomly initialized at the beginning of training. All the calculations presented in the following section were carried out using the hyperparameters in table 3.3.

## 5.3 Results

We have chosen to benchmark the accuracy of NNVMC in describing positronic bound states by comparing ground-state energies, positron binding energies, and positron annihilation rates for some of the positronic systems which have been studied by ECG-SVM, CI, and DMC methods. Additionally, we calculate the positron binding energy of benzene and compare with experiment and the many-body theory of Hofierka et al., and provide a novel benchmark calculation of the positron binding energy of the dilithium molecule,  $\text{Li}_2$ . These systems show a wide variety of positron binding characteristics. Unless geometries are explicitly mentioned, the

computed positron binding energies are ‘vertical’, *i.e.* taken with respect to the same molecular geometry for the bare and positronic molecule. Our results for the positron binding energy of the sodium atom and benzene molecule were obtained by adopting a variance matching procedure similar to that of Entwistle *et al.* [114]. The specific details of our procedure are outlined in appendix D, illustrated using the results obtained for benzene.

### 5.3.1 Positron binding energies

#### Positronium hydride

The Ps atom forms a stable molecular complex with a neutral hydrogen atom. ECG-SVM calculations of the ground-state energy of this complex are available [208]. The ground-state energy obtained by ECG-SVM is  $E_0 = -0.7891794$  Hartrees. Using NNVMC with the FermiNet ansatz, we obtain a ground-state energy of  $E_0 = -0.789144(3)$  Hartrees, in agreement with the ECG-SVM result on the order of tens of microHartrees. Both of these methods exhibit variational error that could be reduced with additional computational expense (larger ECG basis, larger neural network ansatz, *etc.*), but the small difference in the results relative to typical chemical energy scales (milliHartrees) suggests that they are both nearly exact.

#### Sodium and magnesium atoms

The ionization energy of sodium is below the threshold for positronium formation, 0.25 Hartrees. Thus, the positronic bound state of the sodium atom is better described as a molecular complex of a Ps atom and sodium cation. The binding energy of this complex is given by the difference between the ground-state energy of the positronic complex and the sum of the energies of the dissociation fragments (*i.e.* the ground-state energy of the sodium cation and the 0.25 Hartrees binding energy of the Ps atom). We obtain a binding energy of 0.32(15) milliHartrees, in close agreement with the predictions of ECG-SVM calculations of 0.473 milliHartrees. Due to the small magnitude of the binding energy relative to the total energy of a sodium atom, we resort to utilizing a variance matching procedure akin to Entwistle *et al.* [114] to produce the relative binding energy. In the absence of the variance matching procedure, we do not predict binding between the Ps atom and sodium cation, *i.e.* the calculated ground state energy of the  $[\text{Na}^+, \text{Ps}]$  complex is greater than the calculated ground-state energy of  $\text{Na}^+$  and the binding energy of the Ps atom. This is due to the difference in the quality of the optimized trial wavefunction for the positronic system and the bare sodium cation. The need for variance matching could likely be eliminated with greater computational effort.

| LiH                          |                       |                    |                |
|------------------------------|-----------------------|--------------------|----------------|
| Method                       | Bare                  | Positronic         | Binding energy |
| FermiNet-VMC                 | -8.07051(1)           | <b>-8.10775(1)</b> | 0.03723(2)     |
| SJ-VMC                       | -8.06307(3) [188]     | -8.08034(4) [188]  | 0.01727(7)     |
| SJ-FN-DMC                    | -8.070045(38) [188]   | -8.10718(11) [188] | 0.03714(15)    |
| CISD                         | -8.03830 [181]        | -8.05530 [181]     | 0.017          |
| MRD-CI                       | -8.06827 [182]        | -8.09764 [182]     | 0.02937        |
| ECG-SVM                      | <b>-8.07054</b> [174] | -8.10747 [174]     | 0.03693        |
| Hofierka <i>et al.</i> [205] | -                     | -                  | 0.039(1)*      |

Table 5.1: Ground-state energy of LiH and its positronic bound state, obtained via various computational methods at equilibrium bond-length. Statistical errors are omitted where they are smaller than the reported precision or otherwise omitted in the referenced source. The lowest variational energy in each column is bolded. Table reproduced from [209]. (\*) Theoretical (not statistical) uncertainty estimated from the comparison between different levels of theory as described by Hofierka *et al.* [205], caption of Table I.

For the magnesium atom, we obtain a positron binding energy of 16.18(9) milliHartrees. This agrees well with previous ECG-SVM (16.93 milliHartrees) and CI (17.099 milliHartrees) results [178]. Variance matching is unnecessary in this case, due to the large, relative to the variational uncertainty of our calculations, positron binding energy.

### Lithium hydride and beryllium oxide

We have calculated the ground-state potential energy surface of the bare LiH molecule and its positronic complex around the equilibrium interatomic separations, shown in fig. 5.1 (see appendix E for full data tables). Fitting these potential energy surfaces via Nesterov’s algorithm, implemented in the **MOLCAS** package, yields equilibrium interatomic separations of  $3.0196 a_0$ , with ground-state energy -8.07050(1) Hartrees, for the bare molecule, and  $3.371 a_0$  with a ground-state energy of -8.10774(1) Hartrees for the positronic bound state. These deviate very slightly from the widely accepted literature values of  $3.015 a_0$  for the bare molecule and  $3.348 a_0$  for the positronic bound state. These deviations are likely due to differences in the quality of the converged trial wavefunctions across the potential energy surface, which could be improved with additional computational effort, the adoption of transferable wavefunction models, or a denser sampling over the potential energy surface. We have calculated the ground-state energies at the canonical equilibrium separations for comparison with results obtained via other ab initio methods. These results are compared in table 5.1. NNVMC is seen to yield a positron binding energy within

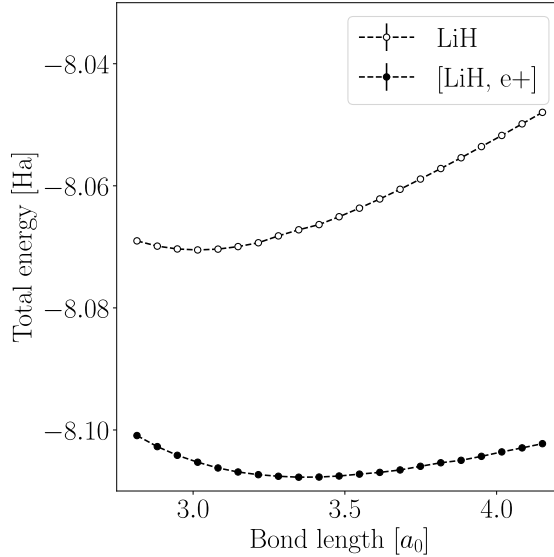


Figure 5.1: Ground-state potential energy surface of the bare LiH molecule and its positronic complex around the equilibrium interatomic separation.

| BeO          |                     |                     |                |
|--------------|---------------------|---------------------|----------------|
| Method       | Bare                | Positronic          | Binding energy |
| FermiNet-VMC | <b>-89.90572(4)</b> | <b>-89.93082(3)</b> | 0.02510(7)     |
| SJ-VMC       | -89.3173(25) [185]  | -89.3365(13) [185]  | 0.0192(38)     |
| SJ-FN-DMC    | -89.7854(13) [185]  | -89.8134(12) [185]  | 0.02800(25)    |
| MRD-CI       | -89.759352 [183]    | -89.773133 [183]    | 0.013781       |

Table 5.2: Ground-state energy of BeO and its positronic bound state at equilibrium interatomic separation. Statistical errors are omitted where they are smaller than the reported precision or otherwise omitted in the referenced source. The lowest variational energy in each column is bolded. Table reproduced from [209].

sub-milliHartree agreement with the most accurate other method considered (ECG-SVM),<sup>3</sup> and a state-of-the-art variational result for the ground-state energy of the positronic bound state.

We have applied the same procedure to the potential energy surface of the bare BeO molecule and its positronic bound state over a wider range of interatomic separations. We obtain an equilibrium interatomic separation of  $2.515 a_0$ , corresponding to a ground-state energy of  $-89.90572(4)$  Hartrees, for the bare molecule, and  $2.530 a_0$ , corresponding to a ground-state energy of  $-89.93082(3)$  Hartrees, for the positronic bound state. These results are compared against SJ-VMC, SJ-DMC,

<sup>3</sup>Note that the binding energy is a non-variational quantity. When we speak of the relative ‘accuracy’ of the compared methods, we are considering which method produces the lowest variational ground-state energies – with the exception of the many-body method of Hofiera *et al.*, which produces predictions for the binding energy directly.

and MRD-CI calculations in table 5.2, which that NNVMC yields the most accurate variational result for the bare molecule and for its positronic bound state.

The equilibrium electronic ground state of BeO is a spin singlet that transitions to a spin triplet at dissociation. The dissociated spin state is simply a consequence of Hund’s rules, yielding a triplet state for the oxygen atom and a singlet state for the beryllium atom (leaving an overall triplet state for the combined system). We enforce the appropriate ground-state spin configuration at wide interatomic separations by a spin assignment of  $s_z = 1$  (preventing convergence to the singlet state), and locate the singlet-triplet transition by comparing against the potential energy surface for a spin assignment of  $s_z = 0$ . These are referred to as ‘singlet’ and ‘triplet’ states in fig. 5.2(b), although this is a slight abuse of nomenclature as the  $s_z = 0$  wavefunction converges to the triplet state above the transition separation. We investigate the positronic density, potential energy surface, dipole moment, and positron binding energy as a function of interatomic separation, shown in fig. 5.2. These reveal several interesting features. At an interatomic separation of  $\sim 4 a_0$ , we find that the positron binding energy drops dramatically (but does not vanish) near the transition between the two ground-state spin configurations. This is due to a sharp change in the static dipole moment of the bare molecule at the transition between the two electronic states. Additionally, at an interatomic separation of  $\sim 6 a_0$ , we find a minimum in the positron binding energy. From inspection of the positronic density in fig. 5.2(a), the minimum corresponds to a smooth transition in the qualitative character of the positronic bound state, which changes from binding to the static dipole field of the bare molecule to binding primarily with the neutral Be atom. Interestingly, the positronic density becomes primarily localized in the *center* of the bond at the transition separation. At the widest interatomic separations, the positron binding energy is in good agreement with the positron binding energy of the lone beryllium atom as predicted by previous CI calculations [175]. This can be anticipated *a priori*, as a lone oxygen atom does not form a positronic bound state.

## Dilithium

We study  $\text{Li}_2$  as a simple example of a non-polar molecule to demonstrate the ability of NNVMC to describe purely correlation driven positron binding to molecules. We obtain a ground-state energy at the equilibrium bond length ( $5.051 a_0$ ) of  $-14.99485(2)$  Hartrees for the bare molecule, and  $-15.06162(2)$  Hartrees for the positronic bound state, yielding a positron binding energy of  $66.75(2)$  milliHartrees. The positronic density of the positronic bound-state of  $\text{Li}_2$  is shown in fig. 5.3. We see that the positronic component of the wavefunction is localized around the covalent bond of the molecule in a toroidal geometry. The positronic bound state of this molecule has not been studied using any other method, and we expect these results

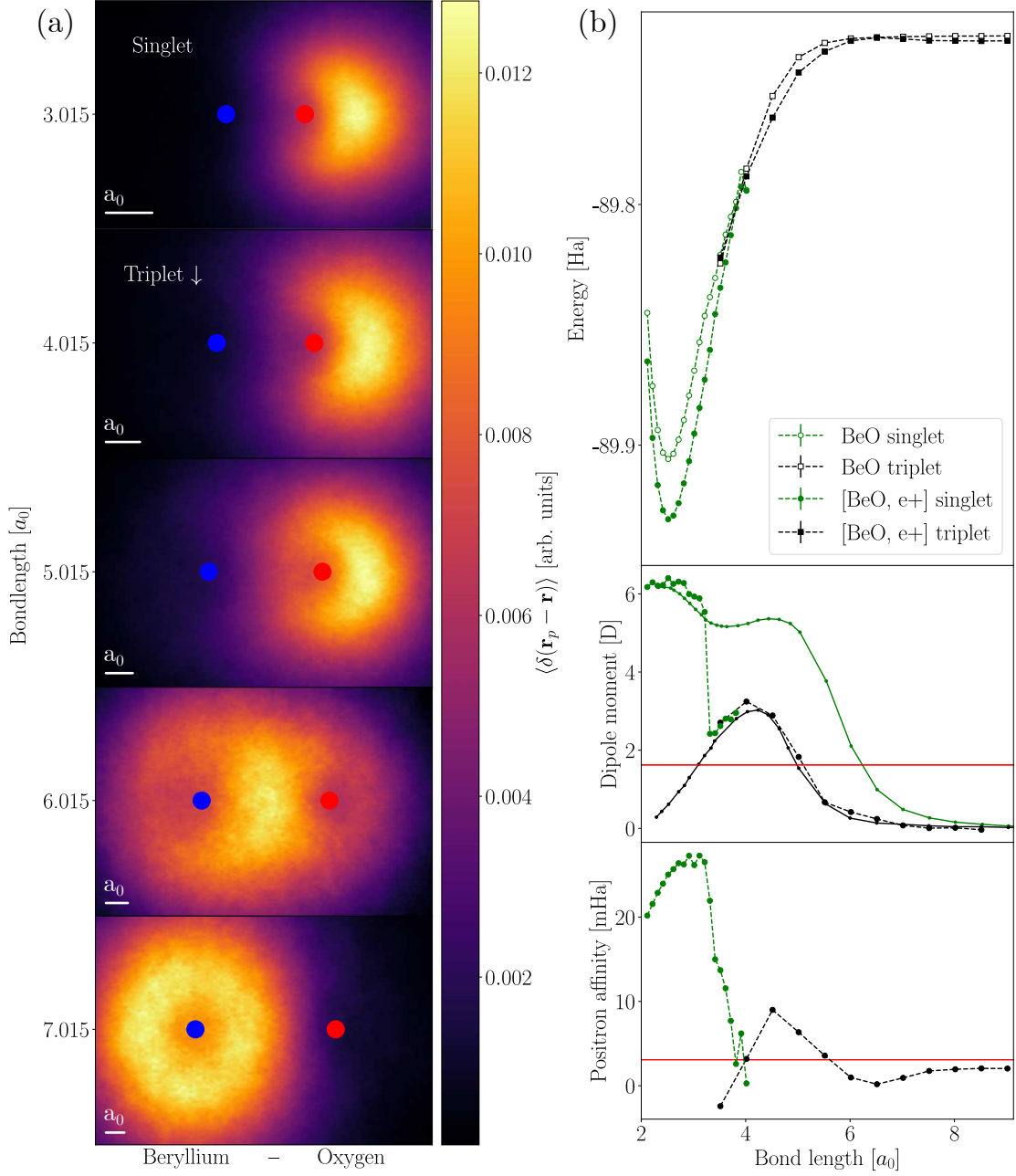


Figure 5.2: **(a)** Ground-state positronic one-particle density of BeO as a function of bond length (interatomic separation). Beryllium and oxygen nuclei are indicated by blue and red dots, respectively. The density is accumulated by histogramming positron positions from MCMC samples drawn from an optimized FermiNet trial wavefunction. Distances have been rescaled to enable visual comparison; the scale bar at the bottom left of each panel indicates one Bohr radius,  $a_0$ . **(b)** Energy (of the bare molecule and positronic complex), bare molecular dipole moment, and positron affinity of BeO as a function of bond length. The solid lines and dots in the dipole moment plot are taken from Bunker *et al.* [183]. The horizontal red lines in the dipole moment and positron affinity plots indicate the critical dipole moment for positron binding to the static dipole field, and the positron affinity of a lone beryllium atom [175], respectively. Figure reproduced from [209].

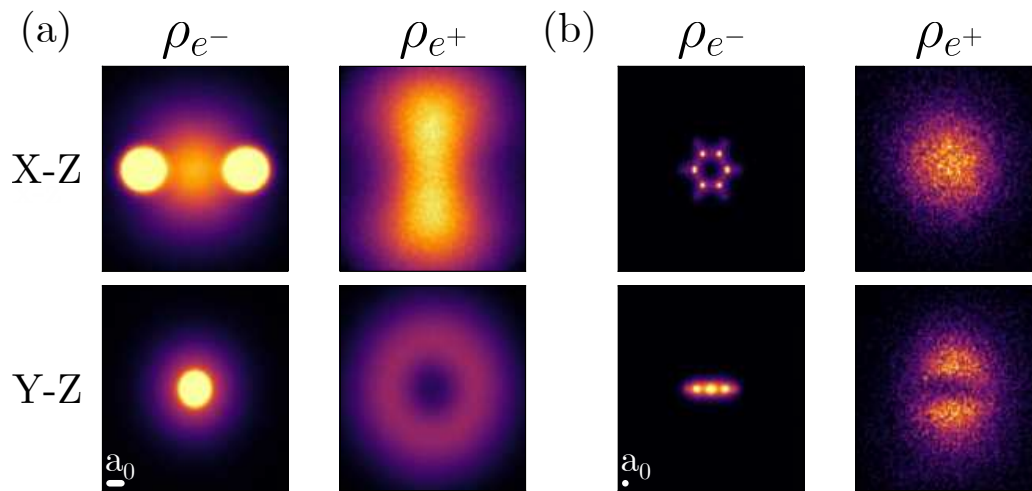


Figure 5.3: Orthographic projections of the ground-state density for positronic **(a)** dilithium and **(b)** benzene molecules. The left and right columns of each sub-figure show the electronic and positronic density,  $\rho_{e-}$  and  $\rho_{e+}$ . In the dilithium molecule, the positronic component of the wavefunction is localized around the covalent bond of the molecule in a toroidal geometry. In the benzene molecule, the positronic component of the wavefunction forms a highly diffuse cloud resembling a  $p$ -type orbital ‘sandwiching’ the delocalized  $\pi$ -bonding orbitals in the aromatic ring. The Monte Carlo accumulation of the positronic density for benzene is noisy due to the highly diffuse nature of the positronic component of the wavefunction. The scale bar in the bottom left of each subfigure indicates one Bohr radius,  $a_0$ . Figure reproduced from [209].

will provide a simple benchmark for the ability of other computational methods to describe purely correlation driven positronic bound states.

## Benzene

To demonstrate the ability of NNVMC to obtain positron binding energies for experimentally relevant non-polar organic species, we calculate the positron binding energy of the benzene molecule. Utilizing the variance matching procedure of Entwistle [114], we obtain a positron binding energy of 4.1(3) milliHartrees. This is in close agreement with the positron binding energy for benzene obtained by Hofierka *et al.* of  $\sim 4.26$  milliHartrees [205], later refined to  $\sim 5.44$  milliHartrees [207], and the most recent experimental measurement of the positron binding energy of benzene of 4.85(11) milliHartrees [207].

### 5.3.2 Annihilation rates

The dominant annihilation process in positronic bound-states is due to two-body collisions between a positron and an electron with opposing spin, producing two  $\gamma$ -ray photons. This is known, appropriately, as  $2\gamma$  annihilation. Given access to the many-body wavefunction, the  $2\gamma$  annihilation rate is proportional to the limit of the pair density – calculated for particles  $i$  and  $j$  as the expectation value of the operator  $\delta(\mathbf{r}_i - \mathbf{r}_j)$  – between a positron and an electron as their mutual separation tends to zero (the ‘pair contact density’). For a positronic bound state containing a single positron, the  $2\gamma$  annihilation is written

$$\Gamma_{2\gamma} = \pi r_0^2 c \sum_{i=1}^N \frac{\langle \Psi | \mathcal{O}_i^s \delta(\mathbf{r}_i - \mathbf{r}^+) | \Psi \rangle}{\langle \Psi | \Psi \rangle}. \quad (5.5)$$

where  $r_0$  is the classical electron radius,  $c$  is the speed of light,  $\mathbf{r}_i$  is the co-ordinate of the  $i^{\text{th}}$  electron,  $\mathbf{r}^+$  is the co-ordinate of the positron, and  $\mathcal{O}_i^s$  is a projection operator for the singlet state of the positron-electron pair. For a spin-assigned wavefunction, the action of the projection operator is equivalent to evaluating the expectation value over only one of the two spin species. To avoid wasting statistics, this can be enacted by averaging the operator over both spin species.

Accumulating eq. (5.5) directly via Monte Carlo integration presents a challenge: the exact coalescence point between the electrons and a positron will never be sampled by a finite Markov chain. Instead, we follow the procedure described by Simula [210]. We evaluate the spherically averaged pair density between electrons

and positrons at finite separation during Monte Carlo sampling,

$$g(r) = \frac{1}{2} \left\langle \frac{\sum_{i=1}^N \delta(|\mathbf{r}_i - \mathbf{r}^+| - r)}{4\pi r^2} \right\rangle_p, \quad (5.6)$$

where the factor of  $1/2$  arises because of the averaging over spin species resulting from the singlet projection operator,  $\mathcal{O}_i^s$ . The  $2\gamma$ -annihilation rate is proportional to the limit of  $g(r)$  as  $r \rightarrow 0$ :

$$\Gamma_{2\gamma} = 100.9 \lim_{r \rightarrow 0} g(r), \quad (5.7)$$

where the prefactor has been expressed such that the final expression for  $\Gamma_{2\gamma}$  is given in terms of inverse nanoseconds, *i.e.*  $\pi r_0^2 c = 100.9 a_0^3 \text{ ns}^{-1}$ ,<sup>4</sup> and  $g(r)$  is calculated in atomic units (*i.e.*, units of  $a_0^{-3}$ ). To evaluate the limit of the pair density in practice, we accumulate electron-positron separations into a histogram of fixed-width bins with appropriate normalization, and fit the resulting data to provide an extrapolated value as  $r \rightarrow 0$ . Extrapolation is necessary due to the vanishing volume of the spherical bins as  $r \rightarrow 0$ . Errors in the bin counts are estimated as

$$\sigma_r \simeq \sqrt{C_r} \quad (5.8)$$

where  $C(r)$  is the number of counts in the bin centered at  $r$ . We fit  $g(r)$  to a fifth-order polynomial via the Levenberg-Marquadt algorithm as implemented in the `SciPy` package, excluding bins with zero counts. The first-order coefficient of the polynomial fit is fixed at  $-1$  to satisfy the Kato cusp condition on the pair density. The final error in  $g(r \rightarrow 0)$  is obtained from the variance of the 0<sup>th</sup> order parameter in the polynomial fit.

---

<sup>4</sup>In atomic units,  $c = \alpha^{-1}$  where  $\alpha \simeq 1/137$  is the fine-structure constant. Using the relation  $e^2/r_0 = m_e c^2$ , and setting  $e = m_e = 1$ , we find that  $r_0 = c^{-2}$  and  $\pi r_0^2 c = \pi c^{-3} = \pi \alpha^3 \simeq 1.22 \times 10^{-6} a_0^3 E_h / \hbar$ . Converting  $\hbar / E_h$  to nanoseconds, this becomes  $\sim 50.46 a_0^3 \text{ ns}^{-1}$ . Finally, multiplying by a factor of 2 to account for the redundancy in the spin assignment of the positron yields  $\sim 100.9$ .

| Method                       | HPs          | [Na <sup>+</sup> , Ps] | 2 $\gamma$ annihilation rate [ns <sup>-1</sup> ] |                        |                        |                                     |                            |
|------------------------------|--------------|------------------------|--------------------------------------------------|------------------------|------------------------|-------------------------------------|----------------------------|
|                              |              |                        | [Mg, e <sup>+</sup> ]                            | [LiH, e <sup>+</sup> ] | [BeO, e <sup>+</sup> ] | [Li <sub>2</sub> , e <sup>+</sup> ] | [Benzene, e <sup>+</sup> ] |
| CI                           | 2.0183 [177] | —                      | 1.000991 [178]                                   | 0.8947* [211]          | —                      | —                                   | —                          |
| FN-DMC                       | 2.32 [186]   | —                      | —                                                | 1.3602* [187]          | —                      | —                                   | —                          |
| ECG-SVM                      | 2.4361 [208] | 1.898 [214]            | 0.955 [215]                                      | 1.375 [174]            | —                      | —                                   | —                          |
|                              | 2.4722 [212] |                        | 1.0249 [216]                                     |                        |                        |                                     |                            |
|                              | 2.4685 [213] |                        |                                                  |                        |                        |                                     |                            |
| Hofierka <i>et al.</i> [205] | —            | —                      | —                                                | 2.083                  | —                      | —                                   | 0.666                      |
| NNVMC                        | 2.440(1)     | 1.870(1)               | 1.076(1)                                         | 1.3391(7)              | 0.9533(8)              | 1.962(2)                            | 0.5247(6)                  |

\* Contact density at equilibrium bond length of 3.348  $a_0$  obtained by quadratic interpolation of three values around the minimum in the potential energy surface.

Table 5.3: 2 $\gamma$  annihilation rates for positronic atoms and molecules calculated via NNVMC and other computational methods. Statistical errors are omitted where they are smaller than the reported precision, or otherwise omitted in the referenced source. ECG results for all species besides HPs utilize the fixed-core approximation. Table reproduced from [209].

We have calculated annihilation rates according to this procedure for the same systems whose positron binding energies we presented in the previous section. These are compared with existing estimates, where available, in table 5.3. Of the methods compared, we expect that ECG-SVM best captures the features of the wavefunction that govern the positron annihilation rates. We find that the annihilation rates produced by NNVMC are in close agreement with the ECG-SVM results for positronium hydride, lithium hydride, and the alkali metal atoms. This suggests that the NNVMC optimized trial wavefunction accurately describes the electron-positron correlations. In contrast, the many-body theory of Hofierka *et al.* appears to overestimate the annihilation rate in lithium hydride compared to ECG-SVM. On this basis, we conclude that NNVMC offers a promising competitor to their method, yielding accurate annihilation rates in addition to binding energies in close agreement with benchmarks and experiments.

## 5.4 Discussion and conclusion

The selection of systems we have presented covers the full range of modes of positron binding. In all cases, NNVMC yields highly accurate, in some cases state-of-the-art, positronic ground-state energies, and the resulting binding energies are in close agreement with theoretical and (where available) experimental estimates. This is achieved with an identical wavefunction ansatz for every system studied, again highlighting the flexibility of neural network wavefunctions. As in chapter 4, we do not employ any fine-tuning or system-specific treatments; the neural network is not supplied with any information about the nature of the ground-state wavefunction.

These results advance the state-of-the-art in applying QMC methods to the treatment of positronic bound states. Crucially, neural-network trial wavefunctions produce sufficiently accurate approximations of the ground-state wavefunction at the VMC level that we do not have to resort to an additional DMC calculation. In addition, our results for benzene represent the first QMC calculation of a positron binding energy for an experimentally relevant non-polar organic species. This clearly demonstrates that NNVMC can yield accurate ground-state properties of positronic bound states for modes of positron binding governed entirely by strong electron-positron correlations. Indeed, fig. 5.3 offers an intuitive understanding of how positron binding occurs in such systems: binding is enabled by the presence of regions of increased electronic density away from atomic nuclei. In  $\text{Li}_2$ , the increased electron density is the covalent bond; in benzene, it is the delocalized  $\pi$ -bonding electrons.

There are two primary drawbacks of the NNVMC method for describing positronic bound states. The first is computational expense, a problem in all applications of NNVMC. The large computational prefactor makes calculations for molecules con-

taining more than  $\sim 50$  quantum particles prohibitively expensive. This would represent the main bottleneck in undertaking a large scale NNVMC survey of all experimentally relevant species, including large organic molecules and organic molecules containing heavy elements such as sulfur. In the latter case, it is possible to substantially reduce the computational burden by employing pseudopotentials to remove most of the core electrons from the calculation. This approach has been successfully employed in the calculation of positron annihilation rates in solids with QMC [210]. The second drawback of NNVMC is the need to employ variance matching in cases where the positron binding energy is much smaller than the total ground-state energy of the host molecule. The development of methods to treat the uncontrolled error in relative energies due to a difference in quality of two optimized trial wavefunctions remains an open problem for the NNVMC method. Further refining the convergence of the optimized trial wavefunction by increasing the size of the neural network, tuning hyperparameters, or training for longer are all possible approaches for minimizing this error. However, none of these provides a general solution.

Our NNVMC approach to positronic chemistry is straightforward to extend to systems beyond those presented here. We expect that the flexibility of neural network wavefunctions will enable the treatment of more exotic positronic bound states, such as the larger polyelectronic clusters imagined by Wheeler, or molecules binding multiple positrons. Additionally, the extension of NNVMC to periodic boundary conditions described in chapter 4 can be combined directly with the multicomponent FermiNet ansatz we have described. This would allow the calculation of, *e.g.*, the ground-state wavefunction of a positron bound to semiconductor defects. From this, one could obtain the main quantities of interest in positron annihilation spectroscopy studies of defect populations in semiconductors, such as the angle-dependent momentum density of emitted  $\gamma$ -rays.

More generally, the ease with which the FermiNet was extended to describe multicomponent wavefunctions suggests that *any* continuous-space Hamiltonian is a suitable target for the application of the NNVMC method; no property specific to electronic wavefunctions underlies the success of NNVMC in describing ordinary molecules. Rather, NNVMC promises to offer a one-size-fits-all method for describing the ground-state properties of any system of quantum particles. We emphasize that this is due to NNVMC (when utilizing FermiNet-like ansätze) being a basis-set free method.

# Chapter 6

## Nuclear quantum effects

Most ground-state quantum chemical calculations are carried out with the kinetic energy of the nuclei neglected entirely, or treated under the Born-Oppenheimer approximation. However, in certain systems, this is insufficient. The behavior of systems containing light elements can be substantially influenced by ‘nuclear quantum effects’ due to the zero-point motion of the nuclei.

Nuclear quantum effects are most important in hydrogen-rich systems. The zero-temperature high-pressure phase diagram of solid hydrogen is a prominent example: computational studies have demonstrated that contributions from the nuclear kinetic energy have substantial qualitative effects on this phase diagram [217–219]. This is a topic of great interest due to a prediction by Ashcroft [220] that atomic metallic hydrogen, realized by subjecting solid molecular hydrogen to high pressures to induce the so-called ‘Wigner-Huntington’ transition, is superconducting at room temperature. There have been several recent experimental observations of near-room-temperature hydride superconductors [221–223]. As they are believed to be conventional BCS-type (*i.e.*, phonon mediated) superconductors, proper treatment of nuclear quantum effects may be crucial for predictive modelling of these materials.

Our treatment of positronic wavefunctions in the previous chapter demonstrated that NNVMC is a powerful method for approximating the ground-state wavefunction of multicomponent quantum systems. In this chapter, we will extend this approach to the treatment of the proton (the nucleus of the hydrogen atom) as a quantum particle. After all, in non-relativistic quantum mechanics we may treat a proton as naught but a heavy positron. First, we will lay out the mathematical formalism for the Born-Oppenheimer approximation, and make explicit the contributions to the Hamiltonian that are usually discarded, or treated approximately. Then, we will discuss traditional approaches to the treatment of nuclear quantum effects. Finally, we will describe how the PsiFormer ansatz can be modified to describe the multicomponent wavefunction of systems containing electrons and quantum protons. We will show that this ansatz, combined with the NNVMC method, can treat non-

relativistic molecular and solid-state systems consisting of electrons and protons without resorting to the Born-Oppenheimer approximation. We will devote particular attention to treating the impact of nuclear quantum effects on the ground-state phase diagram of solid hydrogen at high pressures.

## 6.1 The Born-Oppenheimer approximation redux

To derive the Born-Oppenheimer approximation, we will first reintroduce the nuclear kinetic energy operator into the molecular Hamiltonian. This is written,

$$\mathcal{H} = \mathcal{H}_e + \mathcal{T}_N \quad (6.1)$$

where  $\mathcal{H}_e$  is the molecular Hamiltonian of eq. (2.3), and  $\mathcal{T}_N$  is the nuclear kinetic energy operator. The nuclear kinetic energy operator is given by

$$\mathcal{T}_N = -\frac{1}{2} \sum_{i=1}^M \frac{1}{M_i} \nabla_{\mathbf{R}_i}^2, \quad (6.2)$$

where  $M_i$  is the mass of the  $i^{\text{th}}$  nucleus, and we have adjusted the subscript of the Laplacian operator to  $\mathbf{R}_i$  to make it clear that this acts on the *nuclear* co-ordinates.

At every nuclear geometry,  $\mathbf{R}$ , the electronic Hamiltonian permits a set of many-body electronic eigenfunctions,

$$\mathcal{H}_e \psi_n(\mathbf{r}; \mathbf{R}) = E_n(\mathbf{R}) \psi_n(\mathbf{r}; \mathbf{R}), \quad (6.3)$$

where  $\psi_n(\mathbf{r}; \mathbf{R})$  is the  $n^{\text{th}}$  electronic eigenfunction and  $E_n(\mathbf{R})$  is the corresponding electronic eigenvalue. As a function of the nuclear geometry, the electronic eigenvalues define a set of ‘potential energy surfaces’. The notation for the electronic eigenfunctions indicates that they are explicit functions of the electronic co-ordinates,  $\mathbf{r}$ , and that the form of these functions exhibits a (differentiable) dependence on the nuclear co-ordinates,  $\mathbf{R}$ .

For every nuclear geometry,  $\mathbf{R}$ , the electronic eigenfunctions form a complete basis. Thus, we may expand the eigenfunctions of the total Hamiltonian,  $\mathcal{H}$ , as

$$\Psi_m(\mathbf{r}, \mathbf{R}) = \sum_n \psi_n(\mathbf{r}; \mathbf{R}) \phi_{nm}(\mathbf{R}), \quad (6.4)$$

such that

$$\mathcal{H} \Psi_m(\mathbf{r}, \mathbf{R}) = \epsilon_m \Psi_m(\mathbf{r}, \mathbf{R}). \quad (6.5)$$

We identify the expansion coefficients  $\phi_{nm}(\mathbf{R})$  as the nuclear wavefunctions, for reasons that will be made concrete shortly. Note that while the index  $n$  is a good

quantum number for the electronic Hamiltonian, it is no longer a good quantum number for the total Hamiltonian.

Let us consider the substitution of eq. (6.4) into eq. (6.5), one term at a time. First, the electronic Hamiltonian,

$$\begin{aligned}\mathcal{H}_e \sum_n \psi_n \phi_{nm} &= \sum_n (\mathcal{H}_e \psi_n) \phi_{nm} \\ &= \sum_n E_n \psi_n \phi_{nm},\end{aligned}\tag{6.6}$$

where we have omitted the arguments  $\mathbf{r}$  and  $\mathbf{R}$  for brevity. Next, the nuclear kinetic energy operator,

$$\mathcal{T}_N \sum_n \psi_n \phi_{nm} = - \sum_i \sum_n \frac{1}{2M_i} [(\nabla_{\mathbf{R}_i}^2 \psi_n) \phi_{nm} + 2(\nabla_{\mathbf{R}_i} \psi_n) \cdot (\nabla_{\mathbf{R}_i} \phi_{nm}) + \psi_n (\nabla_{\mathbf{R}_i}^2 \phi_{nm})].\tag{6.7}$$

Substituting eqs. (6.6) and (6.7) into eq. (6.5) yields

$$\begin{aligned}\sum_n \left[ E_n \psi_n \phi_{nm} - \sum_i \frac{1}{2M_i} [(\nabla_{\mathbf{R}_i}^2 \psi_n) \phi_{nm} + 2(\nabla_{\mathbf{R}_i} \psi_n) \cdot (\nabla_{\mathbf{R}_i} \phi_{nm}) + \psi_n (\nabla_{\mathbf{R}_i}^2 \phi_{nm})] \right] \\ = \epsilon_m \sum_n \psi_n \phi_{nm}.\end{aligned}\tag{6.8}$$

Closing on the left with  $\psi_k$  and integrating over the electronic co-ordinates yields a set of coupled equations for the nuclear wavefunctions,

$$\begin{aligned}E_k \phi_{km} - \sum_i \frac{1}{2M_i} \left[ \nabla_{\mathbf{R}_i}^2 \phi_{km} + \sum_n [\langle \psi_k | \nabla_{\mathbf{R}_i}^2 | \psi_n \rangle \phi_{nm} + 2 \langle \psi_k | \nabla_{\mathbf{R}_i} \psi_n \rangle \cdot \nabla_{\mathbf{R}_i} \phi_{nm}] \right] \\ = \epsilon_m \phi_{km}.\end{aligned}\tag{6.9}$$

To simplify our discussion, we can write this as a matrix equation and group the diagonal and off-diagonal terms:

$$\mathbf{H} \phi_m = (\mathbf{H}_d + \mathbf{H}_{od}) \phi_m = \epsilon_m \phi_m,\tag{6.10}$$

where  $\phi_m$  is a vector with entries  $\phi_{km}$ . The diagonal terms are

$$(\mathbf{H}_d)_{kn} = \left[ E_k - \sum_i \frac{1}{2M_i} (\nabla_{\mathbf{R}_i}^2 + \langle \psi_k | \nabla_{\mathbf{R}_i}^2 | \psi_n \rangle) \right] \delta_{kn}.\tag{6.11}$$

Note that the  $\langle \psi_k | \nabla_{\mathbf{R}_i} \psi_n \rangle$  term does not contribute to the diagonal entries: assuming

time reversal invariance (and, therefore, a real wavefunction) the gradient of an orthonormalized eigenfunction must be orthogonal to the eigenfunction itself, in order to preserve normalization. The off-diagonal terms are

$$(\mathbf{H}_{\text{od}})_{kn} = - \sum_i \frac{1}{2M_i} [\langle \psi_k | \nabla_{\mathbf{R}_i}^2 | \psi_n \rangle (1 - \delta_{kn}) + 2 \langle \psi_k | \nabla_{\mathbf{R}_i} \psi_n \rangle \cdot \nabla_{\mathbf{R}_i} ] . \quad (6.12)$$

The second term on the right-hand side of eq. (6.12) is known as the ‘vibronic coupling’. This can be further simplified by considering the identity,

$$\langle \psi_k | \nabla_{\mathbf{R}_i} \psi_n \rangle = \frac{\langle \psi_k | [\mathcal{H}_e, \nabla_{\mathbf{R}_i}] | \psi_n \rangle}{E_k - E_n}, \quad (6.13)$$

where  $[A, B]$  is the commutator of  $A$  and  $B$ . The only term in  $\mathcal{H}_e$  that does not commute with  $\nabla_{\mathbf{R}_i}$  is the electro-nuclear attraction. Thus, we find that

$$[\mathcal{H}_e, \nabla_{\mathbf{R}_i}] = -Z_i \sum_{j=1}^N \frac{\mathbf{r}_j - \mathbf{R}_i}{|\mathbf{r}_j - \mathbf{R}_i|^3}. \quad (6.14)$$

This commutator is nothing but the sum of Coulomb forces between the  $i^{\text{th}}$  nucleus and the electrons. After substitution, the right-hand side of eq. (6.13) is a vector whose components are given by the matrix element between electronic eigenstates due to the components of the Coulomb force. That is, we may interpret the vibronic coupling as the mixing between eigenstates of the electronic Hamiltonian due to the mutual drag between electrons and nuclei under the nuclear zero-point motion.

Now let us consider the terms in  $\langle \psi_k | \nabla_{\mathbf{R}_i}^2 | \psi_n \rangle$ . Given the identity,

$$\nabla_{\mathbf{R}_i} \cdot \langle \psi_k | \nabla_{\mathbf{R}_i} \psi_n \rangle = \langle \nabla_{\mathbf{R}_i} \psi_k | \nabla_{\mathbf{R}_i} \psi_n \rangle + \langle \psi_k | \nabla_{\mathbf{R}_i}^2 | \psi_n \rangle, \quad (6.15)$$

we find, after rearranging and inserting a resolution of the identity, that

$$\langle \psi_k | \nabla_{\mathbf{R}_i}^2 | \psi_n \rangle = \nabla_{\mathbf{R}_i} \cdot \langle \psi_k | \nabla_{\mathbf{R}_i} \psi_n \rangle - \sum_l \langle \nabla_{\mathbf{R}_i} \psi_k | \psi_l \rangle \cdot \langle \psi_l | \nabla_{\mathbf{R}_i} \psi_n \rangle. \quad (6.16)$$

For the diagonal entries,  $k = n$ , the first term on the right-hand side vanishes, and we retain only the square of the vibronic coupling matrix, eq. (6.13). This is known as the ‘diagonal Born-Oppenheimer correction’. The general relationship between the off-diagonal entries of eq. (6.16) and eq. (6.13) is not so clear-cut, but it is typical to assume that these entries are small when the entries of the vibronic coupling matrix are small.

With the preceding analysis in hand, we are in a position to define successive approximations to the solution of the coupled electro-nuclear problem, eq. (6.5). The first approximation we will consider – that is so ubiquitous in quantum chemistry

and electronic structure theory that it is usually taken implicitly – is the Born-Oppenheimer approximation.

From eq. (6.13), the entries of the vibronic coupling matrix are inversely proportional to the separation between the electronic potential energy surfaces. When these potential energy surfaces are well separated, the vibronic coupling can be neglected. Thus, we may neglect all the off-diagonal terms, eq. (6.12), and the diagonal Born-Oppenheimer correction; the matrix equation, eq. (6.10), simplifies to a set of uncoupled equations for the nuclear wavefunctions associated with each potential energy surface,

$$[E_n(\mathbf{R}) + \mathcal{T}_N] \phi_{nm}(\mathbf{R}) = \epsilon_m \phi_{nm}(\mathbf{R}). \quad (6.17)$$

The reason for calling  $E_n(\mathbf{R})$  a ‘potential energy surface’ now becomes clear: this is a TISE for the nuclear wavefunctions, with the potential energy term given by the (nuclear-configuration dependent) eigenvalues of the electronic Hamiltonian.

It might appear that something mysterious has happened here:  $n$  is a good quantum number for the solutions of eq. (6.17), but does not index the eigenvalues  $\epsilon_m$ . To reconcile this fact, recall that the total wavefunction, eq. (6.4), contains contributions from the nuclear wavefunction on every potential energy surface. We solve eq. (6.17) for all of these nuclear wavefunctions, and use them to construct the total eigenfunction  $\Psi_m(\mathbf{r}, \mathbf{R})$  with associated total eigenvalue  $\epsilon_m$ . To restrict ourselves to the consideration of a single potential energy surface, we argued that the uncoupled nature of eq. (6.17) allows us to prepare the wavefunction in a single electronic state. The neglect of the off-diagonal terms in eq. (6.10) is equivalent to making an adiabatic approximation that the electronic state remains unchanged under the zero-point motion of the nuclei. Thus, the total wavefunction takes the product form,

$$\Psi_{nm}^{\text{BO}}(\mathbf{r}, \mathbf{R}) = \psi_n(\mathbf{r}; \mathbf{R}) \phi_{nm}(\mathbf{R}), \quad (6.18)$$

and the associated total energy eigenvalues,  $\epsilon_{nm}$ , are labelled by the electronic and nuclear principal quantum numbers,  $n$  and  $m$ . This is allowed by the decoupling of the electronic and nuclear motion under the Born-Oppenheimer approximation. We could have alternatively derived the Born-Oppenheimer approximation by invoking this form of the wavefunction as a variational ansatz for the solution of eq. (6.5) and neglecting the action of the nuclear kinetic energy operator,  $\mathcal{T}_N$ , on the electronic component of the wavefunction.

Obtaining solutions of the electro-nuclear TISE under the Born-Oppenheimer approximation is a two-step process. First, the TISE for the electronic Hamiltonian is solved to obtain the electronic potential energy surfaces. Then, the nuclear TISE, eq. (6.17), is solved to obtain the associated nuclear eigenstates and total energy eigenvalues. While the neglect of the vibronic couplings has substantially simplified

the problem by decoupling the electronic and nuclear degrees of freedom, this is still a formidable task. Equation (6.17) is still a  $3M$ -dimensional TISE with  $M$  nuclei, that may possess mutual fermionic or bosonic exchange statistics. Even neglecting any exchange interaction due to the quantum statistics of the nuclei – an approximation that is typically justified by the near vanishing overlap between localized orbitals corresponding to separate nuclei – we must resort to further approximations or sophisticated numerical techniques to solve the nuclear eigenproblem.

The most common, and simplest, approximation used to solve for the nuclear wavefunctions is the ‘harmonic approximation’. Here, the electronic potential energy surfaces are taken to be approximately quadratic around the equilibrium positions of the nuclei,

$$\mathbf{R}_0 = \underset{\mathbf{R}}{\operatorname{argmin}} E_n(\mathbf{R}). \quad (6.19)$$

This is equivalent to truncating the Taylor expansion of  $E_n(\mathbf{R})$ , about  $\mathbf{R}_0$ :

$$E_n(\mathbf{R}) \simeq E_n(\mathbf{R}_0) + \frac{1}{2} \sum_{ij} \left. \frac{\partial^2 E_n}{\partial R_i \partial R_j} \right|_{\mathbf{R}_0} (R_i - R_{0,i})(R_j - R_{0,j}), \quad (6.20)$$

where  $i$  indexes the components of the  $3M$ -dimensional vector  $\mathbf{R}$ . Note that the first-order term in this expansion vanishes at the equilibrium positions of the nuclei. Under this second-order truncation, eq. (6.17) describes a set of coupled quantum harmonic oscillators. The normal modes and associated fundamental frequencies of the nuclear vibrational motion are obtained by diagonalizing the mass-weighted Hessian,

$$H_{ij} = \frac{1}{\sqrt{M_i M_j}} \left. \frac{\partial^2 E_n}{\partial R_i \partial R_j} \right|_{\mathbf{R}_0}. \quad (6.21)$$

A significant advantage of the harmonic approximation is that the electronic eigenvalues only need to be calculated near the equilibrium positions of the nuclei, and their derivatives evaluated via, *e.g.*, finite differences.

The harmonic approximation is only valid for small displacements of the nuclear co-ordinates around their equilibrium positions, and breaks down completely when the potential energy surface is strongly anharmonic near the equilibrium positions. The former condition is violated for highly excited nuclear vibrational states. In such cases, the effect of anharmonicity is calculated via perturbation theory, often using model potentials to approximate the electron potential energy surfaces [224]. When the potential is strongly anharmonic, as is often the case in systems containing hydrogen bonds, perturbative approaches are no longer useful. Instead, one must obtain high-quality potential energy surfaces over larger regions of the nuclear configuration space, and resort to techniques like vibrational self-consistent field theory [225] to solve eq. (6.5). Path integral methods [226] are also commonly employed to

study the dynamics of such systems.

In the solid state, the effects of nuclear zero-point motion are incorporated by treating the crystal lattice as a set of coupled harmonic oscillators. This is precisely equivalent to the harmonic approximation outlined above. The normal modes of this system of oscillators are the ‘phonons’ of the crystal. Of course, this treatment implies the invocation of the Born-Oppenheimer approximation, and with it, the associated pitfalls. The effect of electron-phonon coupling, under the harmonic approximation, on response functions in solids is typically incorporated via perturbative approaches [227, 228]. Anharmonicity of the potential energy surface is included in this treatment by introducing interactions between the phonons. A thorough review of electron-phonon interactions, including non-adiabatic and anharmonic effects, is given by Giustino [229]. Numerous *ab initio* methods have been developed in the last two decades for the treatment of anharmonic nuclear quantum effects in solids [230–236].

The question remains: what to do when the Born-Oppenheimer approximation breaks down entirely? This occurs whenever the separation between electronic potential energy surfaces is small, either at avoided crossings or conical intersections. Often, this corresponds to nuclear geometries that permit degenerate electronic eigenvalues due to a point-group symmetry. Until relatively recently, it was a widely held belief that such cases are extremely rare in molecular systems, which generally lack point-group symmetries [237]. In the last two decades, this belief has been upended. Non-adiabatic processes are now understood to be common in many fields, including: the photochemistry of organic molecules [238], charge- and excitation-transfer reaction in biological processes [226, 239], and charge separation in photovoltaic materials [240]. Recently, there have been promising advances in the extension of traditional quantum chemistry to the treatment of ground-state non-Born-Oppenheimer effects in molecular systems through the nuclear electronic orbital family of methods [241].

The development of generic, fully *ab initio* methods for the treatment of the coupled electro-nuclear problem, incorporating non-adiabatic and anharmonic effects and the quantum statistics of the nuclei, is a formidable challenge. In the following sections, we present some modest progress in this direction within the NNVMC framework. We show that the NNVMC framework is straightforward to adapt to the description of multicomponent many-body wavefunctions of electrons and nuclei, and present results demonstrating that the method is capable of solving the TISE for the ground-state properties of these systems. We do not invoke the Born-Oppenheimer approximation, or require subsequent corrections to obtain the non-adiabatic contributions to the ground-state energy.

## 6.2 Quantum nuclei in NNVMC

In the previous chapter, we made a minimal set of modifications to the FermiNet architecture to enable the description of multicomponent wavefunctions. In this chapter, we will be working with the PsiFormer. Since we carried out the work on positronic chemistry, the PsiFormer (and derived architectures) has become the *de facto* ansatz for NNVMC, as attention-based architectures seem to yield superior ground-state energies versus message-passing architectures.<sup>1</sup>

In contrast to the implementation of the multicomponent FermiNet, we do not define additional parameters in the neural network backbone corresponding to individual particle species. Instead, we distinguish the treatment of the particle species in the neural network backbone by modifying the *input features*. This seems better justified in an attention-based neural network: we reason that the attention matrices are conditioned on the input features, such that the feature vectors of any given particle species will learn to attend to the salient components of the feature vectors of other particle species in the course of training. However, we still use distinct parameters per particle species in the linear projection layer used to produce the orbitals.

There are three properties of particles that we need to consider in treating nuclear quantum effects. These are charge, spin, and mass. Spin is already incorporated into the input features for the PsiFormer (eq. (3.50)). We include charge in exactly the same way, concatenating the charge of a particle to the corresponding input-feature vector. To incorporate mass, we do not want to use the bare masses of the particles, as these differ by several orders of magnitude for electrons and nuclei. Instead, we normalize the bare masses according to

$$\mu_i = \frac{m_i - m_{\min}}{m_{\max} - m_{\min}}, \quad (6.22)$$

where

$$m_{\max} = \max_i m_i \quad \text{and} \quad m_{\min} = \min_i m_i \quad (6.23)$$

and  $m_i$  is the mass of the  $i^{\text{th}}$  particle. The mass features,  $\mu_i$ , are thus restricted to the interval  $[0, 1]$ . For example, in a system containing electrons, protons, and deuterons:  $\mu_{\text{electron}} = 0$ ,  $\mu_{\text{proton}} = 0.5$ , and  $\mu_{\text{deuteron}} = 1.0$ . Like the spin and charge, the normalized mass is concatenated to the input-feature vector for each particle.

In developing the multicomponent FermiNet architecture, we did not explicitly consider enforcing the Kato cusp conditions between positrons and electrons. This

---

<sup>1</sup>We hypothesize that this is because the attention mechanism in the PsiFormer is a general approximation of a permutation invariant function over the electrons for ‘message construction’ – as opposed to the fixed choice of the sum over electron embeddings used in the FermiNet.

is because the cusp conditions are left to be learned by the FermiNet. However, in the PsiFormer, the cusp conditions are enforced explicitly by a Jastrow factor. In principle, we could modify the explicit Jastrow factor to account for the Kato cusp conditions, eq. (3.31), generalized to particles with arbitrary mass and charge [242]. Instead, we draw inspiration from the two-particle stream in the FermiNet and represent the Jastrow factor with an MLP, allowing the cusps to be learned. We have found this to be advantageous in describing systems containing particles with mutually attractive Coulomb interactions. We reason that, in such systems, there are substantial contributions to the energy expectation value from regions of configuration space where two mutually attracting particles are close together. Thus, accurately describing the wavefunction in these regions is of paramount importance in obtaining accurate ground-state energies (and stable computation of gradients).

Using an explicit Jastrow factor guarantees satisfaction of the Kato cusp conditions, but introduces a kind of ‘stiff’ curvature into the wavefunction *near* the cusp, that the neural network must learn to compensate. In contrast, using a neural network Jastrow factor allows the appropriate cusps in the wavefunction to emerge naturally during optimization, and offers greater flexibility in the representation of the wavefunction near the cusps. This is further justified by a related phenomenon due to the exponentially decaying envelopes used in systems with open boundary conditions. It has been observed that optimization of the parameters of the envelope actually corresponds to the most substantial changes in the energy expectation value during training.<sup>2</sup> These envelopes also introduce the kind of stiff curvature we are hoping to avoid by removing the explicit Jastrow.<sup>3</sup>

The MLP Jastrow factor is written,

$$J(r) = \sum_{i \neq j} \text{MLP}(\mathbf{j}_{ij}), \quad (6.24)$$

where

$$\mathbf{j}_{ij} = (q_i q_j, \sigma_i \sigma_j, \tilde{\mu}_{ij}, \mathbf{h}_{0,ij}, |\mathbf{h}_{0,ij}|), \quad (6.25)$$

with  $q_i$  and  $\sigma_i$  the charge and spin of the  $i^{\text{th}}$  particle, and

$$\tilde{\mu}_{ij} = \frac{\mu_{ij} - \mu_{\min}}{\mu_{\max} - \mu_{\min}} \quad (6.26)$$

---

<sup>2</sup>To our knowledge, this phenomenon has not been discussed in any published works, but we refer to private communications with Nicholas Gao and Michael Scherbela at the *Deep Learning meets Variational Monte Carlo* workshop, 2023.

<sup>3</sup>It may be desirable to explore more flexible forms of the envelope for the same reason; however, it is unclear how to enforce the decay of the wavefunction in open boundary conditions using, *e.g.*, an MLP envelope.

the normalized reduced mass of particles  $i$  and  $j$ . The reduced masses are defined,

$$\mu_{ij} = \frac{m_i m_j}{m_i + m_j}, \quad (6.27)$$

and the minimum and maximum reduced masses are defined,

$$\mu_{\min} = \min_{ij} \mu_{ij} \quad \text{and} \quad \mu_{\max} = \max_{ij} \mu_{ij}. \quad (6.28)$$

The  $\mathbf{h}_{0,ij}$  are the FermiNet two-particle input features defined in eq. (3.40). We use the input features instead of the bare particle separations, in order to automatically enforce the boundary conditions in periodic systems.

The final consideration in the design of our multicomponent ansatz for nuclear quantum effects is the treatment of the exchange statistics of quantum nuclei. In this work we will only be considering protons, which are fermions. If we wish to treat the fermionic statistics, we can do so in precisely the same way as we discussed for positrons, or opposite spin electrons: with either block or dense Pfau-Spencer determinants. However, the flexibility of our ansatz affords another option: we can also treat the protons as *distinguishable* particles. In that case, there is no restriction on the wavefunction under the exchange of the individual protons. We will refer to particles treated in this way as ‘boltzmannons’. We can implement this case by taking the protonic component of the wavefunction as a Hartree product, given by the trace of the protonic block(s) of the Pfau-Spencer determinant. The resulting ansatz could be written,

$$\Psi(\mathbf{r}) = \exp(J(\mathbf{r})) \sum_k \left( \det \left[ \phi_i^{k,e}(\mathbf{r}_j^e) \right] \right) \prod_{j=1}^{N_p} \phi_j^p(\mathbf{r}_j^p), \quad (6.29)$$

where the superscripts e and p are understood to distinguish electronic and protonic orbitals and co-ordinates, and  $N_p$  is the number of protons. We have suppressed the many-body notation for simplicity, but the  $\phi$  should be understood as the many-body Pfau-Spencer orbitals. This form of the protonic component of the wavefunction is strictly more general than a Hartree-product wavefunction due to the many-body nature of the orbitals. Note that the Pfau-Spencer orbitals remain equivariant under the exchange of boltzmannons; that is, each orbital is still a function of one particle’s coordinates and the *sets* of coordinates of other particles of each species, even when said species is composed of distinguishable particles.

Many treatments of nuclear quantum effects implicitly take boltzmannonic statistics by neglecting the nuclear exchange (for instance, taking the nuclear wavefunction as a Hartree product of harmonic oscillator wavefunctions in the harmonic approximation). Here, the primary utility of this approximation is a reduction in

computational expense: removing the requirement to evaluate a determinant for the nuclear component of the wavefunction reduces the asymptotic expense of a forward evaluation of this component from  $O(N^3)$  to  $O(N^2)$  (where  $N$  denotes the number of boltzmannons).

We will not treat any nuclei with bosonic exchange statistics<sup>4</sup> in the present work. However, it is simple to extend our treatment to this case. Instead of a determinant or Hartree-product, we draw inspiration from other NNVMC treatments of bosonic systems and propose writing the bosonic component of the wavefunction in the form of a Jastrow factor. For example, the wavefunction of a system consisting of electrons and deuterons could be written,

$$\Psi(\mathbf{r}) = \exp(J(\mathbf{r})) \sum_k \left( \det \left[ \phi_i^{k,e}(\mathbf{r}_j^e) \right] \right) \exp \left( \sum_{j=1}^{N_d} \phi_i^d(\mathbf{r}_j^d) \right), \quad (6.30)$$

where the superscript  $d$  labels deuteronic orbitals and co-ordinates.

At the start of training, we randomly initialize the nuclear co-ordinates in the MCMC walkers. In solid systems, this random initialization is drawn from a uniform distribution over the simulation cell. In molecular systems, this random initialization is drawn from a Gaussian distribution centered on the origin with a width chosen to approximately match the spatial extent of the Born-Oppenheimer geometry of the molecule. No other information about the Born-Oppenheimer co-ordinates of the nuclei is supplied during initialization of the network parameters or the co-ordinates of the MCMC walkers.

We did not employ pre-training in any of the computations presented here. It is not impossible to imagine how one would obtain appropriate electronic orbitals for the pre-training strategy adopted in chapter 5; however, this would require biasing the nuclear geometry. The parameters of the neural network are all randomly initialized according to Kaiming initialization [92] at the beginning of training. All the calculations presented were carried out using the hyperparameters in table 3.3.

## 6.3 Results and discussion

We have chosen three distinct test cases for NNVMC applied to multicomponent systems containing quantum protons and electrons: (i) the hydrogen dimer, the simplest system exhibiting non-trivial nuclear quantum effects; (ii) the water, heavy water, ammonia, and methane molecules, exhibiting distinct nuclear geometries to be learned by the trial wavefunction; and (iii) a small simulation cell of hydro-

---

<sup>4</sup>We will briefly discuss deuterons in treating isotopic substitution effects, but these will be treated as boltzmannons.

gen in periodic boundary conditions, exhibiting both structural (*i.e.*, protonic) and electronic phase transitions.

### 6.3.1 The hydrogen dimer

The hydrogen dimer represents an interesting special case. When the protons are treated as quantum particles, there are no fixed nuclei. Thus, the Hamiltonian is invariant under the simultaneous translation and rotation of all the electrons and protons. In practice, we break the translational symmetry by retaining an isotropic envelope in our trial wavefunction, centered with respect to an arbitrary origin. Note, however, that we modify the envelope to remove the usual electron-nuclear cusp. This is achieved by interpolating between a Gaussian envelope at short distances, and an exponentially decaying envelope at large distances to retain the correct boundary conditions. Written in the same form as eq. (3.47), this reads

$$\phi_i^{k\alpha}(\mathbf{r}_j) = \varphi_i^{k\alpha}(\mathbf{r}_j) [C(\mathbf{r}_j)\pi_{i,\text{long}}^{k\alpha} \exp(-\sigma_{i,\text{long}}^{k\alpha}|\mathbf{r}_j|) + (1 - C(\mathbf{r}_j))\pi_{i,\text{short}}^{k\alpha} \exp(-\sigma_{i,\text{short}}^{k\alpha}|\mathbf{r}_j|^2)], \quad (6.31)$$

where we have suppressed the many-body notation for the neural network orbitals,  $\alpha$  is a meta-index representing particle species (charge, spin, mass), and

$$C(\mathbf{r}) = \tanh(\gamma_i^{k\alpha}|\mathbf{r}|) \quad (6.32)$$

is a smooth cut-off function. The numbers  $\{\pi_{i,\text{short}}^{k\alpha}, \sigma_{i,\text{short}}^{k\alpha}, \pi_{i,\text{long}}^{k\alpha}, \sigma_{i,\text{long}}^{k\alpha}, \gamma_i^{k\alpha}\}$  are all variational parameters.

We have carried out NNVMC calculations for the two nuclear spin isomers of the hydrogen dimer: ‘parahydrogen’, where the nuclear spins form a singlet, and ‘orthohydrogen’, where the nuclear spins form a triplet. For orthohydrogen, we treat the protons with fermionic exchange statistics. The energy splitting between these isomers is known from experiment to be  $\sim 0.55$  milliHartrees [243]. We obtain a ground-state energy of -1.16366(1) Hartrees for parahydrogen, and -1.16317(1) Hartrees for orthohydrogen, predicting a splitting of 0.49(2) milliHartrees. This splitting results from the difference in ground-state nuclear angular momentum required by fermionic exchange statistics in each case –  $l = 1$  for the triplet and  $l = 0$  for the singlet.

Extremely accurate ECG calculations of the non-adiabatic ground-state energy of parahydrogen yield a ground-state energy of -1.164025 Hartrees [244], in close agreement with our results. At this level of precision, it could be that the kinetic energy penalty incurred by localizing the center-of-mass of the wavefunction explains the discrepancy: the kinetic energy penalty (versus a zero-momentum plane-wave) of creating a stationary Gaussian wavepacket of size  $\ell$  is on the order of  $1/m\ell^2$ . For

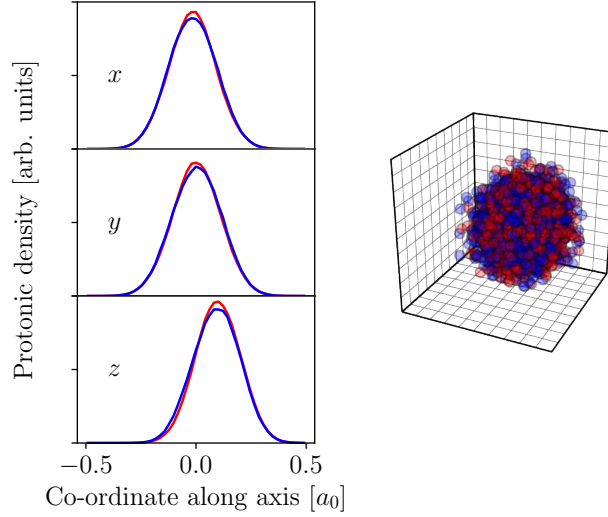


Figure 6.1: Visualisation of the protonic density of an optimized neural-network trial wavefunction for the orthohydrogen dimer with quantum protons. The line plots on the left represent histograms of the protonic density along the three Cartesian principal axes. The scatter plot on the right is produced from points accumulated from  $\sim 400$  thousand Monte Carlo samples of the proton configurations. The red and blue lines and markers correspond to the two spin species.

a particle of mass  $m \simeq 4000$ <sup>5</sup> and localization length  $\ell \simeq 1$ , this is of the right order of magnitude. Future work might disambiguate this possibility by carrying out a calculation in which only the relative coordinates of the particles with respect to the center-of-mass of the system are supplied to the network.

Despite its simplicity, the hydrogen dimer offers a reasonably comprehensive proof-of-principle for treating nuclear quantum effects with NNVMC. The close agreement between ECG and NNVMC for the ground-state energy of parahydrogen suggests that we are able to obtain accurate non-adiabatic ground-state energies. Furthermore, the close agreement between our calculated value and the experimental value for the energy splitting between the spin isomers of hydrogen demonstrates the ability of NNVMC to accurately treat nuclear exchange effects.

The one-particle protonic density of the optimized trial wavefunction for parahydrogen is plotted in fig. 6.1. We do not show the one-particle protonic density for orthohydrogen, which is nearly indistinguishable modulo the spin labelling. We see that, despite our attempts to enforce a fixed origin, the protonic density does not seem to be centered at the origin. Additionally, the rotational symmetry of the Hamiltonian is approximately obeyed. This is in stark contrast to the HEG results in chapter 4. There, we saw that the expected translational invariance of the one-electron density was broken in the Wigner crystal phase (*i.e.*, we did not observe the floating crystal). Given those results, we might have expected to see the typical

---

<sup>5</sup>Double the mass of a proton in atomic units, which is  $\sim 1836$ .

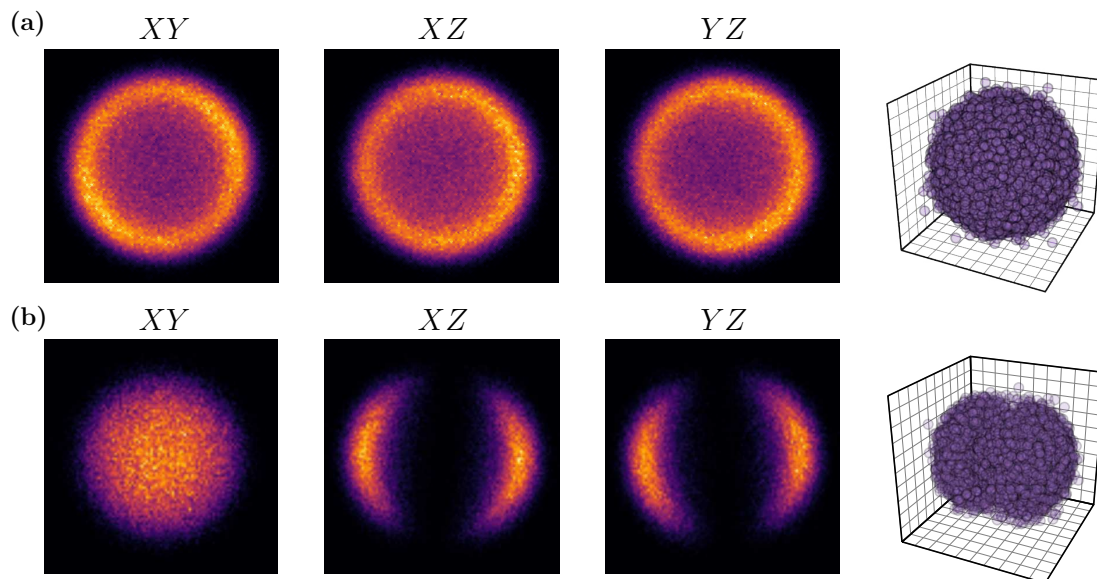


Figure 6.2: Histograms of the two-particle protonic pair density for **(a)** parahydrogen and **(b)** orthohydrogen optimized trial wavefunctions, projected into planes normal to the three Cartesian principal axes. Scatter plot on the right shows the samples used to produce the histograms, accumulated from  $\sim 400$  thousand Monte Carlo samples of the proton configurations.

symmetry-broken dumbbell shaped dimer associated with diatomic molecules in the protonic density. Instead, we find an almost spherical protonic density.

Figure 6.2 shows the two-particle protonic pair density (*i.e.*, histograms of the separations of protons) for parahydrogen and orthohydrogen. In the absence of an anisotropic external potential, this should be rotationally symmetric for the true ground-state wavefunction. This is (approximately) the case for two-particle protonic pair density of parahydrogen, but we find that the two-particle density of orthohydrogen exhibits the typical symmetry broken dimer arrangement of the protons. In fact, this is due to adopting a real-valued trial wavefunction. The protonic component of the orthohydrogen wavefunction only contains odd angular momentum components, primarily  $l = 1$ , due to fermionic antisymmetry. Complex-valued spherical harmonics in the pairing coordinate (‘p-orbitals’ for  $l = 1$ ) would exhibit an isotropic two-particle density. However, our real wavefunction must consist of superpositions of the complex-valued spherical harmonics. The interference pattern due to this superposition gives rise to the favored axis in the two-particle density. A detailed discussion of this effect, in the context of two-dimensional harmonic oscillator potentials, can be found in [245].

The proper treatment of high-symmetry systems remains an interesting open problem for NNVMC calculations. One might imagine following an approach similar to that adopted in ECG calculations. That is, express the wavefunction on a system

|                                                  | Molecule         |                  |                 |                 |
|--------------------------------------------------|------------------|------------------|-----------------|-----------------|
|                                                  | H <sub>2</sub> O | D <sub>2</sub> O | NH <sub>3</sub> | CH <sub>4</sub> |
| Energy                                           |                  |                  |                 |                 |
| $\epsilon_0$ [Ha]                                | -76.3914(1)      | -76.3940(1)      | -56.5152(1)     | -40.4649(1)     |
| Virial ratio                                     | 1.0020(2)        | 1.0032(2)        | 1.0019(2)       | 1.0020(2)       |
| Structure                                        |                  |                  |                 |                 |
| $\langle r_{\text{X-H}} \rangle$ [ $a_0$ ]       | 1.8419(1)        | 1.8246(1)        | 1.9448(4)       | 2.0969(1)       |
| $\sqrt{\text{Var}(r_{\text{X-H}})}$ [ $a_0$ ]    | 0.13063          | 0.10815          | 0.13386         | 0.14464         |
| Expt. $r_{\text{X-H}}$ [ $a_0$ ]                 | 1.8376 [246]     | 1.8305 [246]     | 1.9192 [247]    | 2.054(2) [248]  |
| $\theta_{\text{H-X-H}}$ [deg]                    | 105.62(1)        | 104.87(1)        | 107.70(4)       | 109.34(1)       |
| $\sqrt{\text{Var}(\theta_{\text{H-X-H}})}$ [deg] | 9.232            | 7.717            | 8.860           | 8.393           |
| Expt. $\theta_{\text{H-X-H}}$ [deg]              | 104.5 [246]      | 104.35 [246]     | 106.67 [243]    | 109.471 [249]   |

Table 6.1: Energy, ground-state virial ratio, hydrogen bond lengths and angles of water, heavy water, ammonia, and methane, evaluated via NNVMC. Experimental hydrogen bond lengths and angles from infrared spectroscopy are included for comparison.

of relative co-ordinates between quantum particles. However, it is difficult to express general antisymmetrizers for  $N$  particles in a system of relative co-ordinates. This is not a problem for systems with a few quantum particles like parahydrogen, because there are no particles with mutual fermionic exchange statistics present; however, the approach will not scale to general systems.

### 6.3.2 Small molecules

We have carried out NNVMC calculations for water, heavy water, ammonia, and methane with the heavy atomic nuclei (oxygen, nitrogen, and carbon) treated under the Born-Oppenheimer approximation, and the protons and deuterons treated as quantum particles with boltzmannonic exchange statistics. This kind of mixed treatment is common in non-Born-Oppenheimer quantum chemical approaches, like the nuclear electronic orbital family of methods [241]. The contribution of the nuclear kinetic energy operator for the heavy nuclei is much smaller than that of the protons, justifying its neglect. This can still be a faulty approximation in cases where electronic potential energy surfaces are very narrowly separated in regions near the equilibrium position of the heavy nuclei, but that is not the case here (shifting the single heavy nucleus being equivalent to making an arbitrary translation of the coordinate system).

Selected ground-state properties are listed in table 6.1. Experimental heavy atom-hydrogen bond lengths and angles obtained from infrared or microwave spectroscopy are also included for comparison. We find that our calculations produce virial ratios very close to unity, suggesting that we have obtained a good approx-

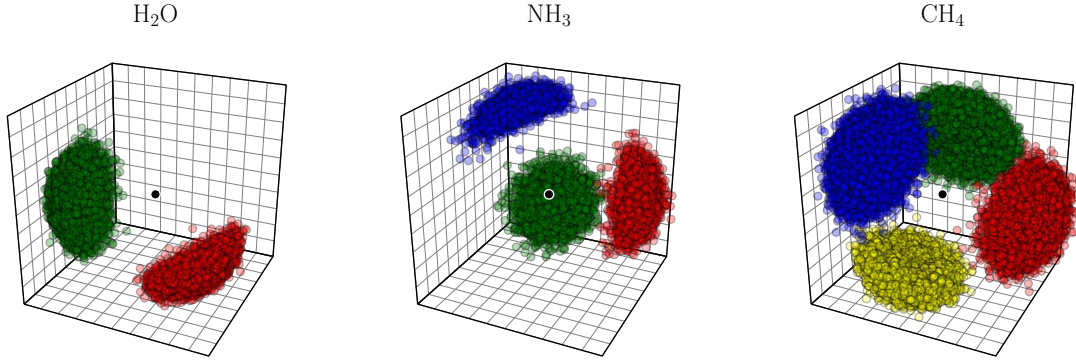


Figure 6.3: Scatter plots of the protonic density of three small molecules: water, ammonia, and methane. Points are accumulated from  $\sim 40$  thousand Monte Carlo samples of the proton configurations from optimized neural-network trial wavefunctions. Colours correspond to labels assigned by K-Means clustering of the proton positions, to distinguish, by their spatial location, what would be considered ‘distinguishable’ protons under the Born-Oppenheimer approximation. Central black dot with white outline indicates the (fixed) position of the heavy nucleus.

imation of the ground-state wavefunction. However, the expectation values of the heavy nucleus to proton bond lengths are slightly different from those obtained experimentally. We believe this is due to a fundamental difference in the definition of the ‘bond length’: experimental values are obtained by fitting the parameters of a potential energy surface to spectroscopic measurements, and locating the minima of those potential functions. This is more akin to the Born-Oppenheimer notion of equilibrium bond length, and does not necessarily coincide with the expectation value obtained from the ground-state wavefunction. In fact, in our calculations, there is no sensible notion of a potential energy surface: integrating out the electronic degrees of freedom in the local energy results in a sum over the Born-Oppenheimer potential energy surfaces of the ground and excited states, weighted by the nuclear component of the wavefunction on each potential energy surface.<sup>6</sup> Alternatively, the discrepancy between our bond lengths and the experimental values may suggest that our wavefunctions are not totally converged, as structural expectation values like bonding lengths and angles are more sensitive to errors in the wavefunction than

<sup>6</sup>To see why this is the case, consider the expectation value of the electronic Hamiltonian for the exact expansion of an eigenstate of the total Hamiltonian in terms of the Born-Oppenheimer electronic eigenstates, eq. (6.4). This is

$$\begin{aligned}
 \langle \Psi_m | \mathcal{H}_e | \Psi_m \rangle &= \sum_{nn'} \phi_{nm}(\mathbf{R}) \phi_{n'm}^*(\mathbf{R}) \langle \psi_n | \mathcal{H}_e | \psi_{n'} \rangle \\
 &= \sum_{nn'} \phi_{nm}(\mathbf{R}) \phi_{n'm}^*(\mathbf{R}) E_{n'} \delta_{nn'}(\mathbf{R}) \\
 &= \sum_n |\phi_{nm}(\mathbf{R})|^2 E_n(\mathbf{R}).
 \end{aligned}$$

the energy.

These systems are very well described by the Born-Oppenheimer approximation; our interest is primarily in demonstrating the ability of NNVMC to produce correct molecular structures without resorting to the Born-Oppenheimer approximation. Water, ammonia, and methane – due to the differing electronic structure of the heavy atom – display distinct molecular conformations: dihedral, trihedral, and tetrahedral, respectively. From the protonic density plots in fig. 6.3, we see that these classic structures emerge spontaneously during optimization of the wavefunction in an NNVMC calculation. Additionally, these results demonstrate the possibility of treating isotopic effects with NNVMC, through our observation of bond-shortening upon deuteration of the water molecule.

### 6.3.3 Eight protons in a box

As an initial foray into the solid state, we consider a system of ‘eight protons in a box’. This is a system in periodic boundary conditions, with a cubic simulation cell. We fix the spin-polarization of the electrons *and* the protons to zero, *i.e.* we assign 4 up-spin protons and electrons, and 4 down-spin protons and electrons, and we treat the protons with fermionic exchange statistics. The ground-state of this system is then governed by a single parameter, the Wigner-Seitz radius  $r_s$ . The Wigner-Seitz radius is still defined with respect to the average *electronic* (or, equivalent in this case, protonic) density, despite the number of quantum particles being doubled. Note that we are not fixing the *pressure* in our calculations, but the *volume* of the cell. While this system is far too small to enable quantitative claims about the zero-temperature phase diagram of solid hydrogen, it nonetheless (as we shall show) exhibits the essential physics of the real system, and serves as a useful prototype for demonstrating the capability of the NNVMC method in studying nuclear quantum effects in solid hydrogen.

As  $r_s$  is varied, we unambiguously observe the onset of the Wigner-Huntington transition. We observe both a *structural* transition from a simple cubic arrangement of the protons to a face-centered cubic arrangement of the protons, and an *electronic* metal-insulator transition. The ground-state energies we obtain as a function of  $r_s$  are listed in appendix E.

The structural phase transition is most easily observed by direct visual inspection of the single-particle protonic density. Orthographic and perspective projections of the protonic density, as a function of  $r_s$ , are shown in fig. 6.4.<sup>7</sup> At the two lowest values of  $r_s$  (high density), the protons assume a simple cubic arrangement

---

<sup>7</sup>We note that, akin to our results in chapter 4, we do not observe a homogeneous single-particle density despite the symmetry of the Hamiltonian under the simultaneous translation of all particles. We will not re-tread our discussion of this symmetry breaking here.

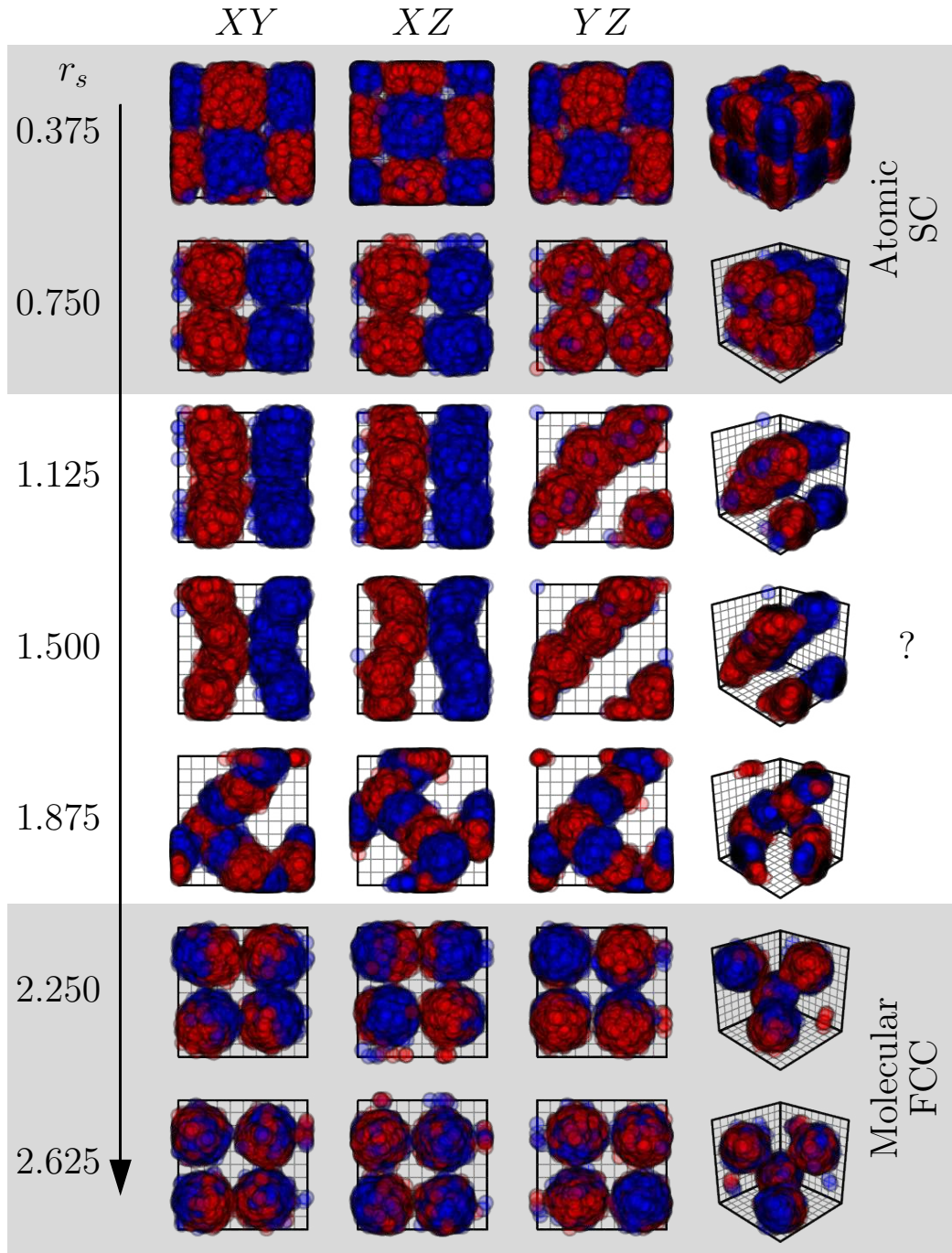


Figure 6.4: Scatter plots showing the protonic density as a function of Wigner-Seitz radius for the 8-protons-in-a-box system. Each point corresponds to a proton position obtained from Monte Carlo samples of proton configurations from an optimized neural-network trial wavefunction. Red and blue coloring distinguishes the two spin species of protons. The columns labelled  $XY$ ,  $XZ$ , and  $YZ$  show orthographic projections along the three crystal axes, and the rightmost column shows a perspective projection. In the high- and low-density regimes, the crystal structure is clear: atomic simple-cubic (SC) at high density and molecular face-centered cubic (FCC) at low density. In the intermediate regime, no clear crystal structure can be identified.

within the simulation cell with a single-atom basis. At the highest density, the protonic spins assume an antiferromagnetic arrangement. At the second highest density, the protonic spins assume planar ferromagnetic domains with the spins of adjacent planes ordered antiferromagnetically. It is difficult to conclude if the precise arrangement of the nuclear spins is real physics, or an artefact of an imperfect optimization due to the close proximity in energy of different nuclear spin states. However, it *is* clear that the spin species occupy mutually distinct lattice sites in both cases (*i.e.*, the nuclear spin state is not paramagnetic). At the two highest values of  $r_s$  (low density), the protons assume a face-centered cubic arrangement within the simulation cell. The basic units of the crystal structure are now hydrogen molecules exhibiting the *para* arrangement of the nuclear spins: each proton’s nearest neighbour possesses an oppositely aligned spin, consistent with the favored nuclear spin state of isolated hydrogen dimers. Indeed, the expectation value of the nearest-neighbour distance between protons tends to the value of a lone hydrogen molecule as the Wigner-Seitz radius increases (see fig. 6.5). Furthermore, we find that each molecule exhibits an *approximately* isotropic protonic density, and the axis of the molecules is not ‘locked’ by the crystal structure. It is worth noting that these effects could not be observed without a method able to account for the spin of the proton.

At the highest density, we see that the ‘blobs’ – which we might think of as ‘protonic orbitals’ in a mean-field treatment – in the proton density are no longer spherical, and are distorted into rounded cubes. We hypothesize that what is being observed here is similar to the Wigner transition: it is no longer energetically favorable for the localization length of the protonic orbitals at each site to continue to shrink with the size of the simulation cell. At some point, the potential energy penalty incurred by the ‘protonic orbitals’ overlapping on neighbouring sites is less than the kinetic energy penalty of further localization of the orbitals. Following Wigner’s argument [125, 126], we might expect that, under sufficient compression, we would observe another structural phase transition to a ‘protonic gas’. We leave the study of this amusing possibility to future work.

At intermediate densities, the story is less clear. In our opinion, the crystal structures obtained at these densities look non-physical. We offer two possible explanations for this. First, it is possible that in this small system, the structural transition between the atomic and molecular phases occurs over a broad range of densities at which there are a large number of nearly-degenerate crystal structures, hindering the convergence of training. Second, it is possible that the true ground-state crystal arrangement for eight-protons-in-a-box is not commensurate with a cubic simulation cell. The simple cubic arrangement is commensurate with a cubic simulation cell, forming a  $2\times 2\times 2$  arrangement of the single atom basis; and a single conventional unit cell of the face-centered cubic arrangement of molecules is

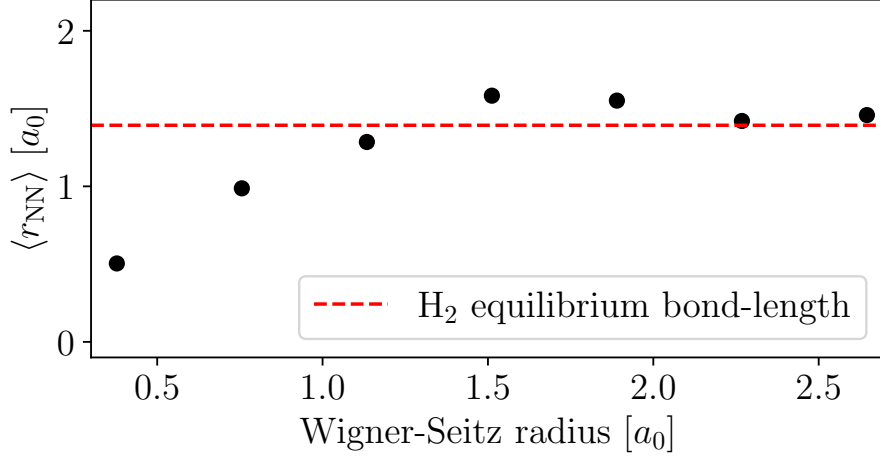


Figure 6.5: Nearest-neighbour separation as a function of Wigner-Seitz radius for the 8-protons-in-a-box-system, obtained as an expectation value over  $\sim 400$  thousand Monte Carlo samples from optimized-neural network trial wavefunctions.

also commensurate with a cubic simulation cell. However, the ground-state phase diagram of solid hydrogen is known to contain a crystal structures which are not commensurate with a cubic simulation cell containing 8 atoms. Disambiguating these possibilities would require a NNVMC calculation in which the simulation cell and Wigner-Seitz radius are not held fixed, and are instead continuously updated alongside the ground-state wavefunction to minimize the total free-energy of the system at a fixed *pressure*. We believe such calculations are feasible today: accumulating stress tensors via the Hellman-Feynman theorem is straightforward in the VMC framework [250], has been briefly explored by other authors [251]. This approach can be straightforwardly applied to minimize the Gibbs free energy with respect to the simulation cell lattice vectors [252]. However, we have not explored this possibility in the present work.

To demonstrate the electronic metal-insulator transition in a many-body context, we calculate the ground-state expectation value of the trace of the localization tensor of the electrons [253]. The diagonal components of the localization tensor are defined as

$$\langle r_\alpha^2 \rangle = -\frac{1}{N|\mathbf{G}_\alpha^2|} \log |Z^{(\alpha)}|^2, \quad (6.33)$$

where  $N$  is the number of electrons,  $\mathbf{G}_\alpha$  is a simulation-cell reciprocal lattice vector, and  $Z^\alpha$  is the expectation value of an operator which we dub the ‘many-body phase amplitude’,

$$Z^{(\alpha)} = \langle \Psi | \exp i \mathbf{G}_\alpha \cdot \left( \sum_{i=1}^N \mathbf{r}_i \right) | \Psi \rangle, \quad (6.34)$$

where the  $\mathbf{r}_i$  are understood to be the electronic co-ordinates. The traces of the localization tensor and the expectation value of the many-body phase amplitude are

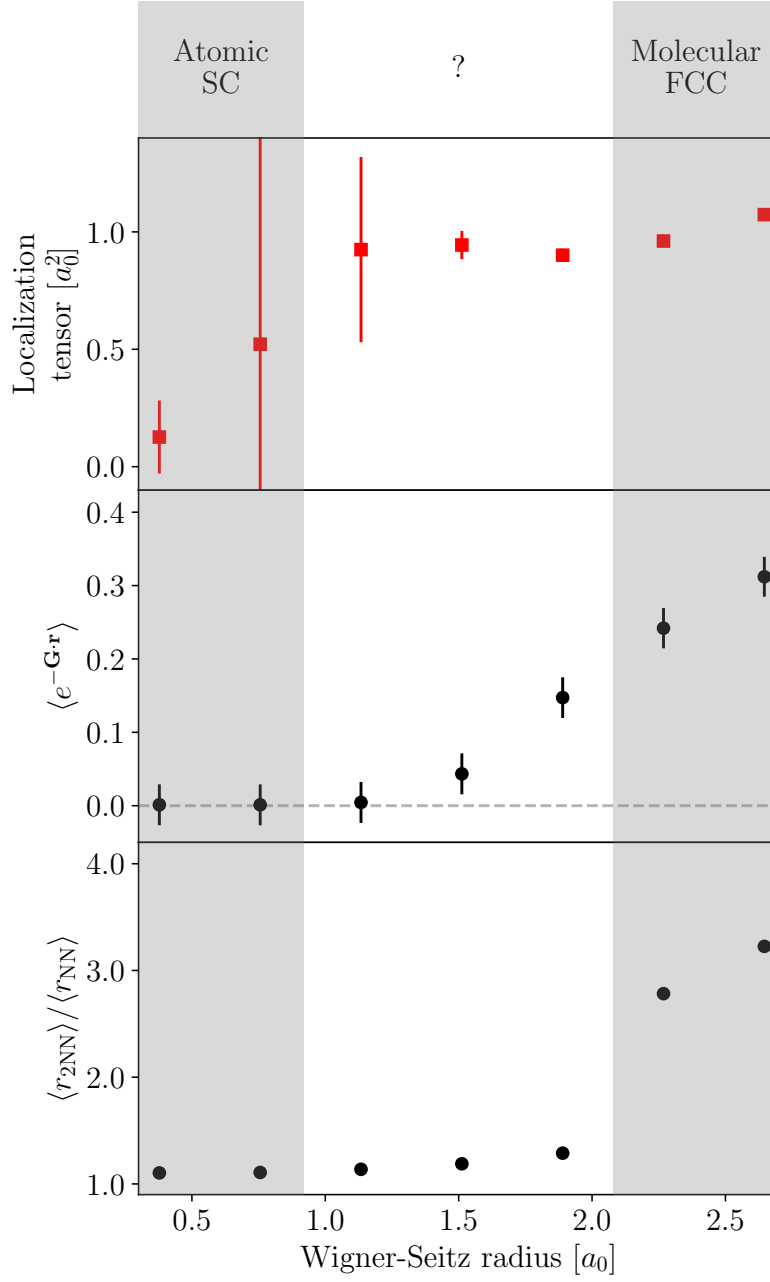


Figure 6.6: From top-to-bottom: trace of the electronic localization tensor, trace of the expectation value of the electronic many-body phase amplitude, and ratio of the expected protonic second- and first- nearest-neighbour separation as a function of Wigner-Seitz radius for the 8-protons-in-a-box system. All quantities are obtained as expectation values over  $\sim 400$  thousand Monte Carlo samples from optimized neural-network trial wavefunctions. The shaded regions indicate the atomic simple-cubic (SC) and molecular face-centered-cubic (FCC) phases obtained from visual inspection of the protonic density. The error bars for the electronic localization tensor and many-body phase amplitude are obtained as the square roots of the corresponding variances evaluated according to appendix B in [253]. The statistical error in the protonic separation ratio results in error bars smaller than the markers.

plotted in fig. 6.6. We see that, between  $r_s = 0.750$  and  $r_s = 1.50$ , the many-body phase amplitude rises from zero to a finite value. This is accompanied by a large spike in the expectation value and error bar of the localization tensor. This is characteristic of the onset of a metal-insulator transition [134, 253]. Intuitively, this corresponds to a characteristic shift in the nature of the electronic component of the ground-state wavefunction from spatially delocalized (*i.e.*, Fermi-gas-like itinerant electrons in a metal) to spatially localized (*i.e.*, tightly bound electrons in an insulator).

Determining the coincidence of the structural and electronic transitions is challenging due to the ambiguity of the crystal structure at intermediate densities. Such an analysis is further complicated by the finite size of the system prohibiting true ‘phase-transitions’ – all changes in the characteristic nature of the ground-state wavefunction occur over a continuous range of densities in a finite simulation cell. To attempt some quantitative comparison between the two transitions, we quantify the degree to which the system can be considered a ‘molecular’ solid by calculating the ratio between the expectation values of the second- and first- nearest neighbour distances for protons. This is plotted on the same axes as the localization tensor and many-body phase amplitude in fig. 6.6. We find that only at the two lowest densities studied – those that a visual inspection of the protonic density suggest are molecular solids – is the ratio significantly different from unity. From this analysis, we would conclude that the intermediate structures are ‘atomic’, and that these atomic structures appear at densities exhibiting the electronic structure of an insulator. That is, we observe three distinct phases: a molecular insulator at low density, an atomic insulator at intermediate densities, and an atomic metal at high density. However, we again remind the reader that these conclusions should not be considered indicative of the real physics of solid hydrogen due to the extremely small system size considered. Instead, one might consider these as properties of a novel form of matter: eight protons in a box.

The state-of-the-art in computing the high-pressure phase diagram of solid hydrogen is well summarized in a recent work of Monacelli *et al.* [219]. They consider various competing candidate phases for the phase diagram between 300 GPa and 700 GPa. They compute the ground-state electronic potential energy surface according to DFT with the B3LYP functional under the Born-Oppenheimer approximation. Atop this potential energy surface, they apply the ‘stochastic self-consistent harmonic approximation’ [235] – a variational method for determining free energy differences due to the nuclear zero-point motion on anharmonic potential energy surfaces – to obtain harmonic and anharmonic vibrational energies, and average proton positions. Finally, they compute the electronic internal energy by DMC with the protons treated as fixed nuclei at their average positions. The sum of the vibrational energies and electronic internal energies is used to compare the stability of

the candidate phases as a function of pressure.

The work of Monacelli *et al.* yields novel results regarding the ground-state phase diagram; however, that is not our primary concern. Rather, we highlight the fact that their work reveals that the phase diagram is qualitatively altered by enthalpy differences between competing phases, due to the nuclear zero-point motion, on the order of a few meV per proton. Specifically, they show that the small contribution of the anharmonic vibrational energy shifts the transition to the atomic phase by  $\sim 100$  GPa, and stabilizes an additional intermediate phase between  $\sim 400$  and  $\sim 600$  GPa. For our purposes, this raises an obvious question: do non-adiabatic and exchange contributions to the internal energy similarly alter this phase diagram? These contributions are neglected in the treatment of Monacelli *et al.*. Indeed, we might argue that, from exchange alone, we can expect a contribution on the order of meV per atom by considering the nuclear spin splitting of the hydrogen dimer. Furthermore, we might expect large non-adiabatic contributions in the *metallic* phase of solid hydrogen, where the gap between the ground and first excited electronic state is vanishingly small by definition. Thus, we argue that a non-Born-Oppenheimer treatment – such as we have presented here – should be further explored to fully disambiguate the nature of the high-pressure ground state phase diagram of solid hydrogen.

## 6.4 Conclusion

In this chapter, we extended a neural-network wavefunction ansatz, the PsiFormer, to describe multi-component systems of electrons and fully quantum protons. This wavefunction has been used to carry out NNVMC calculations of the ground-state properties of the hydrogen dimer, several small molecules, and a small simulation cell of solid hydrogen.

The variational energies we obtain for the hydrogen dimer compare favorably with results obtained with highly-accurate ECG calculations, with a small discrepancy that can be explained by the localization of the center-of-mass of our wavefunction. Furthermore, we calculate the splitting between the ortho and para nuclear spin isomers of the hydrogen molecule to within a few hundredths of a milliHartree of the known experimental value. The ground-state protonic densities we obtain for water, ammonia, and methane demonstrate that the NNVMC method is capable of reproducing the well-known structures of these molecules in a totally *ab initio* fashion.

Finally, our results for the eight-protons-in-a-box system act as a promising proof of principle for the application of the NNVMC method to a non-adiabatic treatment of the ground-state phase diagram of solid hydrogen. Despite the limited system

size studied here, we clearly observe the structural and electronic components of the Wigner-Huntington transition. We believe that with additional development, a proper treatment of simulation cell stresses to accommodate fixed pressure calculations, and much more computational effort, the NNVMC method could be used to understand the impact of non-adiabatic and exchange-driven nuclear quantum effects on the zero-temperature phase diagram of solid hydrogen.

# Chapter 7

## Conclusion

In this thesis, we have explored a series of novel applications of the recently developed NNVMC method. This method combines the flexibility of deep neural networks with the variational Monte Carlo framework to obtain highly accurate solutions to the many-body TISE with a favorable scaling compared to accurate quantum chemistry methods. We have focussed our investigation on novel applications of NNVMC facilitated by the lack of dependence on a set of basis functions for describing the many-body wavefunction.

In chapter 4, we developed an implementation of periodic boundary conditions for neural-network wavefunctions. This was achieved by adapting the common physicists’ intuition of the equivalence of a parallelepiped cell subject to periodic boundary conditions and the 3-torus. Using this intuition, we derived a set of neural network wavefunction ‘input features’ that enforce the many-body equivalent of Bloch’s theorem implied by periodic boundary conditions. These input features are carefully constructed to maintain continuity and smoothness of the wavefunction at the simulation cell boundaries, uniquely label all points in the simulation cell, and include the non-continuous derivative of the Euclidean norm of particle separations required to enforce the Kato cusp conditions.

We applied our implementation of periodic boundary conditions to the study of the homogeneous electron gas, a simplified model of electrons in metals with a long history of being studied via QMC methods. We demonstrated that NNVMC is capable of producing accurate ground-state energies, variationally exceeding the accuracy of traditional SJB-DMC calculations, and closely matching highly accurate i-FCIQMC results for small electron gases. Furthermore, we showed that this accuracy is retained as the electron correlation strength is increased (*i.e.*, as the density of the gas is lowered). Remarkably, we found that neural-network trial wavefunctions are capable of spontaneously converging to the appropriate ground-state symmetry – Fermi liquid or Wigner crystal – at either side of the Wigner transition. This was achieved with an identical wavefunction ansatz at all densities studied, in contrast

to other QMC methods which require a tailored choice of basis functions to converge to the appropriate ground-state symmetry.

Since this work was carried out, our implementation of periodic boundary conditions for NNVMC has also been applied in the study of the unitary Fermi gas [113]. These results, in addition to our own, demonstrate that NNVMC is a powerful tool for studying the ground-state properties of quantum gases. We believe that our work on the HEG could be extended to disambiguate outstanding questions about the ground-state phase diagram of the electron gas.

In chapter 5, we extended the FermiNet neural-network wavefunction ansatz to the description of *multi-component* many-body wavefunctions, enabling calculations for systems consisting of multiple kinds of quantum particles. In particular, we studied the binding between positrons and ordinary molecules. Obtaining accurate positron binding energies and annihilation rates for bound states with molecular matter has proven challenging for traditional quantum chemistry methods. Expanding the mixed electronic-positronic many-body wavefunction in a set of atom-centered basis functions yields very slow convergence with respect to the number of high angular-momentum basis functions in the basis set. This has historically limited the application of post-Hartree-Fock and QMC methods to strongly dipolar molecules (where positronic bound states can be understood at the level of simple electrostatics), and small atoms with only one or two valence electrons. We demonstrated that the NNVMC method is straightforward to adapt to the treatment of multi-component systems, and produces state-of-the-art variational energies for positronic systems across a wide variety of modes of positron binding.

Finally, in chapter 6, we realized a synthesis of the work in the previous two chapters. By extending the PsiFormer neural-network wavefunction ansatz to the description of multi-component wavefunctions and periodic boundary conditions, we studied nuclear quantum effects in hydrogenic systems. In particular, we demonstrated a proof-of-principle for mapping out the ground-state phase diagram of solid hydrogen under pressure through the Wigner-Huntington transition without resorting to the Born-Oppenheimer approximation. We found that our neural-network trial wavefunction was capable of converging to both a high-density atomic metal phase and a low-density molecular insulator phase. Describing this transition requires the trial wavefunction to simultaneously capture changes in the ‘crystal’ structure (*i.e.*, the protonic structure) and the electronic structure. This approach to describing dense solid hydrogen captures non-adiabatic contributions to the internal energy, and the effect of the nuclear spin structure. Therefore, we propose that NNVMC offers a route to conclusively establishing the ground-state phase diagram of dense solid hydrogen.

These applications paint a clear picture of the novel advantages of the NNVMC

method. Due to the flexibility of neural networks, particularly the lack of dependence on a set of basis functions, the method is – in principle – capable of describing the ground-state wavefunction of *any* Hamiltonian without foreknowledge of the nature of the ground-state. Thus, we conclude that the advantages of the method are most prominent in studying novel phases of matter where the relevant physics is not yet completely understood.



# Bibliography

1. Dirac, P. A. M. & Fowler, R. H. Quantum Mechanics of Many-Electron Systems. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **123**, 714–733 (Apr. 1929).
2. Szabo, A. & Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* ISBN: 978-0-486-69186-2 (Courier Corporation, July 1996).
3. Slater, J. C. The Theory of Complex Spectra. *Physical Review* **34**, 1293–1322 (Nov. 1929).
4. Čížek, J. On the Use of the Cluster Expansion and the Technique of Diagrams in Calculations of Correlation Effects in Atoms and Molecules. *Advances in Chemical Physics* **14** (Jan. 1969).
5. Bartlett, R. J. Many-Body Perturbation Theory and Coupled Cluster Theory for Electron Correlation in Molecules. *Annual Review of Physical Chemistry* **32**, 359–401. ISSN: 0066-426X, 1545-1593 (Oct. 1981).
6. Čížek, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *Journal of Chemical Physics* **45**, 4256–4266. ISSN: 0021-9606 (Dec. 1966).
7. Coester, F. & Kümmel, H. Short-Range Correlations in Nuclear Wave Functions. *Nuclear Physics* **17**, 477–485. ISSN: 0029-5582 (June 1960).
8. Raghavachari, K., Trucks, G. W., Pople, J. A. & Head-Gordon, M. A Fifth-Order Perturbation Comparison of Electron Correlation Theories. *Chemical Physics Letters* **157**, 479–483. ISSN: 0009-2614 (May 1989).
9. Feller, D. The Use of Systematic Sequences of Wave Functions for Estimating the Complete Basis Set, Full Configuration Interaction Limit in Water. *The Journal of Chemical Physics* **98**, 7059–7071. ISSN: 0021-9606 (May 1993).
10. Helgaker, T., Klopper, W., Koch, H. & Noga, J. Basis-Set Convergence of Correlated Calculations on Water. *The Journal of Chemical Physics* **106**, 9639–9646. ISSN: 0021-9606 (June 1997).

11. Feller, D., Peterson, K. A. & Hill, J. G. On the Effectiveness of CCSD(T) Complete Basis Set Extrapolations for Atomization Energies. *The Journal of Chemical Physics* **135**, 044102. ISSN: 1089-7690 (July 2011).
12. Dunning Jr., T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *The Journal of Chemical Physics* **90**, 1007–1023. ISSN: 0021-9606 (Jan. 1989).
13. Hohenberg, P. & Kohn, W. Inhomogeneous Electron Gas. *Physical Review* **136**, B864–B871 (Nov. 1964).
14. Levy, M. Electron Densities in Search of Hamiltonians. *Physical Review A* **26**, 1200–1208 (Sept. 1982).
15. Levy, M. Universal Variational Functionals of Electron Densities, First-Order Density Matrices, and Natural Spin-Orbitals and Solution of the v-Representability Problem. *Proceedings of the National Academy of Sciences* **76**, 6062–6065 (Dec. 1979).
16. Kohn, W. & Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **140**, A1133–A1138 (Nov. 1965).
17. Van Noorden, R., Maher, B. & Nuzzo, R. The Top 100 Papers. *Nature News* **514**, 550 (Oct. 2014).
18. Ceperley, D. M. & Alder, B. J. Ground State of the Electron Gas by a Stochastic Method. *Physical Review Letters* **45**, 566–569 (Aug. 1980).
19. Becca, F. & Sorella, S. *Quantum Monte Carlo Approaches for Correlated Systems* ISBN: 978-1-107-12993-1 (Cambridge University Press, Cambridge, 2017).
20. Ceperley, D. M., Chester, G. V., Kalos, M. H. & Whitlock, P. A. Monte Carlo Studies of Crystalline Helium. *Le Journal de Physique Colloques* **39**, C6–1304. ISSN: 0449-1947, 2777-3418 (Aug. 1978).
21. Ceperley, D. M. & Alder, B. J. Ground State of the Electron Gas by a Stochastic Method. *Physical Review Letters* **45**, 566–569 (Aug. 1980).
22. Mitas, L. Electronic Structure by Quantum Monte Carlo: Atoms, Molecules and Solids. *Computer Physics Communications* **96**, 107–117 (Aug. 1996).
23. Foulkes, W. M. C., Mitas, L., Needs, R. J. & Rajagopal, G. Quantum Monte Carlo Simulations of Solids. *Reviews of Modern Physics* **73**, 33–83. ISSN: 0034-6861, 1539-0756 (Jan. 2001).

24. Needs, R. J., Towler, M. D., Drummond, N. D. & López Ríos, P. Continuum Variational and Diffusion Quantum Monte Carlo Calculations. *Journal of Physics: Condensed Matter* **22**, 023201. ISSN: 0953-8984, 1361-648X (Jan. 2010).
25. Kolorenč, J. & Mitas, L. Applications of Quantum Monte Carlo Methods in Condensed Systems. *Reports on Progress in Physics* **74**, 026502. ISSN: 0034-4885 (Jan. 2011).
26. Needs, R. J., Towler, M. D., Drummond, N. D., López Ríos, P. & Trail, J. R. Variational and Diffusion Quantum Monte Carlo Calculations with the CASINO Code. *The Journal of Chemical Physics* **152**, 154106. ISSN: 0021-9606, 1089-7690 (Apr. 2020).
27. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* **21**, 1087–1092. ISSN: 0021-9606 (June 1953).
28. Kent, P. R. C., Needs, R. J. & Rajagopal, G. Monte Carlo Energy and Variance-Minimization Techniques for Optimizing Many-Body Wave Functions. *Physical Review B* **59**, 12344–12351 (May 1999).
29. Lin, X., Zhang, H. & Rappe, A. M. Optimization of Quantum Monte Carlo Wave Functions Using Analytical Energy Derivatives. *The Journal of Chemical Physics* **112**, 2650–2654. ISSN: 0021-9606 (Feb. 2000).
30. Bressanini, D., Morosi, G. & Mella, M. Robust Wave Function Optimization Procedures in Quantum Monte Carlo Methods. *The Journal of Chemical Physics* **116**, 5345–5350. ISSN: 0021-9606 (Apr. 2002).
31. Toulouse, J. & Umrigar, C. J. Optimization of Quantum Monte Carlo Wave Functions by Energy Minimization. *The Journal of Chemical Physics* **126**, 084102. ISSN: 0021-9606 (Feb. 2007).
32. Li, M., Zhang, T., Chen, Y. & Smola, A. J. *Efficient Mini-Batch Training for Stochastic Optimization* in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, New York, NY, USA, Aug. 2014), 661–670. ISBN: 978-1-4503-2956-9.
33. Umrigar, C. J. & Filippi, C. Energy and Variance Optimization of Many-Body Wave Functions. *Physical Review Letters* **94**, 150201 (Apr. 2005).
34. Nightingale, M. P. & Melik-Alaverdian, V. Optimization of Ground- and Excited-State Wave Functions and van Der Waals Clusters. *Physical Review Letters* **87**, 043401 (July 2001).

35. Umrigar, C. J., Toulouse, J., Filippi, C., Sorella, S. & Hennig, R. G. Alleviation of the Fermion-Sign Problem by Optimization of Many-Body Wave Functions. *Physical Review Letters* **98**, 110201 (Mar. 2007).
36. Amari, S.-i. Natural Gradient Works Efficiently in Learning. *Neural Computation* **10**, 251–276. ISSN: 0899-7667 (Feb. 1998).
37. Sorella, S. Wave Function Optimization in the Variational Monte Carlo Method. *Physical Review B* **71**, 241103 (June 2005).
38. Umrigar, C. J., Nightingale, M. P. & Runge, K. J. A Diffusion Monte Carlo Algorithm with Very Small Time-step Errors. *The Journal of Chemical Physics* **99**, 2865–2890. ISSN: 0021-9606 (Aug. 1993).
39. Needs, R. J., Towler, M. D., Drummond, N. D. & Ríos, P. L. Continuum Variational and Diffusion Quantum Monte Carlo Calculations. *Journal of Physics: Condensed Matter* **22**, 023201. ISSN: 0953-8984 (Dec. 2009).
40. Kim, J. *et al.* QMCPACK: An Open Source Ab Initio Quantum Monte Carlo Package for the Electronic Structure of Atoms, Molecules and Solids. *Journal of Physics: Condensed Matter* **30**, 195901. ISSN: 0953-8984 (Apr. 2018).
41. Andrews, J. S., Jayatilaka, D., Bone, R. G. A., Handy, N. C. & Amos, R. D. Spin Contamination in Single-Determinant Wavefunctions. *Chemical Physics Letters* **183**, 423–431. ISSN: 0009-2614 (Sept. 1991).
42. Drummond, N. D. & Needs, R. J. Variance-Minimization Scheme for Optimizing Jastrow Factors. *Physical Review B* **72**, 085124 (Aug. 2005).
43. Jastrow, R. Many-Body Problem with Strong Forces. *Physical Review* **98**, 1479–1484 (June 1955).
44. Kato, T. On the Eigenfunctions of Many-Particle Systems in Quantum Mechanics. *Communications on Pure and Applied Mathematics* **10**, 151–177. ISSN: 1097-0312 (1957).
45. Pack, R. T. & Brown, W. B. Cusp Conditions for Molecular Wavefunctions. *The Journal of Chemical Physics* **45**, 556–559. ISSN: 0021-9606 (July 1966).
46. Feynman, R. P. Atomic Theory of the Two-Fluid Model of Liquid Helium. *Physical Review* **94**, 262–277 (Apr. 1954).
47. Feynman, R. P. & Cohen, M. Energy Spectrum of the Excitations in Liquid Helium. *Physical Review* **102**, 1189–1204 (June 1956).
48. Wigner, E. & Seitz, F. On the Constitution of Metallic Sodium. II. *Physical Review* **46**, 509–524 (Sept. 1934).

49. Cybenko, G. Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control, Signals and Systems* **2**, 303–314. ISSN: 1435-568X (Dec. 1989).
50. Hornik, K., Stinchcombe, M. & White, H. Multilayer Feedforward Networks Are Universal Approximators. *Neural Networks* **2**, 359–366. ISSN: 0893-6080 (Jan. 1989).
51. LeCun, Y., Bengio, Y. & Hinton, G. Deep Learning. *Nature* **521**, 436–444. ISSN: 1476-4687 (May 2015).
52. Lu, Z., Pu, H., Wang, F., Hu, Z. & Wang, L. *The Expressive Power of Neural Networks: A View from the Width in Advances in Neural Information Processing Systems* **30** (Curran Associates, Inc., 2017).
53. Bishop, C. M. & Bishop, H. *Deep Learning: Foundations and Concepts* ISBN: 978-3-031-45468-4 (Springer Nature, Nov. 2023).
54. Raina, R., Madhavan, A. & Ng, A. Y. *Large-Scale Deep Unsupervised Learning Using Graphics Processors in Proceedings of the 26th Annual International Conference on Machine Learning* (Association for Computing Machinery, New York, NY, USA, June 2009), 873–880. ISBN: 978-1-60558-516-1.
55. Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. SchNet – A Deep Learning Architecture for Molecules and Materials. *The Journal of Chemical Physics* **148**, 241722. ISSN: 0021-9606 (Mar. 2018).
56. Noé, F., Tkatchenko, A., Müller, K.-R. & Clementi, C. Machine Learning for Molecular Simulation. *Annual Review of Physical Chemistry* **71**, 361–390. ISSN: 0066-426X, 1545-1593 (Apr. 2020).
57. Carleo, G. & Troyer, M. Solving the Quantum Many-Body Problem with Artificial Neural Networks. *Science* **355**, 602–606 (Feb. 2017).
58. Taddei, M., Ruggeri, M., Moroni, S. & Holzmann, M. Iterative Backflow Renormalization Procedure for Many-Body Ground-State Wave Functions of Strongly Interacting Normal Fermi Liquids. *Physical Review B* **91**, 115106 (Mar. 2015).
59. Ruggeri, M., Moroni, S. & Holzmann, M. Nonlinear Network Description for Many-Body Quantum Systems in Continuous Space. *Physical Review Letters* **120**, 205302 (May 2018).
60. White, S. R. Density Matrix Formulation for Quantum Renormalization Groups. *Physical Review Letters* **69**, 2863–2866 (Nov. 1992).

61. White, S. R. & Martin, R. L. Ab Initio Quantum Chemistry Using the Density Matrix Renormalization Group. *The Journal of Chemical Physics* **110**, 4127–4130. ISSN: 0021-9606 (Mar. 1999).
62. Saito, H. Solving the Bose–Hubbard Model with Machine Learning. *Journal of the Physical Society of Japan* **86**, 093001. ISSN: 0031-9015 (Sept. 2017).
63. Astrakhantsev, N. *et al.* Broken-Symmetry Ground States of the Heisenberg Model on the Pyrochlore Lattice. *Physical Review X* **11**, 041021 (Oct. 2021).
64. Denis, Z. & Carleo, G. *Accurate Neural Quantum States for Interacting Lattice Bosons* Apr. 2024. arXiv: 2404.07869.
65. Nomura, Y., Darmawan, A. S., Yamaji, Y. & Imada, M. Restricted Boltzmann Machine Learning for Solving Strongly Correlated Quantum Systems. *Physical Review B* **96**, 205152 (Nov. 2017).
66. Luo, D. & Clark, B. K. Backflow Transformations via Neural Networks for Quantum Many-Body Wave Functions. *Physical Review Letters* **122**, 226401 (June 2019).
67. Moreno, J. R., Carleo, G., Georges, A. & Stokes, J. Fermionic Wave Functions from Neural-Network Constrained Hidden States. *Proceedings of the National Academy of Sciences* **119**, e2122059119 (Aug. 2022).
68. Choo, K., Neupert, T. & Carleo, G. Two-Dimensional Frustrated J1-J2 Model Studied with Neural Network Quantum States. *Physical Review B* **100**, 125124 (Sept. 2019).
69. Viteritti, L. L., Rende, R. & Becca, F. Transformer Variational Wave Functions for Frustrated Quantum Spin Systems. *Physical Review Letters* **130**, 236401 (June 2023).
70. Viteritti, L. L., Rende, R., Parola, A., Goldt, S. & Becca, F. *Transformer Wave Function for the Shastry-Sutherland Model: Emergence of a Spin-Liquid Phase* Feb. 2024. arXiv: 2311.16889 [cond-mat].
71. Adams, C., Carleo, G., Lovato, A. & Rocco, N. Variational Monte Carlo Calculations of  $A < 4$  Nuclei with an Artificial Neural-Network Correlator Ansatz. *Physical Review Letters* **127**, 022502 (July 2021).
72. Choo, K., Mezzacapo, A. & Carleo, G. Fermionic Neural-Network States for Ab-Initio Electronic Structure. *Nature Communications* **11**, 2368. ISSN: 2041-1723 (May 2020).
73. Barrett, T. D., Malyshev, A. & Lvovsky, A. I. Autoregressive Neural-Network Wavefunctions for Ab Initio Quantum Chemistry. *Nature Machine Intelligence* **4**, 351–358. ISSN: 2522-5839 (Apr. 2022).

74. Zhao, T., Stokes, J. & Veerapaneni, S. Scalable Neural Quantum States Architecture for Quantum Chemistry. *Machine Learning: Science and Technology* **4**, 025034. ISSN: 2632-2153 (June 2023).
75. Liu, A.-J. & Clark, B. K. Neural Network Backflow for Ab Initio Quantum Chemistry. *Physical Review B* **110**, 115137 (Sept. 2024).
76. Han, J., Zhang, L. & E, W. Solving Many-Electron Schrödinger Equation Using Deep Neural Networks. *Journal of Computational Physics* **399**, 108929. ISSN: 0021-9991 (Dec. 2019).
77. Powell, J., Thiede, L., Asparu-Guzik, A. & Duvenaud, D. *Sorting Out Quantum Monte Carlo* Nov. 2023. arXiv: 2311.05598.
78. Pfau, D., Spencer, J. S., Matthews, A. G. D. G. & Foulkes, W. M. C. Ab Initio Solution of the Many-Electron Schrodinger Equation with Deep Neural Networks. *Physical Review Research* **2**, 033429 (Sept. 2020).
79. Spencer, J. S., Pfau, D., Botev, A. & Foulkes, W. M. C. *Better, Faster Fermionic Neural Networks* Nov. 2020. arXiv: 2011.07125 [physics].
80. Hermann, J., Schätzle, Z. & Noé, F. Deep-Neural-Network Solution of the Electronic Schrödinger Equation. *Nature Chemistry* **12**, 891–897. ISSN: 1755-4349 (Oct. 2020).
81. Schätzle, Z., Hermann, J. & Noé, F. Convergence to the Fixed-Node Limit in Deep Variational Monte Carlo. *The Journal of Chemical Physics* **154**, 124108. ISSN: 0021-9606 (Mar. 2021).
82. Hermann, J. *et al.* Ab Initio Quantum Chemistry with Neural-Network Wavefunctions. *Nature Reviews Chemistry* **7**, 692–709. ISSN: 2397-3358 (Oct. 2023).
83. Hutter, M. *On Representing (Anti)Symmetric Functions* July 2020. arXiv: 2007.15298.
84. von Glehn, I., Spencer, J. S. & Pfau, D. *A Self-Attention Ansatz for Ab-initio Quantum Chemistry* Apr. 2023. arXiv: 2211.13672 [physics].
85. Vaswani, A. *et al.* *Attention Is All You Need* in *Advances in Neural Information Processing Systems* **30** (Curran Associates, Inc., 2017).
86. Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. *Neural Message Passing for Quantum Chemistry* in *Proceedings of the 34th International Conference on Machine Learning - Volume 70* (JMLR.org, Sydney, NSW, Australia, Aug. 2017), 1263–1272.
87. He, K., Zhang, X., Ren, S. & Sun, J. *Deep Residual Learning for Image Recognition* in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), 770–778.

88. Lin, J., Goldshlager, G. & Lin, L. Explicitly Antisymmetrized Neural Network Layers for Variational Monte Carlo Simulation. *Journal of Computational Physics* **474**, 111765. ISSN: 0021-9991 (Feb. 2023).
89. Ba, J. L., Kiros, J. R. & Hinton, G. E. *Layer Normalization* July 2016. arXiv: 1607.06450 [cs, stat].
90. Bradbury, J. *et al.* JAX: composable transformations of Python+NumPy programs version 0.3.13. 2018. <http://github.com/google/jax>.
91. Glorot, X. & Bengio, Y. *Understanding the Difficulty of Training Deep Feed-forward Neural Networks* in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (JMLR Workshop and Conference Proceedings, Mar. 2010), 249–256.
92. He, K., Zhang, X., Ren, S. & Sun, J. *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification* Feb. 2015. arXiv: 1502.01852 [cs].
93. Kingma, D. P. & Ba, J. *Adam: A Method for Stochastic Optimization* Jan. 2017. arXiv: 1412.6980 [cs].
94. Martens, J. & Grosse, R. *Optimizing Neural Networks with Kronecker-factored Approximate Curvature* in *Proceedings of the 32nd International Conference on Machine Learning* (PMLR, June 2015), 2408–2417.
95. Flyvbjerg, H. & Petersen, H. G. Error Estimates on Averages of Correlated Data. *The Journal of Chemical Physics* **91**, 461–466. ISSN: 0021-9606 (July 1989).
96. Li, X., Li, Z. & Chen, J. Ab Initio Calculation of Real Solids via Neural Network Ansatz. *Nature Communications* **13**, 7895. ISSN: 2041-1723 (Dec. 2022).
97. Li, R. *et al.* A Computational Framework for Neural Network-Based Variational Monte Carlo with Forward Laplacian. *Nature Machine Intelligence* **6**, 209–219. ISSN: 2522-5839 (Feb. 2024).
98. Chen, A. & Heyl, M. Empowering Deep Neural Quantum States through Efficient Optimization. *Nature Physics* **20**, 1476–1481. ISSN: 1745-2481 (Sept. 2024).
99. Rende, R., Viteritti, L. L., Bardone, L., Becca, F. & Goldt, S. *A Simple Linear Algebra Identity to Optimize Large-Scale Neural Network Quantum States* Oct. 2023. arXiv: 2310.05715 [cond-mat].

100. Goldshlager, G., Abrahamsen, N. & Lin, L. *A Kaczmarz-inspired Approach to Accelerate the Optimization of Neural Network Wavefunctions* Jan. 2024. arXiv: 2401.10190 [physics].
101. Drissi, M., Keeble, J. W. T., Rozalén Sarmiento, J. & Rios, A. Second-Order Optimization Strategies for Neural Network Quantum States. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **382**, 20240057 (June 2024).
102. Li, X., Fan, C., Ren, W. & Chen, J. Fermionic Neural Network with Effective Core Potential. *Physical Review Research* **4**, 013021 (Jan. 2022).
103. Wilson, M., Gao, N., Wudarski, F., Rieffel, E. & Tubman, N. M. *Simulations of State-of-the-Art Fermionic Neural Network Wave Functions with Diffusion Monte Carlo* Mar. 2021. arXiv: 2103.12570 [physics, physics:quant-ph].
104. Ren, W., Fu, W., Wu, X. & Chen, J. Towards the Ground State of Molecules via Diffusion Monte Carlo on Neural Networks. *Nature Communications* **14**, 1860. ISSN: 2041-1723 (Apr. 2023).
105. Cassella, G. *et al.* Discovering Quantum Phase Transitions with Fermionic Neural Networks. *Physical Review Letters* **130**, 036401 (Jan. 2023).
106. Wilson, M. *et al.* Neural Network Ansatz for Periodic Wave Functions and the Homogeneous Electron Gas. *Physical Review B* **107**, 235139 (June 2023).
107. Li, X., Qian, Y. & Chen, J. Electric Polarization from a Many-Body Neural Network Ansatz. *Physical Review Letters* **132**, 176401 (Apr. 2024).
108. Qian, Y., Fu, W., Ren, W. & Chen, J. Interatomic Force from Neural Network Based Variational Quantum Monte Carlo. *The Journal of Chemical Physics* **157**, 164104. ISSN: 0021-9606 (Oct. 2022).
109. Qian, Y., Li, X. & Chen, J. Force and Stress Calculation with Neural Network Wavefunction for Solids. *Faraday Discussions* (2024).
110. Pescia, G., Nys, J., Kim, J., Lovato, A. & Carleo, G. Message-Passing Neural Quantum States for the Homogeneous Electron Gas. *Physical Review B* **110**, 035108 (July 2024).
111. Pescia, G., Han, J., Lovato, A., Lu, J. & Carleo, G. Neural-Network Quantum States for Periodic Systems in Continuous Space. *Physical Review Research* **4**, 023138 (May 2022).
112. Zaheer, M. *et al.* *Deep Sets in Advances in Neural Information Processing Systems* **30** (Curran Associates, Inc., 2017).
113. Lou, W. T. *et al.* Neural Wave Functions for Superfluids. *Physical Review X* **14**, 021030 (May 2024).

114. Entwistle, M. T., Schätzle, Z., Erdman, P. A., Hermann, J. & Noé, F. Electronic Excited States in Deep Variational Monte Carlo. *Nature Communications* **14**, 274. ISSN: 2041-1723 (Jan. 2023).
115. Pfau, D., Axelrod, S., Sutterud, H., von Glehn, I. & Spencer, J. S. Accurate Computation of Quantum Excited States with Neural Networks. *Science* **385**, eadn0137 (Aug. 2024).
116. Xie, H., Null, L. Z. & Wang, L. Ab-Initio Study of Interacting Fermions at Finite Temperature with Neural Canonical Transformation. *Journal of Machine Learning* **1**, 38–59. ISSN: 2790-203X, 2790-2048 (June 2022).
117. Xie, H., Zhang, L. & Wang, L. M\* of Two-Dimensional Electron Gas: A Neural Canonical Transformation Study. *SciPost Physics* **14**, 154. ISSN: 2542-4653 (June 2023).
118. Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S. & Lakshminarayanan, B. Normalizing Flows for Probabilistic Modeling and Inference. *The Journal of Machine Learning Research* **22**, 57:2617–57:2680. ISSN: 1532-4435 (Jan. 2021).
119. Xie, H., Li, Z.-H., Wang, H., Zhang, L. & Wang, L. Deep Variational Free Energy Approach to Dense Hydrogen. *Physical Review Letters* **131**, 126501 (Sept. 2023).
120. Gao, N. & Günnemann, S. *Ab-Initio Potential Energy Surfaces by Pairing GNNs with Neural Wave Functions* in *International Conference on Learning Representations* (Oct. 2021).
121. Gao, N. & Günnemann, S. *Sampling-Free Inference for Ab-Initio Potential Energy Surface Networks* in *The Eleventh International Conference on Learning Representations* (Sept. 2022).
122. Gao, N. & Günnemann, S. *Generalizing Neural Wave Functions* in *Proceedings of the 40th International Conference on Machine Learning* (PMLR, July 2023), 10708–10726.
123. Scherbela, M., Reisenhofer, R., Gerard, L., Marquetand, P. & Grohs, P. Solving the Electronic Schrödinger Equation for Multiple Nuclear Geometries with Weight-Sharing Deep Neural Networks. *Nature Computational Science* **2**, 331–341. ISSN: 2662-8457 (May 2022).
124. Scherbela, M., Gerard, L. & Grohs, P. Towards a Transferable Fermionic Neural Wavefunction for Molecules. *Nature Communications* **15**, 120. ISSN: 2041-1723 (Jan. 2024).

125. Wigner, E. On the Interaction of Electrons in Metals. *Physical Review* **46**, 1002–1011 (Dec. 1934).
126. Wigner, E. Effects of the Electron Interaction on the Energy Levels of Electrons in Metals. *Transactions of the Faraday Society* **34**, 678–685. ISSN: 0014-7672 (Jan. 1938).
127. Smoleński, T. *et al.* Signatures of Wigner Crystal of Electrons in a Monolayer Semiconductor. *Nature* **595**, 53–57. ISSN: 1476-4687 (July 2021).
128. Gell-Mann, M. & Brueckner, K. A. Correlation Energy of an Electron Gas at High Density. *Physical Review* **106**, 364–368 (Apr. 1957).
129. Carr, W. J. Energy, Specific Heat, and Magnetic Properties of the Low-Density Electron Gas. *Physical Review* **122**, 1437–1446 (June 1961).
130. Perdew, J. P. & Zunger, A. Self-Interaction Correction to Density-Functional Approximations for Many-Electron Systems. *Physical Review B* **23**, 5048–5079 (May 1981).
131. Shepherd, J. J., Booth, G., Grüneis, A. & Alavi, A. Full Configuration Interaction Perspective on the Homogeneous Electron Gas. *Physical Review B* **85**, 081103 (Feb. 2012).
132. Ewald, P. P. Die Berechnung Optischer Und Elektrostatischer Gitterpotentiale. *Annalen der Physik* **369**, 253–287. ISSN: 1521-3889 (1921).
133. Tosi, M. P. in *Solid State Physics* (eds Seitz, F. & Turnbull, D.) 1–120 (Academic Press, Jan. 1964).
134. Needs, R. J., Towler, M. D., Drummond, N. D. & Ríos, P. L. *CASINO Manual* Sept. 2019.
135. Saunders, V., Freyria-Fava, C., Dovesi, R., Salasco, L. & Roetti, C. On the Electrostatic Potential in Crystalline Systems Where the Charge Density Is Expanded in Gaussian Functions. *Molecular Physics* **77**, 629–665. ISSN: 0026-8976 (Nov. 1992).
136. Lin, C., Zong, F. H. & Ceperley, D. M. Twist-Averaged Boundary Conditions in Continuum Quantum Monte Carlo Algorithms. *Physical Review E* **64**, 016702 (June 2001).
137. Azadi, S. & Drummond, N. D. Low-Density Phase Diagram of the Three-Dimensional Electron Gas. *Physical Review B* **105**, 245135 (June 2022).
138. Neufeld, V. A. & Thom, A. J. W. A Study of the Dense Uniform Electron Gas with High Orders of Coupled Cluster. *The Journal of Chemical Physics* **147**, 194105. ISSN: 0021-9606 (Nov. 2017).

139. Drummond, N. D., Radnai, Z., Trail, J. R., Towler, M. D. & Needs, R. J. Diffusion Quantum Monte Carlo Study of Three-Dimensional Wigner Crystals. *Physical Review B* **69**, 085116. ISSN: 1098-0121, 1550-235X (Feb. 2004).
140. Bloch, F. Bemerkung zur Elektronentheorie des Ferromagnetismus und der elektrischen Leitfähigkeit. *Zeitschrift für Physik* **57**, 545–555. ISSN: 0044-3328 (July 1929).
141. Misawa, S. Ferromagnetism of an Electron Gas. *Physical Review* **140**, A1645–A1648 (Nov. 1965).
142. Edwards, S. F. & Hillel, A. J. Crystallization and Magnetic Ordering of an Electron Gas. *Journal of Physics C: Solid State Physics* **1**, 61. ISSN: 0022-3719 (Feb. 1968).
143. Care, C. & March, N. Electron Crystallization. *Advances in Physics* **24**, 101–116. ISSN: 0001-8732 (Jan. 1975).
144. Ceperley, D. Ground State of the Fermion One-Component Plasma: A Monte Carlo Study in Two and Three Dimensions. *Physical Review B* **18**, 3126–3138 (Oct. 1978).
145. Alder, B. J., Ceperley, D. M. & Pollock, E. L. Computer Simulation of Phase Transitions in Classical and Quantum Systems. *International Journal of Quantum Chemistry* **22**, 49–61. ISSN: 1097-461X (1982).
146. Ortiz, G., Harris, M. & Ballone, P. Zero Temperature Phases of the Electron Gas. *Physical Review Letters* **82**, 5317–5320 (June 1999).
147. Zong, F. H., Lin, C. & Ceperley, D. M. Spin Polarization of the Low-Density Three-Dimensional Electron Gas. *Physical Review E* **66**, 036703 (Sept. 2002).
148. Drummond, N. D. & Needs, R. J. Phase Diagram of the Low-Density Two-Dimensional Homogeneous Electron Gas. *Physical Review Letters* **102**, 126402 (Mar. 2009).
149. Azadi, S., Drummond, N. D. & Vinko, S. M. *Quantum Monte Carlo Study of the Phase Diagram of the Two-Dimensional Uniform Electron Liquid* May 2024. arXiv: 2405.00425 [cond-mat, physics:physics, physics:quant-ph].
150. Ghosh, A., Ford, C. J. B., Pepper, M., Beere, H. E. & Ritchie, D. A. Possible Evidence of a Spontaneous Spin Polarization in Mesoscopic Two-Dimensional Electron Systems. *Physical Review Letters* **92**, 116601 (Mar. 2004).
151. Falson, J. *et al.* Competing Correlated States around the Zero-Field Wigner Crystallization Transition of Electrons in Two Dimensions. *Nature Materials* **21**, 311–316. ISSN: 1476-4660 (Mar. 2022).

152. Smith, C. *et al.* *Ground State Phases of the Two-Dimension Electron Gas with a Unified Variational Approach* May 2024. arXiv: 2405.19397 [cond-mat, physics:physics, physics:quant-ph].
153. Sung, J. *et al.* *Observation of an Electronic Microemulsion Phase Emerging from a Quantum Crystal-to-Liquid Transition* Dec. 2023. arXiv: 2311.18069 [cond-mat].
154. Dirac, P. a. M. On the Annihilation of Electrons and Protons. *Mathematical Proceedings of the Cambridge Philosophical Society* **26**, 361–375. ISSN: 1469-8064, 0305-0041 (July 1930).
155. Anderson, C. D. The Apparent Existence of Easily Deflectable Positives. *Science* **76**, 238–239 (Sept. 1932).
156. Shukla, A. K. & Kumar, U. Positron Emission Tomography: An Overview. *Journal of Medical Physics / Association of Medical Physicists of India* **31**, 13–21. ISSN: 0971-6203 (2006).
157. Prantzos, N. *et al.* The 511 keV Emission from Positron Annihilation in the Galaxy. *Reviews of Modern Physics* **83**, 1001–1056 (Sept. 2011).
158. Tuomisto, F. & Makkonen, I. Defect Identification in Semiconductors with Positron Annihilation: Experiment and Theory. *Reviews of Modern Physics* **85**, 1583–1631 (Nov. 2013).
159. Hugenschmidt, C. Positrons in Surface Physics. *Surface Science Reports* **71**, 547–594. ISSN: 0167-5729 (Dec. 2016).
160. Ruark, A. E. Positronium. *Physical Review* **68**, 278–278 (Dec. 1945).
161. Wheeler, J. A. Polyelectrons. *Annals of the New York Academy of Sciences* **48**, 219–238. ISSN: 1749-6632 (1946).
162. Cassidy, D. B. & Mills, A. P. The Production of Molecular Positronium. *Nature* **449**, 195–197. ISSN: 1476-4687 (Sept. 2007).
163. Crawford, O. H. Negative Ions of Polar Molecules. *Molecular Physics* **20**, 585–591. ISSN: 0026-8976 (Jan. 1971).
164. Ryzhikh, G. G. & Mitroy, J. Positronic Lithium, an Electronically Stable Li-e<sup>+</sup> Ground State. *Physical Review Letters* **79**, 4124–4126 (Nov. 1997).
165. Gilbert, S. J., Barnes, L. D., Sullivan, J. P. & Surko, C. M. Vibrational-Resonance Enhancement of Positron Annihilation in Molecules. *Physical Review Letters* **88**, 043201. ISSN: 0031-9007, 1079-7114 (Jan. 2002).
166. Barnes, L. D., Gilbert, S. J. & Surko, C. M. Energy-Resolved Positron Annihilation for Molecules. *Physical Review A* **67**, 032706 (Mar. 2003).

167. Barnes, L. D., Young, J. A. & Surko, C. M. Energy-Resolved Positron Annihilation Rates for Molecules. *Physical Review A* **74**, 012706 (July 2006).
168. Young, J. A. & Surko, C. M. Feshbach-Resonance-Mediated Annihilation in Positron Interactions with Large Molecules. *Physical Review A* **77**, 052704 (May 2008).
169. Young, J. A. & Surko, C. M. Feshbach-Resonance-Mediated Positron Annihilation in Small Molecules. *Physical Review A* **78**, 032702 (Sept. 2008).
170. Danielson, J. R., Young, J. A. & Surko, C. M. Dependence of Positron-Molecule Binding Energies on Molecular Properties. *Journal of Physics B: Atomic, Molecular and Optical Physics* **42**, 235203. ISSN: 0953-4075 (Nov. 2009).
171. Danielson, J. R., Gosselin, J. J. & Surko, C. M. Dipole Enhancement of Positron Binding to Molecules. *Physical Review Letters* **104**, 233201 (June 2010).
172. Danielson, J. R., Jones, A. C. L., Natisin, M. R. & Surko, C. M. Comparisons of Positron and Electron Binding to Molecules. *Physical Review Letters* **109**, 113201 (Sept. 2012).
173. Mitroy, J. *et al.* Theory and Application of Explicitly Correlated Gaussians. *Reviews of Modern Physics* **85**, 693–749 (May 2013).
174. Strasburger, K. Adiabatic Positron Affinity of LiH. *The Journal of Chemical Physics* **114**, 615–616. ISSN: 0021-9606 (Jan. 2001).
175. Mitroy, J. & Ryzhikh, G. Positron Binding to Atomic Zinc. *Journal of Physics B: Atomic, Molecular and Optical Physics* **32**, 1375. ISSN: 0953-4075 (Mar. 1999).
176. Bromley, M. W. J. & Mitroy, J. Configuration-Interaction Calculations of Positron Binding to Zinc and Cadmium. *Physical Review A* **65**, 062506 (June 2002).
177. Bromley, M. W. J. & Mitroy, J. Configuration-Interaction Calculations of PsH and E<sup>+</sup>-Be. *Physical Review A* **65**, 012505 (Dec. 2001).
178. Bromley, M. W. J. & Mitroy, J. Large-Dimension Configuration-Interaction Calculations of Positron Binding to the Group-II Atoms. *Physical Review A* **73**, 032507 (Mar. 2006).
179. Bromley, M. W. J. & Mitroy, J. Positron Binding to Calcium, a Configuration Interaction Study. *Journal of Physics B: Atomic, Molecular and Optical Physics* **33**, L325. ISSN: 0953-4075 (May 2000).

180. Mitroy, J., Bromley, M. W. J. & Ryzhikh, G. G. Positron and Positronium Binding to Atoms. *Journal of Physics B: Atomic, Molecular and Optical Physics* **35**, R81. ISSN: 0953-4075 (June 2002).
181. Strasburger, K. Quantum Chemical Study on Complexes of the LiH Molecule with  $E^+$ , Ps and Ps- Including Correlation Energy. *Chemical Physics Letters* **253**, 49–52. ISSN: 0009-2614 (Apr. 1996).
182. Buenker, R. J. *et al.* Positron Binding Energies for Alkali Hydrides. *The Journal of Physical Chemistry A* **109**, 5956–5964. ISSN: 1089-5639 (July 2005).
183. Buenker, R. J., Liebermann, H.-P., Pichl, L., Tachikawa, M. & Kimura, M. Role of the Electric Dipole Moment in Positron Binding to the Ground and Excited States of the BeO Molecule. *The Journal of Chemical Physics* **126**, 104305. ISSN: 0021-9606 (Mar. 2007).
184. Bromley, M. W. J. *Positron-Atom Interactions Studied Using Configuration Interaction Methods* PhD thesis (CDU, Sept. 2002).
185. Bressanini, D., Mella, M. & Morosi, G. Positron Chemistry by Quantum Monte Carlo. II. Ground-state of Positron-Polar Molecule Complexes. *The Journal of Chemical Physics* **109**, 1716–1720. ISSN: 0021-9606 (Aug. 1998).
186. Mella, M., Morosi, G. & Bressanini, D. Positron and Positronium Chemistry by Quantum Monte Carlo. IV. Can This Method Accurately Compute Observables beyond Energy? *The Journal of Chemical Physics* **111**, 108–114. ISSN: 0021-9606 (July 1999).
187. Mella, M., Chiesa, S. & Morosi, G. Annihilation Rate in Positronic Systems by Quantum Monte Carlo:  $e^+LiH$  as Test Case. *The Journal of Chemical Physics* **116**, 2852–2862. ISSN: 0021-9606 (Feb. 2002).
188. Kita, Y., Maezono, R., Tachikawa, M., Towler, M. & Needs, R. J. Ab Initio Quantum Monte Carlo Study of the Positronic Hydrogen Cyanide Molecule. *The Journal of Chemical Physics* **131**, 134310. ISSN: 0021-9606 (Oct. 2009).
189. Kita, Y., Maezono, R., Tachikawa, M., Towler, M. D. & Needs, R. J. Ab Initio Quantum Monte Carlo Study of the Binding of a Positron to Alkali-Metal Hydrides. *The Journal of Chemical Physics* **135**, 054108. ISSN: 0021-9606 (Aug. 2011).
190. Ito, S., Yoshida, D., Kita, Y. & Tachikawa, M. First-Principles Quantum Monte Carlo Studies for Prediction of Double Minima for Positronic Hydrogen Molecular Dianion. *The Journal of Chemical Physics* **153**, 224305. ISSN: 0021-9606 (Dec. 2020).

191. Martinez, J. A. C., Barborini, M. & Tkatchenko, A. Correlated Wave Functions for Electron–Positron Interactions in Atoms and Molecules. *Journal of Chemical Theory and Computation* (Mar. 2022).
192. Jiang, N. & Schrader, D. M. Diffusion Quantum Monte Carlo Calculation of the Binding Energy and Annihilation Rate of Positronium Hydride, PsH. *The Journal of Chemical Physics* **109**, 9430–9433. ISSN: 0021-9606 (Dec. 1998).
193. Gilbert, S. J., Barnes, L. D., Sullivan, J. P. & Surko, C. M. Vibrational-Resonance Enhancement of Positron Annihilation in Molecules. *Physical Review Letters* **88**, 043201 (Jan. 2002).
194. Barnes, L. D., Gilbert, S. J. & Surko, C. M. Energy-Resolved Positron Annihilation for Molecules. *Physical Review A* **67**, 032706 (Mar. 2003).
195. Barnes, L. D., Young, J. A. & Surko, C. M. Energy-Resolved Positron Annihilation Rates for Molecules. *Physical Review A* **74**, 012706 (July 2006).
196. Young, J. A. & Surko, C. M. Feshbach-Resonance-Mediated Annihilation in Positron Interactions with Large Molecules. *Physical Review A* **77**, 052704 (May 2008).
197. Young, J. A. & Surko, C. M. Feshbach-Resonance-Mediated Positron Annihilation in Small Molecules. *Physical Review A* **78**, 032702 (Sept. 2008).
198. Danielson, J. R., Young, J. A. & Surko, C. M. Dependence of Positron–Molecule Binding Energies on Molecular Properties. *Journal of Physics B: Atomic, Molecular and Optical Physics* **42**, 235203. ISSN: 0953-4075 (Nov. 2009).
199. Danielson, J. R., Gosselin, J. J. & Surko, C. M. Dipole Enhancement of Positron Binding to Molecules. *Physical Review Letters* **104**, 233201 (June 2010).
200. Danielson, J. R., Jones, A. C. L., Natisin, M. R. & Surko, C. M. Comparisons of Positron and Electron Binding to Molecules. *Physical Review Letters* **109**, 113201 (Sept. 2012).
201. Tachikawa, M., Kita, Y. & J. Buenker, R. Bound States of the Positron with Nitrile Species with a Configuration Interaction Multi-Component Molecular Orbital Approach. *Physical Chemistry Chemical Physics* **13**, 2701–2705 (2011).
202. Tachikawa, M., Kita, Y. & Buenker, R. J. Bound States of Positron with Simple Carbonyl and Aldehyde Species with Configuration Interaction Multi-Component Molecular Orbital and Local Vibrational Approaches. *New Journal of Physics* **14**, 035004. ISSN: 1367-2630 (Mar. 2012).

203. Koyanagi, K., Takeda, Y., Oyamada, T., Kita, Y. & Tachikawa, M. Positron-Attachment to Nonpolar or Small Dipole CXY (X, Y = O, S, and Se) Molecules: Vibrational Enhancement of Positron Affinities with Configuration Interaction Level of Multi-Component Molecular Orbital Approach. *Physical Chemistry Chemical Physics* **15**, 16208–16213 (2013).
204. Tachikawa, M. Positron-Attachment to Acetonitrile, Acetaldehyde, and Acetone Molecules: Vibrational Enhancement of Positron Affinities with Configuration Interaction Level of Multi-Component Molecular Orbital Approach. *Journal of Physics: Conference Series* **488**, 012053. ISSN: 1742-6596 (Apr. 2014).
205. Hofierka, J., Cunningham, B., Rawlins, C. M., Patterson, C. H. & Green, D. G. Many-Body Theory of Positron Binding to Polyatomic Molecules. *Nature* **606**, 688–693. ISSN: 1476-4687 (June 2022).
206. Cassidy, J. P. *et al.* Many-Body Theory Calculations of Positron Binding to Halogenated Hydrocarbons. *Physical Review A* **109**, L040801 (Apr. 2024).
207. Arthur-Baidoo, E. *et al.* Positron Annihilation and Binding in Aromatic and Other Ring Molecules. *Physical Review A (Atomic, Molecular, and Optical Physics)* **109**. ISSN: 1050-2947 (June 2024).
208. Frolov, A. M. & Smith, V. H. Positronium Hydrides and the Ps<sub>2</sub> Molecule: Bound-state Properties, Positron Annihilation Rates, and Hyperfine Structure. *Physical Review A* **55**, 2662–2673 (Apr. 1997).
209. Cassella, G., Foulkes, W. M. C., Pfau, D. & Spencer, J. S. Neural Network Variational Monte Carlo for Positronic Chemistry. *Nature Communications* **15**, 5214. ISSN: 2041-1723 (June 2024).
210. Simula, K. A., Muff, J. E., Makkonen, I. & Drummond, N. D. Quantum Monte Carlo Study of Positron Lifetimes in Solids. *Physical Review Letters* **129**, 166403 (Oct. 2022).
211. Buenker, R. J. & Liebermann, H.-P. Configuration Interaction Calculations of Annihilation Rates for Positronic Complexes of Alkali Hydrides. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **267**, 763–767. ISSN: 0168-583X (Mar. 2009).
212. Usukura, J., Varga, K. & Suzuki, Y. Signature of the Existence of the Positronium Molecule. *Physical Review A* **58**, 1918–1931 (Sept. 1998).
213. Ryzhikh, G. G. & Mitroy, J. Positron Annihilation Profiles for HPs and He(3Se)E<sup>+</sup>. *Journal of Physics B: Atomic, Molecular and Optical Physics* **32**, 4051. ISSN: 0953-4075 (Aug. 1999).

214. Ryzhikh, G. G., Mitroy, J. & Varga, K. The Structure of Exotic Atoms Containing Positrons and Positronium. *Journal of Physics B: Atomic, Molecular and Optical Physics* **31**, 3965. ISSN: 0953-4075 (Sept. 1998).
215. Mitroy, J. & Ryzhikh, G. G. Improved Binding Energies for LiPs, e+Be, NaPs and e+Mg. *Journal of Physics B: Atomic, Molecular and Optical Physics* **34**, 2001. ISSN: 0953-4075 (May 2001).
216. Bromley, M. W. J. & Mitroy, J. Large-Dimension Configuration-Interaction Calculations of Positron Binding to the Group-II Atoms. *Physical Review A* **73**, 032507 (Mar. 2006).
217. Azadi, S., Monserrat, B., Foulkes, W. M. C. & Needs, R. J. Dissociation of High-Pressure Solid Molecular Hydrogen: A Quantum Monte Carlo and Anharmonic Vibrational Study. *Physical Review Letters* **112**, 165501 (Apr. 2014).
218. Drummond, N. D. *et al.* Quantum Monte Carlo Study of the Phase Diagram of Solid Molecular Hydrogen at Extreme Pressures. *Nature Communications* **6**, 7794. ISSN: 2041-1723 (July 2015).
219. Monacelli, L., Casula, M., Nakano, K., Sorella, S. & Mauri, F. Quantum Phase Diagram of High-Pressure Hydrogen. *Nature Physics* **19**, 845–850. ISSN: 1745-2481 (June 2023).
220. Ashcroft, N. W. Metallic Hydrogen: A High-Temperature Superconductor? *Physical Review Letters* **21**, 1748–1749 (Dec. 1968).
221. Drozdov, A. P., Erements, M. I., Troyan, I. A., Ksenofontov, V. & Shylin, S. I. Conventional Superconductivity at 203 Kelvin at High Pressures in the Sulfur Hydride System. *Nature* **525**, 73–76. ISSN: 1476-4687 (Sept. 2015).
222. Liu, H., Naumov, I. I., Hoffmann, R., Ashcroft, N. W. & Hemley, R. J. Potential High-Tc Superconducting Lanthanum and Yttrium Hydrides at High Pressure. *Proceedings of the National Academy of Sciences* **114**, 6990–6995 (July 2017).
223. Somayazulu, M. *et al.* Evidence for Superconductivity above 260 K in Lanthanum Superhydride at Megabar Pressures. *Physical Review Letters* **122**, 027001 (Jan. 2019).
224. Steinfeld, J. I. *Molecules and Radiation: An Introduction to Modern Molecular Spectroscopy* ISBN: 978-0-06-046417-2 (Harper & Row, 1974).
225. Bowman, J. M. The Self-Consistent-Field Approach to Polyatomic Vibrations. *Accounts of Chemical Research*, 202, 208 (June 1986).

226. Markland, T. E. & Ceriotti, M. Nuclear Quantum Effects Enter the Mainstream. *Nature Reviews Chemistry* **2**, 1–14. ISSN: 2397-3358 (Feb. 2018).
227. Bardeen, J. & Pines, D. Electron-Phonon Interaction in Metals. *Physical Review* **99**, 1140–1150 (Aug. 1955).
228. Grimvall, G. The Electron-Phonon Interaction in Normal Metals. *Physica Scripta* **14**, 63. ISSN: 1402-4896 (July 1976).
229. Giustino, F. Electron-Phonon Interactions from First Principles. *Reviews of Modern Physics* **89**, 015003 (Feb. 2017).
230. Debernardi, A., Baroni, S. & Molinari, E. Anharmonic Phonon Lifetimes in Semiconductors from Density-Functional Perturbation Theory. *Physical Review Letters* **75**, 1819–1822 (Aug. 1995).
231. Monserrat, B., Drummond, N. D. & Needs, R. J. Anharmonic Vibrational Properties in Periodic Systems: Energy, Electron-Phonon Coupling, and Stress. *Physical Review B* **87**, 144302 (Apr. 2013).
232. Errea, I., Rousseau, B. & Bergara, A. Anharmonic Stabilization of the High-Pressure Simple Cubic Phase of Calcium. *Physical Review Letters* **106**, 165501 (Apr. 2011).
233. Calandra, M., Profeta, G. & Mauri, F. Adiabatic and Nonadiabatic Phonon Dispersion in a Wannier Function Approach. *Physical Review B* **82**, 165111 (Oct. 2010).
234. Errea, I., Calandra, M. & Mauri, F. Anharmonic Free Energies and Phonon Dispersions from the Stochastic Self-Consistent Harmonic Approximation: Application to Platinum and Palladium Hydrides. *Physical Review B* **89**, 064302 (Feb. 2014).
235. Monserrat, B. *et al.* Structure and Metallicity of Phase V of Hydrogen. *Physical Review Letters* **120**, 255701 (June 2018).
236. Monacelli, L. *et al.* The Stochastic Self-Consistent Harmonic Approximation: Calculating Vibrational Properties of Materials with Full Quantum and Anharmonic Effects. *Journal of Physics: Condensed Matter* **33**, 363001. ISSN: 0953-8984 (July 2021).
237. Michl, J. *Physical Basis of Qualitative MO Arguments in Organic Photochemistry* in *Photochemistry* (Springer, Berlin, Heidelberg, 1974), 1–59. ISBN: 978-3-540-37908-9.
238. Yarkony, D. R. Nonadiabatic Quantum Chemistry—Past, Present, and Future. *Chemical Reviews* **112**, 481–498. ISSN: 0009-2665 (Jan. 2012).

239. Hammes-Schiffer, S. & Stuchebrukhov, A. A. Theory of Coupled Electron and Proton Transfer Reactions. *Chemical reviews* **110**, 6939–6960. ISSN: 0009-2665 (Dec. 2010).
240. Nelson, T. R. *et al.* Non-Adiabatic Excited-State Molecular Dynamics: Theory and Applications for Modeling Photophysics in Extended Molecular Materials. *Chemical Reviews* **120**, 2215–2287. ISSN: 0009-2665 (Feb. 2020).
241. Pavošević, F., Culpitt, T. & Hammes-Schiffer, S. Multicomponent Quantum Chemistry: Integrating Electronic and Nuclear Quantum Effects via the Nuclear–Electronic Orbital Method. *Chemical Reviews* **120**, 4222–4253. ISSN: 0009-2665 (May 2020).
242. López Ríos, P., Seth, P., Drummond, N. D. & Needs, R. J. Framework for Constructing Generic Jastrow Correlation Factors. *Physical Review E* **86**, 036703 (Sept. 2012).
243. Herzberg, G. *Molecular Spectra and Molecular Structure. Vol.1: Spectra of Diatomic Molecules* (Jan. 1950).
244. Bubin, S., Leonarski, F., Stanke, M. & Adamowicz, L. Non-Adiabatic Corrections to the Energies of the Pure Vibrational States of H<sub>2</sub>. *Chemical Physics Letters* **477**, 12–16. ISSN: 0009-2614 (July 2009).
245. Fadon, A. P., Cassella, G., Sutterud, H. & Foulkes, W. M. C. *Interaction-Induced Symmetry Breaking in Circular Quantum Dots* Oct. 2024. arXiv: 2410.01652 [cond-mat].
246. Cook, R. L., De Lucia, F. C. & Helminger, P. Molecular Force Field and Structure of Water: Recent Microwave Results. *Journal of Molecular Spectroscopy* **53**, 62–76. ISSN: 0022-2852 (Oct. 1974).
247. Helminger, P., De Lucia, F. C. & Gordy, W. Rotational Spectra of NH<sub>3</sub> and ND<sub>3</sub> in the 0.5-Mm Wavelength Region. *Journal of Molecular Spectroscopy* **39**, 94–97. ISSN: 0022-2852 (July 1971).
248. Hirota, E. Anharmonic Potential Function and Equilibrium Structure of Methane. *Journal of Molecular Spectroscopy* **77**, 213–221. ISSN: 0022-2852 (Aug. 1979).
249. Sverdlov, L. M. *Vibrational spectra of polyatomic molecules* ISBN: 978-0-470-83845-7 (Wiley, New York, 1973).
250. Nielsen, O. H. & Martin, R. M. Quantum-Mechanical Theory of Stress and Force. *Physical Review B* **32**, 3780–3791 (Sept. 1985).
251. Qian, Y., Li, X. & Chen, J. Force and Stress Calculations with a Neural-Network Wave Function for Solids. *Faraday Discussions*. ISSN: 1364-5498 (Apr. 2024).

- 252. Knuth, F., Carbogno, C., Atalla, V., Blum, V. & Scheffler, M. All-Electron Formalism for Total Energy Strain Derivatives and Stress Tensor Components for Numeric Atom-Centered Orbitals. *Computer Physics Communications* **190**, 33–50. ISSN: 0010-4655 (May 2015).
- 253. Hine, N. D. M. & Foulkes, W. M. C. Localization Lengths over Metal to Band Insulator Transitions. *Journal of Physics: Condensed Matter* **19**, 506212. ISSN: 0953-8984 (Nov. 2007).
- 254. Sakurai, J. J. & Napolitano, J. *Modern Quantum Mechanics* ISBN: 978-1-108-42241-3 (Cambridge University Press, Sept. 2017).

# Appendix A

## Notational conventions

Vector quantities are labelled by bolded lower-case characters, *e.g.*  $\mathbf{r}$ , and matrix quantities are labelled by bolder upper-case characters, *e.g.*  $\mathbf{W}$ . There is one exception to this convention which we frequently employ: we *always* reserve  $\mathbf{R}_i$  to indicate the co-ordinates of an atomic nucleus indexed by  $i$ . Operators are indicated with either a caligraphic font, *e.g.*  $\mathcal{H}$ , or with a ‘hat’, *e.g.*  $\hat{T}$ . We do not use any notation to distinguish between operators acting in a particular basis.

We frequently discuss many-body wavefunctions which are functions of the spin and spatial co-ordinates of  $N$  quantum particles,

$$\Psi(\mathbf{r}_1, \sigma_1, \mathbf{r}_2, \sigma_2, \dots, \mathbf{r}_N, \sigma_N).$$

To avoid explicitly indexing all  $N$  particles everywhere a many-body quantity appears, we use the bold and italicized shorthands  $\mathbf{r}$ ,  $\boldsymbol{\sigma}$  with no index to indicate the  $N$ -tuple of electron co-ordinates, i.e.

$$\begin{aligned}\mathbf{r} &\equiv (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N), \\ \boldsymbol{\sigma} &\equiv (\sigma_1, \sigma_2, \dots, \sigma_N).\end{aligned}$$

These are sometimes expressed as a single, combined co-ordinate

$$\mathbf{x} \equiv (\mathbf{r}, \boldsymbol{\sigma}) \quad \text{such that} \quad \mathbf{x}_i \equiv (\mathbf{r}_i, \sigma_i).$$

The collective co-ordinates of atomic nuclei are likewise indicated by a bold and italicized shorthand  $\mathbf{R}$ .

The gradient and Laplacian operators are indicated throughout by  $\nabla$  and  $\nabla^2$ . Differential operators carrying a subscripted index, *e.g.*  $\nabla_i^2$ , are understood to act on the 3-dimensional subspace corresponding to the co-ordinate  $\mathbf{r}_i$ . We adopt a special case of this notation in chapter 6, where vector subscripts are used to distinguish differential operators acting on nuclear co-ordinates, *e.g.*  $\nabla_{\mathbf{R}_i}^2$ .

Expectation values of local quantities are indicated using angular brackets, subscripted with the probability distribution the expectation value is taken with respect to, *e.g.*  $\langle E_L \rangle_{p_\theta}$ . We occasionally omit the subscripted  $p_\theta$  in contexts where it is clear that the expectation value is being taken with respect to the probability distribution defined by a trial wavefunction.

Matrix determinants are indicated with the notation  $\det \mathbf{A}$ , or,

$$\det[a_{ij}].$$

Here, we have written dummy indices  $i$  and  $j$ . In general, this determinant notation will always contain two dummy indices within the square brackets that are not the indices of a summation or product, and do not appear on the other side of an equation. These dummy indices label the rows and columns of the matrix. Note that  $\det \mathbf{A} = \det \mathbf{A}^T$ , such that we do not need to specify which dummy index labels rows, and which index labels columns.

## Appendix B

# Monte Carlo integration and the Metropolis algorithm

Evaluating an integral given access only to point evaluations of the integrand requires a numerical quadrature scheme. Concretely, consider an integral

$$I = \int_{\Omega} d\mathbf{x} f(\mathbf{x}), \quad (\text{B.1})$$

where  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $\Omega \subseteq \mathbb{R}^n$ .

Newton-Cotes formulas are the most straightforward set of numerical quadrature rules. Here, the integral is evaluated using point estimates of the integrand at regularly spaced points. These point estimates are used to construct polynomial interpolants whose integrals are tractable, and the full integral is evaluated piecewise. Unfortunately, Newton-Cotes formulas are rendered computationally useless as the dimensionality of the integrand,  $n$ , becomes large.

Consider applying a Newton-Cotes quadrature<sup>1</sup> with a fixed number of evaluation points in the integration domain along each dimension,  $m$ . For an  $n$  dimensional integration domain, the number of point evaluations of the integrand required to implement this scheme is  $m^n$ . In solving the TISE for a small molecule, a modest value is  $n = 30$  (10 electrons in 3 dimensional space). Thus, with as few as  $m = 10$  evaluation points along each dimension, we require  $\sim 10^{30}$  evaluations of the integrand – a clear impossibility. This failure is known as the ‘curse of dimensionality’.

Monte Carlo integration proceeds instead by sampling points randomly within the domain of the integral. Given a set of  $M$  points  $\{\mathbf{x}_i \in \Omega\}$  sampled from a

---

<sup>1</sup>Other numerical quadrature schemes, *e.g.* Gaussian quadrature, do not rely on a fixed spacing of points. However, achieving a fixed level of numerical error with these alternatives still suffers from the curse of dimensionality.

probability distribution  $p(\mathbf{x})$ , we estimate the integral as

$$I = \int_{\Omega} d\mathbf{x} g(\mathbf{x}) p(\mathbf{x}) \simeq \frac{1}{M} \sum_{i=1}^M g(\mathbf{x}_i), \quad g(\mathbf{x}) = \frac{f(\mathbf{x})}{p(\mathbf{x})}. \quad (\text{B.2})$$

This can be understood as evaluating the expectation value of the quantity  $g(\mathbf{x})$  with respect to the probability distribution,  $p(\mathbf{x})$ .

The numerical error in our Monte Carlo estimate of the integral is given by the standard error in the mean,

$$\sigma_M = \sqrt{\frac{\frac{1}{M-1} \sum_{i=1}^M (g(\mathbf{x}_i) - I)^2}{M}}. \quad (\text{B.3})$$

So long as the numerator on the right-hand side remains bounded as  $M \rightarrow \infty$  – *i.e.*, the variance of  $g(\mathbf{x})$  with respect to  $p(\mathbf{x})$  is finite – this error tends to zero as the inverse square root of the number of point estimates of the integrand. At first glance, this is not a great result: to obtain a two-fold reduction in the error requires a four-fold increase in the number of point estimates of the integrand. However, remarkably, this is *completely independent* of the dimensionality of the integral. Thus, so long as we can achieve a reasonable variance, we can obtain accurate estimates of an integral of any dimensionality.<sup>2</sup>

Calculating Monte Carlo estimates, eq. (B.2), requires a practical method for drawing samples from the probability distribution,  $p(\mathbf{x})$ . The Metropolis algorithm is the most commonly used method for generating samples from a probability distribution for use in Monte Carlo integration, and the one which is used in all calculations in this thesis. Samples are generated according to the following procedure:

1. Choose an initial ‘walker position’,  $\mathbf{x}_t$ .
2. Propose a new walker position,  $\mathbf{x}^*$  drawn from a ‘proposal distribution’,  $q(\mathbf{x}^*|\mathbf{x}_t)$ , chosen such that the transition probability is symmetric, *i.e.*,

$$q(\mathbf{x}^*|\mathbf{x}_t) = q(\mathbf{x}_t|\mathbf{x}^*).$$

3. Generate a random number,  $\alpha$ , from a uniform distribution over the range  $[0, 1]$ , then

- if  $\alpha \leq p(\mathbf{x}^*)/p(\mathbf{x}_{t-1})$ , let  $\mathbf{x}_{t+1} = \mathbf{x}^*$ .
- else if  $\alpha > p(\mathbf{x}^*)/p(\mathbf{x}_{t-1})$ , let  $\mathbf{x}_{t+1} = \mathbf{x}_t$ .

---

<sup>2</sup>The variance depends upon the choice of distribution  $p(\mathbf{x})$ . The best distribution to choose is one which closely resembles the integrand,  $f(\mathbf{x})$ . This can be seen by considering what distribution minimizes the difference between  $g(\mathbf{x})$  and its mean value over the domain.

This procedure generates a ‘Markov chain’ – a series of walker positions in which the conditional distribution of each  $\mathbf{x}_{t+1}$  depends only upon  $\mathbf{x}_t$ . As the length of the chain,  $t$ , becomes large, the marginal distribution of the  $\mathbf{x}_t$  approaches  $p(\mathbf{x})$  [27].

It is hard to estimate how many steps of the Markov chain are required before a given walker can be considered distributed according to  $p(\mathbf{x})$ . This is usually handled by adopting a ‘burn-in’ period before utilizing samples from the Markov chain: some heuristically chosen number of walker positions are discarded before beginning to accumulate Monte Carlo estimates.

The samples produced by the Metropolis algorithm with a local proposal distribution (as is often the case) are *autocorrelated*. Subsequent samples are not *i.i.d.*, precluding a straightforward calculation of the variance of an estimate under the central limit theorem. In fact, the variance of an estimate calculated with these samples will be increased, equivalent to evaluating the estimate with a lower ‘effective sample size’. The larger the degree of autocorrelation, the more samples are required to obtain an accurate Monte Carlo estimate. If the integrand function,  $f(\mathbf{x})$ , is expensive to evaluate, it is often more computationally efficient to *discard* some fraction of autocorrelated samples, and only take point estimates of the integrand for temporally well-separated samples – ideally, with a separation greater than the autocorrelation time.<sup>3</sup> This way, a Monte Carlo estimate of similar quality is obtained with less computational effort.

The autocorrelation time is also strongly dependent on the proposal distribution. A good ‘rule of thumb’ for producing samples via the Metropolis algorithm is to choose a proposal distribution which results in an  $\sim 50\%$  acceptance rate of newly proposed walker positions. Often, the proposal distribution is parameterized – for instance, a Gaussian distribution parameterized by its width. In the course of a Monte Carlo integration, the parameters of the proposal distribution can be adjusted to maintain the  $\sim 50\%$  acceptance rate.

---

<sup>3</sup>Here we are using the colloquialism that the index,  $t$ , along the Markov chain corresponds to time.

# Appendix C

## Gradient of the energy functional

Consider the partial derivative with respect to a single parameter (the gradient is obtained by stacking these partial derivatives as a column vector),

$$\partial_\theta E[\Psi_\theta] = \partial_\theta \frac{\int d\mathbf{r} \Psi_\theta^* \mathcal{H} \Psi_\theta}{\int d\mathbf{r} |\Psi_\theta|^2}. \quad (\text{C.1})$$

Applying the product rule, we obtain,

$$\begin{aligned} \partial_\theta E[\Psi_\theta] &= \frac{\int d\mathbf{r} (\partial_\theta \Psi_\theta^*) (\mathcal{H} \Psi_\theta)}{\int d\mathbf{r} |\Psi_\theta|^2} \\ &\quad + \frac{\int d\mathbf{r} (\Psi_\theta^* \mathcal{H}) (\partial_\theta \Psi_\theta)}{\int d\mathbf{r} |\Psi_\theta|^2} \\ &\quad - \left[ \int d\mathbf{r} \partial_\theta |\Psi_\theta|^2 \right] \frac{\int d\mathbf{r} \Psi_\theta^* \mathcal{H} \Psi_\theta}{\int d\mathbf{r} |\Psi_\theta|^4} \\ &= \frac{\int d\mathbf{r} (\partial_\theta \Psi_\theta^*) (\mathcal{H} \Psi_\theta)}{\int d\mathbf{r} |\Psi_\theta|^2} \\ &\quad + \frac{\int d\mathbf{r} (\Psi_\theta^* \mathcal{H}) (\partial_\theta \Psi_\theta)}{\int d\mathbf{r} |\Psi_\theta|^2} \\ &\quad - \left[ \int d\mathbf{r} (\partial_\theta \Psi_\theta^*) \Psi_\theta + \int d\mathbf{r} \Psi_\theta^* (\partial_\theta \Psi_\theta) \right] \frac{\int d\mathbf{r} \Psi_\theta^* \mathcal{H} \Psi_\theta}{\int d\mathbf{r} |\Psi_\theta|^4}, \end{aligned} \quad (\text{C.2})$$

where we have exploited the Hermitian property of the Hamiltonian to ‘act on the left’ and pull through the partial derivative. Next, we apply the same trick used to

obtain the local energy operator to each term,

$$\begin{aligned}
\partial_\theta E[\Psi_\theta] &= \frac{\int d\mathbf{r} |\Psi_\theta|^2 \frac{\partial_\theta \Psi_\theta^*}{\Psi_\theta^*} \frac{\mathcal{H} \Psi_\theta}{\Psi_\theta}}{\int d\mathbf{r} |\Psi_\theta|^2} \\
&+ \frac{\int d\mathbf{r} |\Psi_\theta|^2 \frac{\Psi_\theta^* \mathcal{H}}{\Psi_\theta^*} \frac{\partial_\theta \Psi_\theta}{\Psi_\theta}}{\int d\mathbf{r} |\Psi_\theta|^2} \\
&- \left[ \int d\mathbf{r} |\Psi_\theta|^2 \frac{\partial_\theta \Psi_\theta^*}{\Psi_\theta^*} + \int d\mathbf{r} |\Psi_\theta|^2 \frac{\partial_\theta \Psi_\theta}{\Psi_\theta} \right] \frac{\int d\mathbf{r} \Psi_\theta^* \mathcal{H} \Psi_\theta}{\int d\mathbf{r} |\Psi_\theta|^4} \\
&= \langle \partial_\theta \log \Psi_\theta^* E_L \rangle_{p_\theta} + \langle \partial_\theta \log \Psi_\theta E_L^* \rangle_{p_\theta} \\
&- \langle \partial_\theta \log \Psi_\theta^* + \partial_\theta \log \Psi_\theta \rangle_{p_\theta} \langle E_L \rangle_{p_\theta} \\
&= \langle E_L^* \partial_\theta \log \Psi_\theta + E_L \partial_\theta \log \Psi_\theta^* - 2 \langle E_L \rangle_{p_\theta} \partial_\theta \log |\Psi_\theta| \rangle_{p_\theta}.
\end{aligned} \tag{C.3}$$

Thus, we obtain eq. (3.13). In the final line, we have used the fact that for a complex number,  $z = ae^{ib}$ , we have that,

$$\partial_\theta \log z = \partial_\theta \log a + i\partial_\theta b, \tag{C.4}$$

and

$$\partial_\theta \log z^* = \partial_\theta \log a - i\partial_\theta b, \tag{C.5}$$

such that

$$\partial_\theta \log z + \partial_\theta \log z^* = 2\partial_\theta \log(a) = 2\partial_\theta \log|z|. \tag{C.6}$$

Note that the quantity inside the expectation value in eq. (C.3) is strictly real-valued. The final product term contains only real numbers, and the complex component of the first two terms cancels exactly. Writing  $\Psi_\theta = |\Psi_\theta|e^{i\phi_\theta}$ ,

$$\begin{aligned}
E_L^* \partial_\theta \log \Psi_\theta &= (\partial_\theta \log |\Psi_\theta| + i\partial_\theta \phi_\theta) (\text{Re} E_L - i\text{Im} E_L), \\
&= \partial_\theta \log |\Psi_\theta| \text{Re} E_L - i\partial_\theta \log |\Psi_\theta| \text{Im} E_L + i\partial_\theta \phi_\theta \text{Re} E_L - \partial_\theta \phi_\theta \text{Im} E_L,
\end{aligned} \tag{C.7}$$

and,

$$\begin{aligned}
E_L \partial_\theta \log \Psi_\theta^* &= (\partial_\theta \log |\Psi_\theta| - i\partial_\theta \phi_\theta) (\text{Re} E_L + i\text{Im} E_L), \\
&= \partial_\theta \log |\Psi_\theta| \text{Re} E_L + i\partial_\theta \log |\Psi_\theta| \text{Im} E_L - i\partial_\theta \phi_\theta \text{Re} E_L - \partial_\theta \phi_\theta \text{Im} E_L,
\end{aligned} \tag{C.8}$$

such that

$$E_L^* \partial_\theta \log \Psi_\theta + E_L \partial_\theta \log \Psi_\theta^* = 2(\partial_\theta \log |\Psi_\theta| \text{Re} E_L - \partial_\theta \phi_\theta \text{Im} E_L), \tag{C.9}$$

which is real-valued.

We now proceed to prove that the ground-state wavefunction of a Hamiltonian  $\mathcal{H} = \mathcal{H}^*$  can always be taken as a real-valued function. Consider,

$$(\mathcal{H}\Psi_0)^* = \mathcal{H}^*\Psi_0^* = E_0\Psi_0^*, \quad (\text{C.10})$$

where in the second equality we have evaluated the action of  $\mathcal{H}$  on  $\Psi_0$  before applying the conjugation. By  $\mathcal{H} = \mathcal{H}^*$ , the second equality implies that

$$\mathcal{H}\Psi^* = E_0\Psi_0^*, \quad (\text{C.11})$$

such that  $\Psi_0^*$  is another ground state of  $\mathcal{H}$ . If we assume the ground-state is non-degenerate, we require that  $\Psi_0^* = c\Psi_0$  for some unit complex number  $c$ , *i.e.* complex conjugation only enacts a global phase rotation of  $\Psi_0$ . This is possible if and only if the phase of  $\Psi_0$  is constant everywhere, such that a purely real-valued ground-state can always be obtained by applying a global phase.

Time-reversal symmetry is more generally understood as the symmetry of  $\mathcal{H}$  under the application of an anti-unitary time-reversal operator,  $\mathcal{T}$ , such that  $\mathcal{T}\mathcal{H} = \mathcal{H}\mathcal{T}$ . For a detailed discussion of time-reversal symmetry, and a more general proof that the ground-state wavefunction of a time-reversal symmetric Hamiltonian can be taken as a real-valued function (even in the degenerate case), we direct the interested reader to section 4.4 of ref. [254].

Thus, in obtaining the ground-state wavefunction of a time-reversal symmetric Hamiltonian, we can restrict the wavefunction ansatz,  $\Psi_\theta$ , to always be real-valued. In this case, the gradient of the energy function becomes,

$$\partial_\theta E[\Psi_\theta] = 2 \langle (E_L - \langle E_L \rangle) \partial_\theta \log |\Psi_\theta| \rangle_{\rho_\theta}, \quad (\text{C.12})$$

which is obtained from eq. (C.3) by recognizing that  $E_L$  is purely real, and that  $\partial_\theta \log \Psi_\theta = \partial_\theta \log \Psi_\theta^* = \partial_\theta \log |\Psi_\theta|$ , when  $\Psi_\theta$  is real-valued.<sup>1</sup>

---

<sup>1</sup>This might appear puzzling, due to the branch cut of the complex logarithm along the negative real number line. Note, however, that the *derivative* of the complex logarithm is everywhere well-defined on the complex plane.

## Appendix D

# Variance matching procedure for benzene

In VMC calculations, trial wavefunctions optimized for systems with different potentials, or containing different numbers of quantum particles, are unlikely to be similarly well converged. This introduces a source of uncontrolled error. Entwistle *et al.* [114] demonstrate that this error can be reduced by the use of ‘variance matching’. In calculating positron binding energies, we wish to calculate relative energies between a bare molecule and the same molecule with an additional positron. Typically, the trial wavefunction will be less well-converged for the positronic system.

At the end of training, the variance of the ground-state energy for the positronic system is estimated from inference samples. In this case, we estimate the variance using  $\sim 40$  million MCMC samples. Then, we carry out inference for the trial wavefunction of the bare molecule at snapshots of the trial wavefunction parameters obtained during training. To calculate the relative energy, we take the ground-state energy of the bare molecular system as the value obtained when the variance of the ground-state energy is equal to that obtained for the positronic system at the end of training.

The results of this procedure are shown in fig. D.1 for the benzene molecule. Due to the stochastic nature of the training procedure, the variance and ground-state energy obtained from the trial wavefunction are not strictly monotonic as a function of training iteration. However, they do generally decrease. We find that the variance of the ground-state energy of the positronic system most closely matches the variance of the ground-state energy of the bare molecule after 175000 training iterations of the trial wavefunction for the bare molecule. Thus, we use the ground-state energy of the bare molecule obtained from this checkpointed trial wavefunction in calculating the positron binding energy.

It is to be emphasized that this procedure is *not* theoretically justified. There is no guarantee that equal variances between wavefunctions containing different

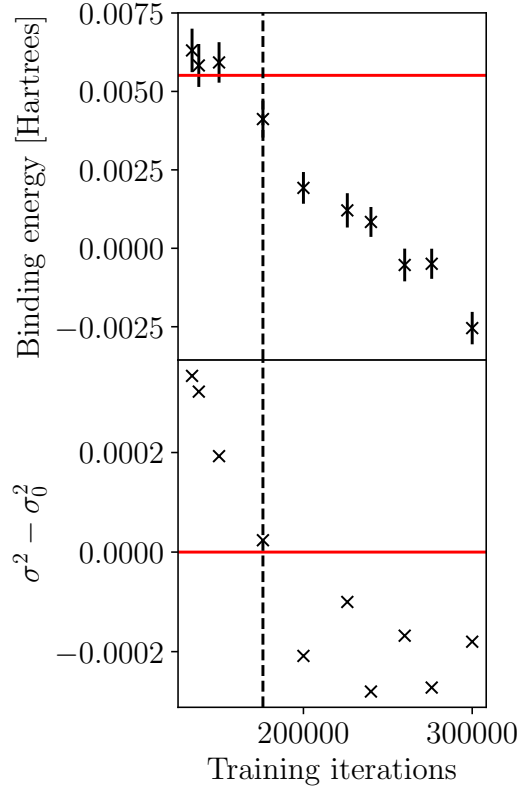


Figure D.1: Quantities used in obtaining the bare molecule ground-state energy used in the variance-matched calculation of the positron binding energy of the benzene molecule. **Top:** Binding energy as a function of training iteration for the bare benzene molecule. Horizontal red line indicates the positron binding energy from ref. [207]. **Bottom:** Difference between the variance obtained from training checkpoints of the bare-molecular trial wavefunction, and the variance obtained at the end of training for the converged positronic trial wavefunction.

numbers of quantum particles should indicate equally well converged ground-state energies. This procedure is loosely based on the intuition that the variance is linear in the error of the ground-state energy, but the slope of this linear relationship is not generally equal for different systems. Despite the ‘fast and loose’ nature of the variance matching procedure, it seems to work well in practice. We would prefer a more rigorous procedure.

# Appendix E

## Data tables

Positron binding tables are reproduced verbatim from ref. [209].

### Positron binding to lithium hydride

| Bondlength [ $a_0$ ] | Energy [Hartrees] |                        |            |                              |
|----------------------|-------------------|------------------------|------------|------------------------------|
|                      | Bare              | $\sigma_{\text{Bare}}$ | Positronic | $\sigma_{\text{Positronic}}$ |
| 2.815                | -8.069021         | 0.000004               | -8.100914  | 0.000004                     |
| 2.882                | -8.069882         | 0.000005               | -8.102727  | 0.000009                     |
| 2.948                | -8.070345         | 0.000003               | -8.104151  | 0.000010                     |
| 3.015                | -8.070507         | 0.000004               | -8.105283  | 0.000004                     |
| 3.082                | -8.070371         | 0.000007               | -8.106226  | 0.000007                     |
| 3.148                | -8.069958         | 0.000003               | -8.106885  | 0.000005                     |
| 3.215                | -8.069331         | 0.000004               | -8.107324  | 0.000003                     |
| 3.281                | -8.068203         | 0.000278               | -8.107583  | 0.000003                     |
| 3.348                | -8.067207         | 0.000287               | -8.107737  | 0.000004                     |
| 3.415                | -8.066352         | 0.000007               | -8.107712  | 0.000011                     |
| 3.481                | -8.065065         | 0.000005               | -8.107542  | 0.000004                     |
| 3.548                | -8.063663         | 0.000005               | -8.107234  | 0.000005                     |
| 3.615                | -8.062156         | 0.000004               | -8.106949  | 0.000006                     |
| 3.682                | -8.060571         | 0.000004               | -8.106551  | 0.000008                     |
| 3.749                | -8.058894         | 0.000008               | -8.105952  | 0.000010                     |
| 3.816                | -8.057171         | 0.000005               | -8.105373  | 0.000132                     |
| 3.883                | -8.055407         | 0.000007               | -8.104972  | 0.000006                     |
| 3.950                | -8.053575         | 0.000007               | -8.104321  | 0.000010                     |
| 4.017                | -8.051748         | 0.000008               | -8.103593  | 0.000004                     |
| 4.084                | -8.049849         | 0.000008               | -8.102936  | 0.000005                     |
| 4.151                | -8.047965         | 0.000005               | -8.102246  | 0.000010                     |

Table E.1: Ground-state energy as a function of bondlength for bare and positronic LiH.

# Positron binding to beryllium oxide

| Bondlength [ $a_0$ ] | Energy [Hartrees] |                        |            |                              |
|----------------------|-------------------|------------------------|------------|------------------------------|
|                      | Bare              | $\sigma_{\text{Bare}}$ | Positronic | $\sigma_{\text{Positronic}}$ |
| Singlet              |                   |                        |            |                              |
| 2.115                | -89.845004        | 0.000024               | -89.865185 | 0.000027                     |
| 2.215                | -89.875412        | 0.000024               | -89.896963 | 0.000019                     |
| 2.315                | -89.893706        | 0.000014               | -89.916586 | 0.000019                     |
| 2.415                | -89.903115        | 0.000042               | -89.927073 | 0.000033                     |
| 2.515                | -89.905721        | 0.000014               | -89.930761 | 0.000057                     |
| 2.615                | -89.903574        | 0.000075               | -89.929260 | 0.000088                     |
| 2.715                | -89.897711        | 0.000070               | -89.924101 | 0.000020                     |
| 2.815                | -89.889610        | 0.000022               | -89.915860 | 0.000043                     |
| 2.915                | -89.879387        | 0.000073               | -89.906657 | 0.000029                     |
| 3.015                | -89.869008        | 0.000073               | -89.895173 | 0.000032                     |
| 3.115                | -89.857212        | 0.000018               | -89.884496 | 0.000039                     |
| 3.215                | -89.846269        | 0.000021               | -89.872785 | 0.000118                     |
| 3.315                | -89.838491        | 0.000082               | -89.860431 | 0.000095                     |
| 3.415                | -89.830544        | 0.000024               | -89.845536 | 0.000027                     |
| 3.515                | -89.820883        | 0.000027               | -89.834588 | 0.000032                     |
| 3.615                | -89.812532        | 0.000032               | -89.824085 | 0.000029                     |
| 3.715                | -89.805062        | 0.000019               | -89.812757 | 0.000077                     |
| 3.815                | -89.798934        | 0.000097               | -89.801516 | 0.000026                     |
| 3.915                | -89.786483        | 0.000030               | -89.792682 | 0.000042                     |
| 4.015                | -89.794044        | 0.000051               | -89.794320 | 0.000043                     |
| Triplet              |                   |                        |            |                              |
| 3.515                | -89.824472        | 0.000018               | -89.822049 | 0.000031                     |
| 4.015                | -89.784895        | 0.000023               | -89.788294 | 0.000031                     |
| 4.515                | -89.754943        | 0.000024               | -89.763941 | 0.000031                     |
| 5.015                | -89.738802        | 0.000023               | -89.745154 | 0.000032                     |
| 5.515                | -89.732889        | 0.000022               | -89.736450 | 0.000033                     |
| 6.015                | -89.731046        | 0.000022               | -89.732024 | 0.000033                     |
| 6.515                | -89.730538        | 0.000023               | -89.730718 | 0.000034                     |
| 7.015                | -89.730307        | 0.000022               | -89.731238 | 0.000031                     |
| 7.515                | -89.730150        | 0.000021               | -89.731904 | 0.000030                     |
| 8.015                | -89.730040        | 0.000021               | -89.731992 | 0.000036                     |
| 8.515                | -89.730035        | 0.000025               | -89.732091 | 0.000031                     |
| 9.015                | -89.729969        | 0.000022               | -89.732011 | 0.000030                     |

Table E.2: Ground-state energy as a function of bondlength for bare and positronic BeO.

## Positron binding to benzene

| System                  | $\langle E \rangle$ [Hartrees] | $\sigma_E^2$ |
|-------------------------|--------------------------------|--------------|
| Positronic @ 300k iter. | -232.223293                    | 0.000847     |
| Bare @ 134k iter.       | -232.216986                    | 0.001202     |
| Bare @ 138k iter.       | -232.217466                    | 0.001170     |
| Bare @ 150k iter.       | -232.217369                    | 0.001040     |
| Bare @ 176k iter.       | -232.219174                    | 0.000871     |
| Bare @ 200k iter.       | -232.221368                    | 0.000639     |
| Bare @ 226k iter.       | -232.222084                    | 0.000747     |
| Bare @ 240k iter.       | -232.222453                    | 0.000568     |
| Bare @ 260k iter.       | -232.223825                    | 0.000680     |
| Bare @ 276k iter.       | -232.223786                    | 0.000576     |
| Bare @ 300k iter.       | -232.225837                    | 0.000668     |

Table E.3: Ground-state energy at the end of training for positronic benzene, and as a function of training iteration for bare benzene.

## Ground-state energies for eight protons in a box

| $r_s$    | $\langle E \rangle$ [Hartrees] |
|----------|--------------------------------|
| 0.377976 | 46.6842(2)                     |
| 0.755953 | 2.7882(4)                      |
| 1.133929 | -2.9633(1)                     |
| 1.511905 | -4.22487(9)                    |
| 1.889882 | -4.57283(7)                    |
| 2.267858 | -4.66273(9)                    |
| 2.645834 | -4.68056(6)                    |

Table E.4: Ground-state energy as a function of Wigner-Seitz radius for eight protons in a box.