Imperial College of Science, Technology and Medicine
Department of Physics

Neural Network Wave Functions for Superfluid Fermi Gases

Hunting down quantum phases of matter from first principles

Wan Tong Lou

Submitted in part fulfilment of the requirements for the degree of Doctor of Philosophy
in Physics of Imperial College of Science, Technology and Medicine
October 2025

Statement of Originality

This is to declare that all content in this thesis is the original work of the author, except where otherwise acknowledged, and that all else is appropriately referenced. Part of the contents in Chapter 4 and Chapter 5 draws upon the following published work by the author:

> Wan Tong Lou et al. "Neural Wave Functions for Superfluids". In: *Physical Review X* 14.2 (May 2024), p. 021030. DOI: `10.1103/PhysRevX.14.021030`.

The article was originally published by the American Physical Society (APS). Reproduction in this thesis is permitted under APS policy, which allows authors to include their own published work in theses or dissertations provided that full bibliographic citation and the APS copyright credit line are included. These policies can be found at `https://journals.aps.org/copyrightFAQ.html`:

> [· · ·] *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.*

In addition, this work is published 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

The search for exotic phases of matter in quantum systems remains a central goal in condensed matter physics. Recent progress in machine learning has driven the development of powerful new computational methods in quantum chemistry and physics. In particular, the application of neural networks to solve the Schrödinger equation within the variational Monte Carlo (VMC) framework has led to the development of the successful neural network VMC (NNVMC) method. This thesis extends and explores the application of NNVMC to the study of strongly interacting Fermi gases.

We first investigated the unitary Fermi gas (UFG)—a system with strong, short-range attractions known to possess a superfluid ground state that is difficult to describe quantitatively. By applying the NNVMC method with the FermiNet Ansatz, we first identified its key limitations before proposing a simple modification. This change led to a substantial improvement in accuracy over the original FermiNet in the UFG, outperforming results from the state-of-the-art simulations while being consistent with experiments.

Next, we explored the phases of the 2D spin-imbalanced Fermi gas. The combined power of NNVMC and our modified FermiNet enabled us not only to observe the long-predicted Fulde-Ferrell-Larkin-Ovchinnikov phase but also to discover a previously unknown phase, characterised by the crystallisation of tightly bound Cooper pairs within a uniform sea of unpaired fermions.

Lastly, we studied an anisotropic Fermi gas featuring spin-dependent splitting in momentum space — a prime signature of the recently discovered form of collinear magnetism, termed altermagnetism. We mapped its 2D phase diagram, providing unbiased estimations for the locations of various phases beyond the mean-field approximation.

In conclusion, our work has demonstrated that the NNVMC method is a powerful and versatile tool for investigating strongly correlated systems, confirming its potential to uncover novel phenomena in the vast landscape of quantum systems.

Acknowledgements

I would like to express my deepest gratitude to my supervisors, Professor Matthew Foulkes and Professor Johannes Knolle, for their invaluable guidance, encouragement, and patience throughout the course of my PhD. Their insights and support have been fundamental to the development of this work.
I am also grateful to Dr. David Pfau and Dr. James Spencer for their generous feedback, thoughtful discussions, and unwavering support, which were instrumental in shaping and refining my research.
I would like to acknowledge all my lovely colleagues and friends in the CMTH group at Imperial for providing a stimulating and supportive environment, as well as for the many fun discussions and good times we have shared.
Special thanks go to Carolyn Dale and Juraci Didone, whose assistance behind the scenes has been indispensable.
On a personal note, I am profoundly thankful to my parents for their enduring support, encouragement, and love. To my friends, thank you for your companionship and for reminding me of life beyond research.
Finally, to Wing Fung, who always stands beside me in my brightest and darkest days.

To my parents, whose support made the beginning possible.

To Wing Fung, whose strength carried me through to the end.

Contents

Statement of Originality **3**

Copyright Declaration **5**

Abstract **7**

Acknowledgements **9**

1 Introduction **23**

2 Background **27**

2.1 BCS-BEC Crossover 27

2.2 Microscopic Model of Ultracold Fermi Gases 28

2.3 BCS Theory of Superconductivity 29

2.3.1 BCS Ground State Wave Function 30

2.3.2 BCS Gap Equation 31

2.3.3 Condensation Energy 32

2.4 BCS Theory for Zero Temperature Dilute Fermi Gases 33

2.4.1 The Gap and The Chemical Potential in 3D 34

2.4.2 The Gap and The Chemical Potential in 2D 38

2.5 Spin-Imbalanced Fermi Gas 39

2.5.1 The Clogston-Chandrasekhar Limit 40

2.5.2 The FFLO Phase 42

2.5.3 The Sarma Phase 46

2.6 Anisotropic Fermi Gas 48

2.6.1 Altermagnetism 48

2.7 Observables 51

2.7.1 Density Matrices and the Condensate Fraction 51

2.7.2 Momentum Densities 52

2.7.3 Mean Field Results 53

3 Neural Network Variational Monte Carlo **55**
3.1 The Many-Body Problem . . . 55
3.2 Variational Monte Carlo . . . 56
3.3 The Slater-Jastrow(-Backflow) Trial Wave Function . . . 58
3.3.1 The Backflow Transformation . . . 59
3.3.2 Spin Independent Observables . . . 60
3.3.3 Multi-determinant Wave Functions . . . 61
3.4 Neural Network Wave Functions . . . 61
3.4.1 Neural Networks . . . 61
3.4.2 Neural Network Quantum States . . . 63
3.4.3 Neural Wave Functions in Continuous Systems . . . 64
3.4.4 Dense Determinant . . . 67
3.4.5 Periodic Systems in NNVMC . . . 67
3.5 Neural Wave Function Optimisation . . . 68
3.5.1 The Stochastic Reconfiguration Method . . . 68
3.5.2 The Kronecker-Factored Approximate Curvature (KFAC) Method . . . 69

4 Neural AGP Wave Functions **73**
4.1 Fixed Particle Number BCS Wave Function . . . 73
4.2 AGP Singlet Wave Function . . . 75
4.2.1 AGPs Wave Function with Unpaired States . . . 75
4.3 AGP Wave Function . . . 76
4.3.1 AGP Wave Function with Unpaired States . . . 78
4.3.2 Relation to the AGPs Wave Function with Unpaired States . . . 79
4.4 Neural Network Pairing Orbitals . . . 80
4.4.1 More Compact Versions of Geminals . . . 81
4.4.2 Other Forms of Geminals . . . 82
4.5 The Neural AGP Wave Functions . . . 83
4.5.1 Relationship between the AGPs FermiNet and the block-diagonal Slater FermiNet . . . 83
4.5.2 Relationship between the Pfaffian FermiNet and the dense Slater FermiNet 87
4.6 Benchmarking Neural AGP Wave Functions on Beryllium and Nitrogen . . . 90
4.6.1 Beryllium . . . 90
4.6.2 Nitrogen . . . 93
4.6.3 Discussion . . . 95

5 Unitary Fermi Gas **97**
5.1 Results . . . 98
5.1.1 Slater FermiNet and AGPs FermiNet . . . 99

5.1.2 AGPs FermiNet and FN-DMC . . . 101
5.1.3 Condensate Fraction . . . 103
5.1.4 Scaling behaviours between the two Ansätze . . . 105
5.2 Discussion . . . 107

6 Spin-Imbalanced Fermi Gas 111
6.1 Results . . . 112
6.1.1 Pair-Momentum Density . . . 114
6.1.2 One-Particle Density . . . 116
6.2 Discussion . . . 116

7 Anisotropic Fermi Gas 121
7.1 Results . . . 121
7.1.1 Condensate Fraction . . . 122
7.1.2 Two-Body Density Matrix and Pair-Momentum Density . . . 123
7.2 Discussion . . . 127

8 Conclusion 131

A Scattering Theory 149
A.1 Lippmann-Schwinger Equation . . . 149
A.1.1 Scattering in dilute Fermi gases . . . 150
A.2 Method of Partial Waves . . . 152
A.2.1 Partial wave expansions in 3D . . . 152
A.2.2 Partial wave expansions in 2D . . . 155
A.3 Regularising the Interaction . . . 157

B Expectation Values in QMC 159
B.1 Density Matrices . . . 159
B.1.1 Improved Estimators for the Condensate Fraction . . . 160

C Supplementary Data and Figures 163
C.1 Anisotropic Fermi Gas . . . 163
C.1.1 Two-Body Density Matrix . . . 163
C.1.2 Pair-Momentum Density . . . 163

List of Figures

2.1 A microscopic illustration of the BCS-BEC crossover as a function of $1/k_F a_s$. The colours of the circles (red and blue) indicates the spins of the particles, and the dashed ellipses indicates pairing between the particles. On the BCS side, the sizes of the pairs are much bigger than the inter-particle separation and the pairs overlap, whereas on the BEC side, the pairs are much more tightly bounded. At unitarity, the sizes of the pairs are roughly the same as the inter-particle separation. 28

2.2 Comparisons various expressions for the gap Δ and the chemical potential μ as functions of the inverse scattering length $1/k_F a_s$ are plotted. The left hand side is the BCS region and the right hand side is the BEC region, with the unitarity limit at $k_F a_s = 0$. Note that while we have plotted the asymptotes for the whole range of $1/k_F a_s$, one should keep in mind that the BCS asymptotes are only valid as $1/k_F a_s \to -\infty$ and the BEC asymptotes are only valid as $1/k_F a_s \to \infty$. . . . 37

2.3 The gap Δ and the chemical potential μ as functions of the negative natural logarithm of the scattering length $-\ln(k_F a_s)$. The negative logarithm of $k_F a_s$ is a convenient choice of independent variable because it passes through zero in the BCS-BEC crossover regime, aiding characterisation; the negative sign ensures that the interaction strength increases from left to right across the graph (as was also the case in Fig. (2.2)). 39

2.4 The blocked regions of Cooper pairing in momentum space. The concentric circles, from inner to outer, indicate the Fermi surface for spin-down particles, the mean chemical potential $\bar{\mu}$, and the Fermi surface for spin-up particles, respectively. The shaded areas indicate regions where Cooper pairing is blocked. Here, we set $\bar{\mu} = 10$, $h = 2$ and $\Delta_{\mathbf{q}} = 4$, varying only the momentum $\mathbf{q}$. 45

2.5 The blocked regions of Cooper pairing in momentum space for the Sarma phase. The concentric circles, from inner to outer, indicate the Fermi surface for spin-down particles, the mean chemical potential $\bar{\mu}$, and the Fermi surface for spin-up particles, respectively. The shaded areas indicate the regions of momentum space where Cooper pairing is blocked. Here, we set $\bar{\mu} = 10$, $\mathbf{q} = 0$ and $h = 5$, and only tune the gap $\Delta_{\mathbf{q}}$ to obtain different shaded regions. 47

2.6 Fermi surfaces of the two spin species for ferromagnetism (left) and altermagnetism (right) 49

3.1 A schematic diagram for a simple multi-layer perceptron neural network. It is consisted of an input layer (red), two hidden layers (orange), and an output layer (blue). Each layer contain different number of perceptrons and the neighbouring layers are fully connected by the edges. 62

3.2 Schematic diagram of the operations in the intermediate layers of the FermiNet architecture in Eq. (3.23) and Eq. (3.24). The sets of arrays in the one and two-electron streams are first averaged and concatenated before being passed to the next layer. 66

4.1 Comparison of beryllium atom energy between the PfaffNet with diagonal geminals and full geminals, and the Slater FermiNet with dense determinants. The reference energy of the isolated Be atom is taken from Ref. [28]. 91

4.2 Comparison of beryllium dimer energy between the PfaffNet with either one full or one diagonal geminal, and the Slater FermiNet with both 1 and 16 dense determinants. The FCIQMC results are taken from Ref. [67]. 92

4.3 Comparison of beryllium chain energy between the PfaffNet with either one full or one diagonal geminal, and the Slater FermiNet with 1, 2, 4, 8 and 16 dense determinants. The reference energy of the isolated Be atom is taken from Ref. [28]. 93

4.4 Comparison of nitrogen atom energy between the PfaffNet with either one full or one diagonal geminal, and the Slater FermiNet with different number of dense determinants. 94

4.5 Full dissociation curve of the nitrogen dimer molecule. Energy is displayed relative to the spectroscopy value. 95

5.1 Comparison between results obtained using the AGPs FermiNet and the Slater FermiNet for different numbers of particles, N, with $r_s = 1$ and $\mu = 12$. All simulations used 32 determinants, 300,000 optimisation steps, and the same hyperparameters. 100

5.2 Comparison of the system-size dependent values of the Bertsch parameter, ξ, as calculated using the AGPs FermiNet and FN-DMC, with $k_F = 1$ and $\mu = 12$. According to the variational principle, lower values are better. The error bars on the AGPs FermiNet results are smaller than the sizes of the crosses. Inset: difference between the AGPs and FN-DMC values of the Bertsch parameter. The errors in the inset are obtained by adding the standard errors of the FN-DMC and AGPs FermiNet results in quadrature. The latter are obtained by computing the standard error of the MCMC-averaged Bertsch parameter accumulated over 50,000 inference steps. 102

5.3 Pairing gaps calculated with the Slater FermiNet and the AGPs FermiNet for different numbers of particles N, with $r_s = 1$ and $\mu = 12$. 103

5.4 Comparison of the TBDM estimators calculated using the AGPs FermiNet and the Slater FermiNet with $N = 38$, $r_s = 1$ and $\mu = 12$. The error bars show the standard error of the TBDM estimator, accumulated over 2,000 inference steps. Most of the error bars are so small that they are obscured the symbols. 105

5.5 Percentage difference between the Slater FermiNet with various number of block-diagonal determinants and the AGPs FermiNet with one determinant at different system sizes, from 4 to 14 particles. 106

5.6 Number of block-diagonal determinants required for the multi-determinant Slater FermiNet energy to fall within 5% and 10% of the energy obtained using a single-determinant/single-geminal AGPs FermiNet. 107

6.1 The occupied momentum states in the non-interacting (25, 9) Fermi gas. The red crosses and the blue circles denote the occupied states by the spin-up and spin-down particles, respectively. The wave vectors k_x and k_y are measured in units of $G = 2\pi/L$, where L is the length of the side of the square real-space simulation cell. 113

6.2 Pair-momentum density with signatures of the FFLO phase. 114

6.3 Pair-momentum density at the intermediate strength of interaction. 115

6.4 Pair-momentum density with signatures of a standard BEC s-wave superfluid. . . 115

6.5 One-particle density of the $(25, 9)$ spin-imbalanced Fermi gas for a range of different interaction strengths near the weakly interacting BCS limit, plotted separately for different spin channels. The plots on the left-hand side are the densities of spin-up particles and the plots on the right-hand side are the densities of the spin-down particles. 118

6.6 One-particle densities and their 1D projections for the $(25, 9)$ spin-imbalanced Fermi gas at $v_0 = 0.3$, plotted separately for different spin channels. The first row shows the majority spin (spin-up) density and the second row represent the minority spin (spin-down) density. The first column shows a horizontal slice of the 2D density with a fixed $y = y_0 \approx -0.06L$, indicated by the *red dash-dotted horizontal* line on the figure in the third column. The second column shows a vertical slice of the 2D density with a fixed $x = x_0 \approx 0.17L$, indicated by the *red dashed vertical* line on the figure in the third column. In the 1D projections, the black dashed horizontal line represents the average number density, $\bar{n}^\sigma = \frac{N^\sigma}{\Omega}$, of fermions of spin σ in the simulation cell with volume Ω. The blue curves are the actual values of the density across the cell, with the orange curves being running average of the blue curves. 119

6.7 One-particle density of the $(25, 9)$ spin-imbalanced Fermi gas for a range of interaction strengths in the intermediate to strongly interacting regime, plotted separately for different spin channels. The plots on the left-hand side are the densities of spin-up particles and the plots on the right-hand side are the densities of the spin-down particles. 120

7.1 The occupied momentum states in the weakly-interacting (11, 11) anisotropic Fermi gas. The red crosses and the blue circles denote the occupied states by the spin-up and spin-down particles, respectively. The coloured regions at the two Fermi surfaces denote the region bounded by the Debye frequency. The dashed wavy line in magenta and the dash-dotted wavy line in olive indicates possible finite-momentum and zero-momentum pairing, respectively. This arrangement of states is the same for any of our chosen anisotropy $\delta \in \{0.1, 0.3, 0.5, 0.7\}$. 122

7.2 The superfluid condensate fraction $c_{\uparrow\downarrow}$ for a range of values of v_0 and δ. The white number in each square is the value of the condensate fraction at the corresponding v_0 and δ. The single digits in parentheses indicate the standard error in the final decimal place of the corresponding value of the condensate fraction. 124

7.3 The pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^\uparrow, N^\downarrow) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.1, 0.2$ and $\delta = 0.1$. The left-hand figures are the absolute values of the pair-momentum density with the dominant peak at $\mathbf{k} = 0$ masked out, such that other weaker signals can be seen. The middle figures are the original unmasked absolute values of the pair-momentum density. The right-hand figures are the phases of the pair-momentum density. 125

7.4 The real space, translationally averaged two-body density matrix, $\bar{\rho}^{(2)}_{\uparrow\downarrow}(\mathbf{r})$ of the $(N^\uparrow, N^\downarrow) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.1, 0.2$ and $\delta = 0.1$. 126

7.5 The pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.4, 0.5$ and $\delta = 0.1$. The left-hand figures are the absolute values of the pair-momentum density with the dominant peak at $\mathbf{k} = 0$ masked out, such that other weaker signals can be seen. The middle figures are the original unmasked absolute values of the pair-momentum density. The right-hand figures are the phases of the pair-momentum density. 127

7.6 The absolute values of the pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.2, 0.3, 0.35, 0.4$ and fixed $\delta = 0.5$. Note that we have cropped out the regions outside $\mathbf{q} = (\pm 2G, \pm 2G)$, as the signals in those regions are essentially zero. The whole uncropped version of the pair-momentum density can be found in the figures in Appendix C.1. 128

7.7 The absolute values of the pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.2, 0.3, 0.35, 0.4$ and fixed $\delta = 0.7$. Note that we have cropped out the regions outside $\mathbf{q} = (\pm 2G, \pm 2G)$, as the signals in those regions are essentially zero. The whole uncropped version of the pair-momentum density can be found in the figures in Appendix C.1. 129

C.1 The real space, translationally averaged two-body density matrix, $\bar{\rho}_{\uparrow\downarrow}^{(2)}(\mathbf{r})$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.05, 0.1, 0.2$ and $\delta = 0.1, 0.3, 0.5, 0.7$. 164

C.2 The real space, translationally averaged two-body density matrix, $\bar{\rho}_{\uparrow\downarrow}^{(2)}(\mathbf{r})$, of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.3, 0.35, 0.4, 0.5$ and $\delta = 0.1, 0.3, 0.5, 0.7$. 165

C.3 The absolute values and the phases of the pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.05$ and $\delta = 0.1, 0.3, 0.5, 0.7$ 166

C.4 The absolute values and the phases of the pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.1$ and $\delta = 0.1, 0.3, 0.5, 0.7$ 167

C.5 The absolute values and the phases of the pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.2$ and $\delta = 0.1$ 168

C.6 The absolute values and the phases of the pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.3$ and $\delta = 0.1, 0.3, 0.5, 0.7$ 169

C.7 The absolute values and the phases of the pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.35$ and $\delta = 0.1, 0.3, 0.5, 0.7$ 170

C.8 The absolute values and the phases of the pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.4$ and $\delta = 0.1, 0.3, 0.5, 0.7$ 171

C.9 The absolute values and the phases of the pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.5$ and $\delta = 0.1, 0.3, 0.5, 0.7$ 172

List of Tables

5.1 Total energy of the UFG with 36 to 38 particles and their corresponding pairing gap using different wave functions. 101

5.2 Estimates of the superfluid condensate fraction at unitarity using various methods. The quantity $k_F r_e$ is a dimensionless number, indicating the deviation of the simulated system from a perfect UFG with zero-range interaction. 104

6.1 Values of v_0 and their corresponding values of $\eta = \ln(k_F a_s)$, as well as the binding energy per particle $|\epsilon_b|/\epsilon_F = 8e^{-2(\gamma+\eta)}$ at a fixed $\mu = 24$. 113

Chapter 1

Introduction

> "The beauty and clearness of the dynamical theory, which asserts heat and light to be modes of motion, is at present obscured by two clouds. "
>
> *Lord Kelvin*

The urge to understand the physical world around us has been deeply ingrained in human nature. Up until the last century, our understanding to the physical world was thought to be mostly complete. Nevertheless, the discovery of the quantum realm has added a whole new dimension to our understanding of the physical world. Although the basis of the whole quantum theory only depends on a few postulates, the picture they are able to paint is way beyond our imaginations. Many of the modern technologies were grown from this fertilised ground of quantum theories, which could in turn support the development of numerical and experimental techniques, allowing us to explore deeper in the quantum world.

Despite the joint effort of many generations of physicists, our understanding of the quantum world still remains incomplete. In low-energy physics, one of the major challenges is the exponential complexity arises from the inclusion of more and more particles. Indeed, as pointed out by Philip Anderson, one of the founders of the field of condensed matter physics, a collection of many small objects can lead to *emergent* properties of the whole collection as a bulk, just like the concept of "wetness" only exists in water, but not in any individual water molecule. The notion of emergence give raise to vastly different properties and physical phenomena that can be found in the matter around us.

As the effect of thermal fluctuations decreases with temperature, quantum effects will dominate, allowing various interesting and bizarre quantum phases of matter to appear in condensed matter systems. The study of collection of particles at low temperatures has always been a thriving field because of its ability of hosting remarkable complexity and diversity of phases of matter, such as superconductivity and superfluidity. Over the past decade, the substantial development of experimental techniques in dilute ultracold atomic gases sparked the interests in the field even

further [60]. By tuning the effective strength of interactions between the ultracold atoms with an external magnetic field and driving it across the Feshbach resonance [34], it is possible to study the long-predicted phenomenon of the Bardeen-Cooper-Schrieffer (BCS) superfluid phase to Bose-Einstein condensate (BEC) crossover in the laboratory. As the BCS-BEC crossover is a universal behaviour in attractive gases of fermionic particles, the study of ultracold atomic gases can be used to describe vastly different physical systems. This includes high-temperature (high-T_c) superconductors [32], neutron matter in neutron stars [52, 118], and even quark-gluon plasma in QCD [3].

Recent progress in computational science has been driven by the rapid development of the semiconductor industry and concurrent advancements in numerical algorithms. In particular, the widespread availability of powerful hardware such as the graphics processing units (GPUs) has made the large-scale application of artificial neural networks computationally feasible, fuelling the modern revolution in machine learning and artificial intelligence. While its most famous application is the creation of large language models (LLMs), this synergy has also sparked the development of novel computational methods across a wide range of scientific fields, enabling breakthroughs such as DeepMind's AlphaFold in structural biology, or the use of deep neural networks in the detection of gravitational waves from LIGO data.

This same paradigm is now being applied to one of the most formidable challenges in quantum mechanics — solving the many-body Schrödinger equation. Similar to classical systems, where analytical solutions to the coupled differential equations from Newton's laws of motion remain inaccessible beyond two masses, the exact solutions to the many-body Schrödinger equations of an interacting quantum system also impossible to obtain. As such, numerical methods are often employed to obtain approximate solutions to the many-body Schrödinger equation if one wish to go beyond mean-field results. A particularly promising approach, and the focus of this thesis, is to represent the complex quantum wave function itself as a neural network. This technique, combined with the powerful variational Monte Carlo (VMC) framework, is known as the neural network variational Monte Carlo (NNVMC) method, which allows one to find the ground state of a many-body quantum system with remarkable accuracy [19, 23, 24, 95, 96, 117, 121, 122, 138]. We provide a more detailed description of this method in Chapter 3. In this work, we utilise this approach and extend it to study different dilute Fermi gases. In particular, we focus our attention on three different Fermi gas systems: the unitary Fermi gas, the spin-imbalanced Fermi gas, and the anisotropic Fermi gas.

This thesis is organised as follows. In Chapter 2, we present a brief introduction to the theories of dilute Fermi gases with attractive interactions. In Chapter 3, we discuss the neural network variational Monte Carlo method, and how it is differed from the conventional variational Monte Carlo method. In Chapter 4, we discuss the origin of the class of antisymmetric geminal power (AGP) wave functions, and how one could apply these ideas to build neural AGP wave functions. In Chapter 5, we apply the idea of neural network variational Monte Carlo and neural AGP wave functions to study a model system of the unitary Fermi gas that possesses a superfluid

ground state. We also discuss how the standard neural wave function, which is based on the Slater determinants, fails in this system, and how it could be improved by using neural AGP wave functions. In Chapter 6 and Chapter 7, we utilise the neural AGP wave function and apply them to study the more challenging symmetry-broken systems — the dilute Fermi gas with spin population imbalance, and the dilute Fermi gas with anisotropic masses. The latter could be used as an effective model to describe bulk superconductivity in the recently discovered phases of matter — altermagnetism. Finally, we provide a brief conclusion and outlook of our work in the last chapter.

Chapter 2

Background

In this chapter, we present a brief introduction to the theory of attractive Fermi gases, focusing on the exotic phases of matter at zero temperature.

2.1 BCS-BEC Crossover

In ultracold atomic Fermi gases, the system is sufficiently dilute that the effective range of interaction r_e is much smaller than the inter-particle separation r_s, *i.e.*, $r_e \ll r_s$. Under this condition, interactions between same-spin particles can be neglected due to the Pauli exclusion principle, and the interactions between opposite-spin particles are characterised by the s-wave scattering length a_s [128]. Hence, the interaction-driven BCS-BEC crossover in attractive Fermi gases is characterised by two length scales only – the s-wave scattering length a_s, and the Fermi wave vector k_F. In fact, one can characterise the crossover by the dimensionless parameter $1/k_F a_s$ [88], ranging from $-\infty$ to ∞. If $1/k_F a_s \to -\infty$, also known as the BCS limit, the system is a weakly attractive gas of fermions, forming Cooper pairs [8] with size much bigger than the inter-particle separation. If $1/k_F a_s \to \infty$, also known as the BEC limit, the system is a gas of bosonic molecules, formed by strongly bounded pairs of fermions. Between the two limits, at $1/k_F a_s = 0$, all length scales in the system vanish as the scattering length diverges, except for the inter-particle separation $1/k_F$. Hence, all thermodynamical quantities, among them the ground state energy, depend only on this length scale at $T = 0$. This limit is also known as the *unitarity limit*, originated from the unitarity limit known in scattering theory. The system is thus known as the *unitary Fermi gas* (UFG) [128]. An illustration of the BCS-BEC crossover as a function of $1/k_F a_s$ is shown in Fig. (2.1).

While systems at the two limits, the weakly interacting BCS limit and the strongly interacting BEC limit, can be well described theoretically, they require different theories. The major difficulty is to understand the region around unitarity, which is not well described by either theory. Here, numerical methods are required to bridge the gap between theory and experiment. We will discuss the unitary Fermi gas in detail in Chapter 5.

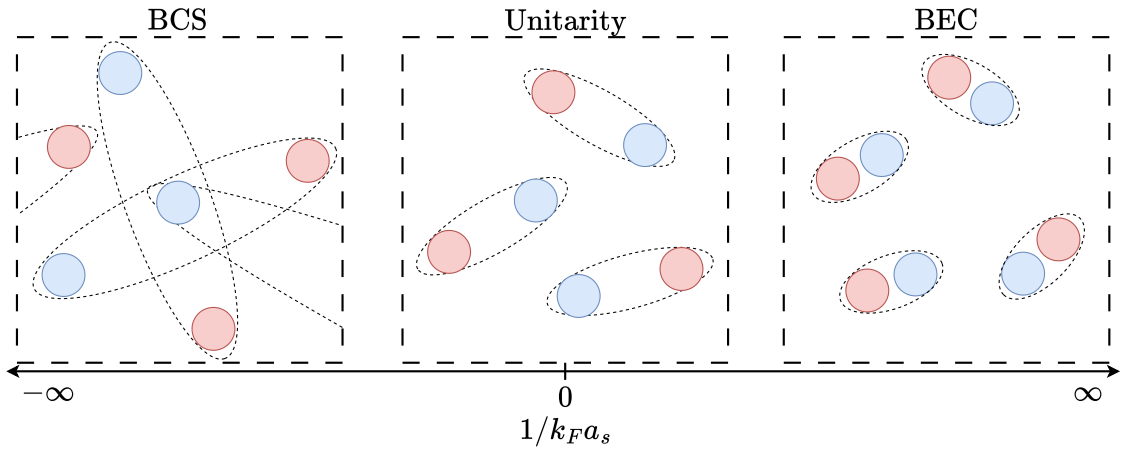


Figure 2.1: A microscopic illustration of the BCS-BEC crossover as a function of $1/k_F a_s$. The colours of the circles (red and blue) indicates the spins of the particles, and the dashed ellipses indicates pairing between the particles. On the BCS side, the sizes of the pairs are much bigger than the inter-particle separation and the pairs overlap, whereas on the BEC side, the pairs are much more tightly bounded. At unitarity, the sizes of the pairs are roughly the same as the inter-particle separation.

2.2 Microscopic Model of Ultracold Fermi Gases

To model ultracold Fermi gases, one can consider a two-component Fermi gas in a box of volume Ω with periodic boundary conditions and the following general Hamiltonian of interacting spin-1/2 fermions in momentum space

$$\hat{\mathcal{H}} = \sum_{\mathbf{k}\sigma} \xi_{\mathbf{k}\sigma} \hat{c}^\dagger_{\mathbf{k}\sigma} \hat{c}_{\mathbf{k}\sigma} + \frac{1}{\Omega} \sum_{\sigma_1\sigma_2} \sum_{\mathbf{k}_1,\mathbf{k}_2,\mathbf{p}} \frac{1}{2} \mathcal{U}^{\sigma_1\sigma_2}_{\mathbf{p}} \hat{c}^\dagger_{\mathbf{k}_1+\mathbf{p}\sigma_1} \hat{c}^\dagger_{\mathbf{k}_2-\mathbf{p}\sigma_2} \hat{c}_{\mathbf{k}_2\sigma_2} \hat{c}_{\mathbf{k}_1\sigma_1}. \tag{2.1}$$

where $\sigma_1, \sigma_2 \in \{\uparrow, \downarrow\}$, Ω is the volume of the system, $\xi_{\mathbf{k}\sigma} = \epsilon_{\mathbf{k}} - \mu_\sigma$, and the summations are over the first Brillouin zone. In spin-balanced (unpolarised) systems, the chemical potentials for the two spin species are identical, so the spin index can be dropped, *i.e.*, $\mu_\uparrow = \mu_\downarrow = \mu$. The interaction $\mathcal{U}^{\sigma_1\sigma_2}_{\mathbf{p}} = \frac{1}{\Omega} \int d\mathbf{r} e^{-i\mathbf{p}\cdot\mathbf{r}} \mathcal{U}^{\sigma_1\sigma_2}(\mathbf{r})$ is the Fourier transform of the real space interaction $\mathcal{U}^{\sigma_1\sigma_2}(\mathbf{r})$. Using the fact that the system is dilute and the effective range of interaction r_e is small, *i.e.* $r_e \ll r_s$, such that the same spins can be treated as non-interacting due to Pauli's principle, and only opposite spins interact, we can impose the conditions $\sigma_1 + \sigma_2 = 0$ and redefine the momenta

$$\mathbf{k}_1 = \mathbf{k}' + \mathbf{q}/2, \qquad \mathbf{k}_2 = -\mathbf{k}' + \mathbf{q}/2, \qquad \mathbf{p} = \mathbf{k} - \mathbf{k}', \tag{2.2}$$

Eq (2.1) becomes

$$\hat{H} = \sum_{\mathbf{k}\sigma} \xi_{\mathbf{k}\sigma} \hat{c}^\dagger_{\mathbf{k}\sigma} \hat{c}_{\mathbf{k}\sigma} + \frac{1}{\Omega} \sum_{\mathbf{k},\mathbf{k}',\mathbf{q}} U_{\mathbf{k},\mathbf{k}'} \hat{c}^\dagger_{\mathbf{k}+\mathbf{q}/2\uparrow} \hat{c}^\dagger_{-\mathbf{k}+\mathbf{q}/2\downarrow} \hat{c}_{-\mathbf{k}'+\mathbf{q}/2\downarrow} \hat{c}_{\mathbf{k}'+\mathbf{q}/2\uparrow} \tag{2.3}$$

where $\sigma \in \{\uparrow, \downarrow\}$ and $U_{\mathbf{k},\mathbf{k}'} = U(\mathbf{k}-\mathbf{k}') = \mathcal{U}^{\uparrow\downarrow}_{\mathbf{k}-\mathbf{k}'}$. Note that the factor of 1/2 in the interaction term is cancelled by the summation over σ_1, σ_2 under the condition $\sigma_1 + \sigma_2 = 0$. It is also convenient to approximate the interaction to be a contact attractive interaction, *i.e.* $U_{\mathbf{k},\mathbf{k}'} = -U_0$, in ultracold Fermi gases. At low energies, the only relevant interaction is s-wave. Hence, we can set $\mathbf{q} = 0$ and arrive at the *reduced BCS Hamiltonian*

$$\hat{H}_{\mathrm{BCS}} = \sum_{\mathbf{k}\sigma} \xi_{\mathbf{k}\sigma} \hat{c}^\dagger_{\mathbf{k}\sigma} \hat{c}_{\mathbf{k}\sigma} - \frac{U_0}{\Omega} \sum_{\mathbf{k},\mathbf{k}'} \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} \hat{c}_{-\mathbf{k}'\downarrow} \hat{c}_{\mathbf{k}'\uparrow}. \tag{2.4}$$

The real space equivalence of this Hamiltonian is what we wish to simulate with the neural network variational Monte Carlo, which we will present in detail in the subsequent chapters. In this chapter, we will provide a brief overview on the analytical methods, mainly the BCS theory of superconductivity, to provide a theoretical picture of the ultracold Fermi gases.

2.3 BCS Theory of Superconductivity

While the general form of Eq. (2.4) at $T = 0$ is not analytical solvable, one can use a *mean-field approximation* to reduce the two-body interaction term to a one-body approximations [1]

$$\hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} \hat{c}_{-\mathbf{k}'\downarrow} \hat{c}_{\mathbf{k}'\uparrow} \approx \left\langle \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} \right\rangle \hat{c}_{-\mathbf{k}'\downarrow} \hat{c}_{\mathbf{k}'\uparrow} + \left\langle \hat{c}_{-\mathbf{k}'\downarrow} \hat{c}_{\mathbf{k}'\uparrow} \right\rangle \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} - \left\langle \hat{c}_{-\mathbf{k}'\downarrow} \hat{c}_{\mathbf{k}'\uparrow} \right\rangle \left\langle \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} \right\rangle \tag{2.5}$$

where terms second-order in the fluctuations have been ignored as we assume the fluctuations are much smaller than the expected values, e.g., $\left| \left\langle \left\langle \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} \right\rangle - \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} \right\rangle \right| \ll \left| \left\langle \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} \right\rangle \right|$. In addition, we define the *superconducting gap function* as

$$\Delta = -\frac{U_0}{\Omega} \sum_{\mathbf{k}'} \left\langle \hat{c}_{-\mathbf{k}'\downarrow} \hat{c}_{\mathbf{k}'\uparrow} \right\rangle \tag{2.6}$$

and the BCS Hamiltonian from Eq. (2.4) becomes

$$\begin{aligned} \hat{H}^{\mathrm{MF}}_{\mathrm{BCS}} &= \sum_{\mathbf{k}} \left(\xi_{\mathbf{k}\uparrow} \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}_{\mathbf{k}\uparrow} + \xi_{-\mathbf{k}\downarrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} \hat{c}_{-\mathbf{k}\downarrow} + \Delta \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} + \Delta^* \hat{c}_{-\mathbf{k}\downarrow} \hat{c}_{\mathbf{k}\uparrow} - \Delta \left\langle \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} \right\rangle \right) \\ &= \sum_{\mathbf{k}} \begin{pmatrix} \hat{c}^\dagger_{\mathbf{k}\uparrow} & \hat{c}_{-\mathbf{k}\downarrow} \end{pmatrix} \begin{pmatrix} \xi_{\mathbf{k}\uparrow} & \Delta \\ \Delta^* & -\xi_{-\mathbf{k}\downarrow} \end{pmatrix} \begin{pmatrix} \hat{c}_{\mathbf{k}\uparrow} \\ \hat{c}^\dagger_{-\mathbf{k}\downarrow} \end{pmatrix} + \sum_{\mathbf{k}} \left[\xi_{-\mathbf{k}\downarrow} - \Delta \left\langle \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} \right\rangle \right] \end{aligned} \tag{2.7}$$

[1] Apart from these two terms, a full mean-field approximation contains other terms such as $\left\langle \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}_{\mathbf{k}'\uparrow} \right\rangle \hat{c}^\dagger_{-\mathbf{k}\downarrow} \hat{c}_{-\mathbf{k}'\downarrow}$, which are not of interest here as they merely shift the energy in the one-body terms. Hence, we can assume that they have already been absorbed into the one-body Hamiltonian.

where the second term is just a constant shift in energy. Eq. (2.7) can be diagonalised using the *Bogoliubov transformation*

$$\begin{pmatrix} \hat{c}_{\mathbf{k}\uparrow} \\ \hat{c}^\dagger_{-\mathbf{k}\downarrow} \end{pmatrix} = \begin{pmatrix} u^*_{\mathbf{k}} & v_{\mathbf{k}} \\ -v^*_{\mathbf{k}} & u_{\mathbf{k}} \end{pmatrix} \begin{pmatrix} \hat{\gamma}_{\mathbf{k}\uparrow} \\ \hat{\gamma}^\dagger_{-\mathbf{k}\downarrow} \end{pmatrix} \qquad \begin{aligned} |u_{\mathbf{k}}|^2 + |v_{\mathbf{k}}|^2 &= 1 \\ u^*_{\mathbf{k}} v_{\mathbf{k}} - u_{\mathbf{k}} v^*_{\mathbf{k}} &= 0 \end{aligned} \tag{2.8}$$

where $\hat{\gamma}_{\mathbf{k}\uparrow}$ and $\hat{\gamma}^\dagger_{-\mathbf{k}\downarrow}$ are the Bogoliubov quasiparticle operators that satisfies the fermionic anti-commutation relations $\{\hat{\gamma}_{\mathbf{k}\sigma_i}, \hat{\gamma}^\dagger_{\mathbf{q}\sigma_j}\} = \delta_{\mathbf{kq}}\delta_{ij}$ and $\{\hat{\gamma}_{\mathbf{k}\sigma_i}, \hat{\gamma}_{\mathbf{q}\sigma_j}\} = \{\hat{\gamma}^\dagger_{\mathbf{k}\sigma_i}, \hat{\gamma}^\dagger_{\mathbf{q}\sigma_j}\} = 0$. The two conditions on $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$ in Eq. (2.8) comes from the anti-commutation relations of the operators. After diagonalisation, the Hamiltonian in Eq. (2.7) becomes

$$\begin{aligned} \hat{H}^{\mathrm{MF}}_{\mathrm{BCS}} &= \sum_{\mathbf{k}} \begin{pmatrix} \hat{\gamma}^\dagger_{\mathbf{k}\uparrow} & \hat{\gamma}_{-\mathbf{k}\downarrow} \end{pmatrix} \begin{pmatrix} E_{\mathbf{k}\uparrow} & 0 \\ 0 & -E_{-\mathbf{k}\downarrow} \end{pmatrix} \begin{pmatrix} \hat{\gamma}_{\mathbf{k}\uparrow} \\ \hat{\gamma}^\dagger_{-\mathbf{k}\downarrow} \end{pmatrix} + \sum_{\mathbf{k}} \left[\xi_{-\mathbf{k}\downarrow} - \Delta \left\langle \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} \right\rangle \right] \\ &= \sum_{\mathbf{k}\sigma} E_{\mathbf{k}\sigma} \hat{\gamma}^\dagger_{\mathbf{k}\sigma} \hat{\gamma}_{\mathbf{k}\sigma} + E_s \end{aligned} \tag{2.9}$$

where the matrix elements are given by

$$\begin{aligned} E_{\mathbf{k}\uparrow} &= \sqrt{\xi^{+2}_{\mathbf{k}} + |\Delta|^2} + \xi^-_{\mathbf{k}} \\ E_{-\mathbf{k}\downarrow} &= \sqrt{\xi^{+2}_{\mathbf{k}} + |\Delta|^2} - \xi^-_{\mathbf{k}} \end{aligned} \qquad \xi^\pm_{\mathbf{k}} = \frac{1}{2}(\xi_{\mathbf{k}\uparrow} \pm \xi_{-\mathbf{k}\downarrow}). \tag{2.10}$$

Note that due to the absence of a magnetic field, we can set $\xi_{\mathbf{k}\uparrow} = \xi_{-\mathbf{k}\downarrow} = \xi_{\mathbf{k}}$, such that $\xi^+_{\mathbf{k}} = \xi_{\mathbf{k}}$ and $\xi^-_{\mathbf{k}} = 0$. This gives

$$E_{\mathbf{k}\uparrow} = E_{-\mathbf{k}\downarrow} = E_{\mathbf{k}} = \sqrt{\xi^2_{\mathbf{k}} + |\Delta|^2}. \tag{2.11}$$

Finally, E_s is a constant shift in energy, given by [2]

$$E_s = \sum_{\mathbf{k}} \left[\left(\xi_{\mathbf{k}} - \sqrt{\xi^2_{\mathbf{k}} + |\Delta|^2} \right) - \Delta \left\langle \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} \right\rangle \right]. \tag{2.12}$$

2.3.1 BCS Ground State Wave Function

Having diagonalised the Hamiltonian, one can now construct the BCS ground state. Firstly, one has to construct the Bogoliubov vacuum $|0_\gamma\rangle$ which satisfies $\hat{\gamma}_{\mathbf{k}\sigma} |0_\gamma\rangle = 0$ for all $\mathbf{k}, \sigma$. From $\hat{\gamma}^2_{\mathbf{k}\sigma} = 0$, one can see that the state that satisfy this condition is $|0_\gamma\rangle = \prod_{\mathbf{k}} \hat{\gamma}_{\mathbf{k}\uparrow} \hat{\gamma}_{-\mathbf{k}\downarrow} |0\rangle$, where $|0\rangle$ is the true vacuum. Transforming this state back to the original fermionic operators $\hat{c}_{\mathbf{k}\sigma}$, we get

$$|0_\gamma\rangle = \prod_{\mathbf{k}} \left(u_{\mathbf{k}} \hat{c}_{\mathbf{k}\uparrow} - v_{\mathbf{k}} \hat{c}^\dagger_{-\mathbf{k}\downarrow} \right) \left(u_{\mathbf{k}} \hat{c}_{-\mathbf{k}\downarrow} + v_{\mathbf{k}} \hat{c}^\dagger_{\mathbf{k}\uparrow} \right) |0\rangle = \prod_{\mathbf{k}} v_{\mathbf{k}} \left(u_{\mathbf{k}} + v_{\mathbf{k}} \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} \right) |0\rangle \tag{2.13}$$

[2] Although this constant shift in energy does not affect the ground state wave function and the gap, this is still needed for subsequent sections when we discuss the stability of the superconducting state under an external magnetic field.

where the identities for $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$ in Eq. (2.8) and the anti-commutation relations are used to arrive at the final expression. Hence, the BCS wave function is given by

$$|\Psi_{\text{BCS}}\rangle = \prod_{\mathbf{k}} \left(u_{\mathbf{k}} + v_{\mathbf{k}} \hat{c}^{\dagger}_{\mathbf{k}\uparrow} \hat{c}^{\dagger}_{-\mathbf{k}\downarrow} \right) |0\rangle \tag{2.14}$$

which is also normalised as one can show $\langle \Psi_{\text{BCS}} | \Psi_{\text{BCS}} \rangle = 1$.

It is also possible to express $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$ in terms of the energies. If we assume the gap Δ to be real, using Eq. (2.7), Eq. (2.8) and Eq. (2.9), we get

$$|u_{\mathbf{k}}|^2 = \frac{1}{2}\left(1 + \frac{\xi_{\mathbf{k}}}{\sqrt{\xi_{\mathbf{k}}^2 + \Delta^2}}\right) \qquad |v_{\mathbf{k}}|^2 = \frac{1}{2}\left(1 - \frac{\xi_{\mathbf{k}}}{\sqrt{\xi_{\mathbf{k}}^2 + \Delta^2}}\right). \tag{2.15}$$

2.3.2 BCS Gap Equation

Using Eq. (2.14) to evaluate $\left\langle \hat{c}^{\dagger}_{\mathbf{k}\sigma} \hat{c}_{\mathbf{k}\sigma} \right\rangle$ and $\left\langle \hat{c}_{-\mathbf{k}\downarrow} \hat{c}_{\mathbf{k}\uparrow} \right\rangle$, we get

$$\left\langle \hat{c}^{\dagger}_{\mathbf{k}\sigma} \hat{c}_{\mathbf{k}\sigma} \right\rangle = |v_{\mathbf{k}}|^2, \qquad \left\langle \hat{c}_{-\mathbf{k}\downarrow} \hat{c}_{\mathbf{k}\uparrow} \right\rangle = -\frac{\Delta}{2\sqrt{\xi_{\mathbf{k}}^2 + \Delta^2}} \tag{2.16}$$

and hence the number density,

$$n = \frac{1}{\Omega} \sum_{\mathbf{k}\sigma} \left\langle \hat{c}^{\dagger}_{\mathbf{k}\sigma} \hat{c}_{\mathbf{k}\sigma} \right\rangle = \frac{1}{\Omega} \sum_{\mathbf{k}} 2|v_{\mathbf{k}}|^2 = \frac{1}{\Omega} \sum_{\mathbf{k}} \left(1 - \frac{\xi_{\mathbf{k}}}{\sqrt{\xi_{\mathbf{k}}^2 + \Delta^2}}\right). \tag{2.17}$$

and the famous BCS gap equation at $T = 0$

$$\Delta = \frac{U_0}{\Omega} \sum_{\mathbf{k}'} \frac{\Delta}{2\sqrt{\xi_{\mathbf{k}'}^2 + \Delta^2}}. \tag{2.18}$$

In particular, the gap equation Eq. (2.18) has a trivial solution, $\Delta = 0$, which corresponds to a normal Fermi liquid state.

To determine the non-trivial solution, we need to make a crude assumption that the interaction only exists within a narrow region near the Fermi surface, such that pairing only happens within a narrow region near the Fermi surface $|\xi_{\mathbf{k}}| < \hbar\omega_D$. This assumption might seems ad hoc, but we will stick with it for now for our discussion here, and do a proper analysis in the next section. Under this assumption, the gap is non-zero $\Delta \neq 0$ only when the particle energy is within a certain range at the Fermi surface, *i.e.*

$$\Delta = \begin{cases} \Delta_0, & |\xi_{\mathbf{k}}| < \hbar\omega_D \\ 0, & \text{otherwise}, \end{cases} \tag{2.19}$$

this turns the summation in Eq. (2.18) into an integral $\sum_{\mathbf{k}'} \to \int d\epsilon g(\epsilon) \approx g(\epsilon_F) \int d\epsilon'$ where $g(\epsilon_F)$ is the density of states at the Fermi surface and $\epsilon' = \epsilon - \mu$ is the relative energy from the chemical potential. Under this approximation, we get

$$1 = \frac{g(\epsilon_F)U_0}{\Omega} \int_{-\hbar\omega_D}^{\hbar\omega_D} d\epsilon' \frac{1}{2\sqrt{\epsilon'^2 + \Delta_0^2}} \approx \frac{g(\epsilon_F)U_0}{\Omega} \ln\left(\frac{2\hbar\omega_D}{\Delta_0}\right) \tag{2.20}$$

where we have assumed $\hbar\omega_D \gg \Delta_0$. Hence, we have obtained an expression for the superconducting gap in BCS theory

$$\Delta_0 = 2\hbar\omega_D e^{-\frac{\Omega}{g(\epsilon_F)U_0}}. \tag{2.21}$$

The cutoff frequency ω_D is known as the Debye frequency. In conventional s-wave superconductors, the Debye frequency is determined by the effective electron-phonon interactions, which are dependent on the material. In particular, at the weakly interacting limit of conventional superconductors, the superconducting gap is much smaller than the Debye energy, $\hbar\omega_D$, implying that the overall effective electron-electron interaction, which consists of the screened Coulomb repulsion and the phonon-induced attraction, can be treated as a delta potential with an energy cutoff [89]. Thus, conventional s-wave superconductors can be well described by the standard BCS theory. However, to extend the BCS theory to describe dilute Fermi gases across the whole BCS-BEC crossover, one would need to get rid of the model-dependent quantities like the Debye cutoff and the contact interaction strength, and replace them with the s-wave scattering length. We will discuss this in more details in Sec. 2.4. For now, we focus on another quantity derived from the standard BCS theory – the condensation energy, which we will need in later sections.

2.3.3 Condensation Energy

The final result we need from the standard BCS theory is the expression for the condensation energy, which will be discussed in this section.

In the energy perspective, the Cooper instability happens because the superconducting state has a lower energy than the normal Fermi liquid state, and the difference in energy is known as the condensation energy. Using the BCS wave function in Eq. (2.14) to calculate the expected value of the Hamiltonian in Eq. (2.9), the BCS ground state energy is given by

$$\langle \Psi_{\mathrm{BCS}} | \hat{H}_{\mathrm{BCS}}^{\mathrm{MF}} | \Psi_{\mathrm{BCS}} \rangle = E_s = \sum_{\mathbf{k}} \left[(\xi_{\mathbf{k}} - E_{\mathbf{k}}) + \frac{\Delta^2}{2E_{\mathbf{k}}} \right] \tag{2.22}$$

which is just the constant shift in energy from Eq. (2.12), since the part of the Hamiltonian from Eq. (2.9) that contains operators has a zero expectation value. On the other hand, the normal state, which can be obtained by solving the Hamiltonian in Eq. (2.7) with $\Delta = 0$, is just a Fermi

sphere of doubly occupied states $\prod_{|\mathbf{k}|<k_F} \hat{c}^\dagger_{\mathbf{k}\uparrow}\hat{c}^\dagger_{\mathbf{k}\downarrow}|0\rangle$. Hence, the normal state energy is given by

$$E_n = \sum_{|\mathbf{k}|<k_F} (\xi_{\mathbf{k}\uparrow} + \xi_{\mathbf{k}\downarrow}) = 2\sum_{|\mathbf{k}|<k_F} \xi_{\mathbf{k}}. \tag{2.23}$$

The condensate energy is the difference between the two $E_c = E_s - E_n$.

With the cutoff $|\xi_{\mathbf{k}}| < \hbar\omega_D$, the condensation energy is given by

$$\begin{aligned} E_c = E_s - E_n &= \sum_{|\mathbf{k}|<k_F} \left[(-\xi_{\mathbf{k}} - E_{\mathbf{k}}) + \frac{\Delta^2}{2E_{\mathbf{k}}}\right] + \sum_{|\mathbf{k}|>k_F} \left[(\xi_{\mathbf{k}} - E_{\mathbf{k}}) + \frac{\Delta^2}{2E_{\mathbf{k}}}\right] \\ &= \sum_{0<\xi_{\mathbf{k}}<\hbar\omega_D} \left[2\left(\xi_{\mathbf{k}} - E_{\mathbf{k}}\right) + \frac{\Delta_0^2}{E_{\mathbf{k}}}\right] \\ &\approx g(\epsilon_F)\Delta_0^2 \int_0^{\hbar\omega_D/\Delta_0} dx \left[2x - 2\sqrt{x^2+1} + \frac{1}{\sqrt{x^2+1}}\right] \\ &\approx -\frac{1}{2} g(\epsilon_F)\Delta_0^2 \end{aligned} \tag{2.24}$$

where we have used the fact that $\frac{\hbar\omega_D}{\Delta_0} \gg 1$ to arrive at the final expression. Note that this expression is valid for *any* spatial dimension, because the only quantity that depends on the dimension d is the density of state $g(\epsilon)$, which is assumed to be a constant, $g(\epsilon) \approx g(\epsilon_F)$, over the region $|\xi_{\mathbf{k}}| < \hbar\omega_D$ for a given d.

2.4 BCS Theory for Zero Temperature Dilute Fermi Gases

The BCS theory is successful in many senses. Not only does it explain the origin of superconductivity in conventional superconductors, but it also makes accurate predictions for different quantities, such as the superconducting gap and the transition temperature.

Despite the triumphs, the theory cannot be applied directly to explain superfluidity in dilute Fermi gases. The main problem comes from the Debye cutoff energy as none of the physical processes in dilute Fermi gases produce such a cutoff. The two model-dependent quantities, the interaction strength U_0 and the cutoff ω_D, are unphysical in dilute Fermi gases. However, if one does not assume the interaction to be zero at energies higher than some cutoff energy ($\hbar\omega_D$ in this case), the summation term in the gap equation in Eq. (2.18) becomes divergent. This is a pure artificial divergence since any physical interaction, no matter how short-range and strong it is, must die off at some finite momentum, which effectively avoids the problem of divergence in the gap equation. Therefore, if one wishes to solve the gap equation with the delta interaction without making an artificial cutoff for ultracold Fermi gases, an extra step of regularising the gap equation is needed, which we discuss in the following section.

Although Eq. (2.21) depends on the two model-dependent and unphysical quantities, U_0 and ω_D, we can take a step further and replace them with the physical s-wave scattering length a_s. In

dilute Fermi gases, particles are sufficiently far apart that the interaction between two particles at average separation is weak, even though the interaction can be orders of magnitude stronger when they are close to each other. Therefore, it is convenient to replace the bare interaction between particles by an effective interaction. This is done by a regularisation technique that integrates out the short-wavelength part of the bare interaction, leaving its long-wavelength component as the effective interaction [89]. Detailed discussions of this technique can be found in multiple references, such as Ref. [39, 89, 119, 128, 130, 135, 158, 174], and also in Appendix A. Here, we only provide a brief outline.

2.4.1 The Gap and The Chemical Potential in 3D

In three-dimensional space ($d = 3$), it is possible to relate the interaction, U_0, with the scattering length, which is given by (see Appendix A.3) [115]

$$\frac{m}{4\pi\hbar^2 a_s} = -\frac{1}{U_0} + \frac{1}{\Omega}\sum_{\mathbf{k}} \frac{1}{2\epsilon_{\mathbf{k}}}, \tag{2.25}$$

where $\epsilon_{\mathbf{k}} = \frac{\hbar^2 k^2}{2m}$. While the summation on the right hand side of Eq. (2.25) is divergent, it is possible to substitute it into the gap equation Eq. (2.18) and get rid of the term containing U_0. The gap equation Eq. (2.18) then becomes

$$-\frac{m}{4\pi\hbar^2 a_s} = \frac{1}{2\Omega}\sum_{\mathbf{k}'} \left(\frac{1}{\sqrt{\xi_{\mathbf{k}'}^2 + \Delta^2}} - \frac{1}{\epsilon_{\mathbf{k}'}} \right) \tag{2.26}$$

This summation is now finite due to the cancellation of the divergences. We now let the system size tend to infinity, so that $\sum_{\mathbf{k}'} \to \int \frac{\Omega}{(2\pi)^3} d\mathbf{k}$, and Eq. (2.26) becomes [101, 174]

$$\frac{1}{k_F a_s} = -\frac{2\pi\hbar^2}{m k_F} \int \frac{d\mathbf{k}}{(2\pi)^3} \left(\frac{1}{\sqrt{\xi_{\mathbf{k}}^2 + \Delta^2}} - \frac{1}{\epsilon_{\mathbf{k}}} \right) = -\frac{2}{\pi}\sqrt{\frac{\Delta}{\epsilon_F}} I_1\left(\frac{\mu}{\Delta}\right) \tag{2.27}$$

where $\epsilon_F = \frac{\hbar^2 k_F^2}{2m}$ is the Fermi energy. The integral function $I_1(z)$ is defined as

$$I_1(z) = \int_0^\infty dx\, x^2 \left(\frac{1}{\sqrt{(x^2 - z)^2 + 1}} - \frac{1}{x^2} \right) \tag{2.28}$$

which is finite for any real z. The number density in Eq. (2.17) can also be written as a function of the chemical potential and the gap

$$n = \int \frac{d\mathbf{k}}{(2\pi)^3} \left(1 - \frac{\xi_{\mathbf{k}}}{\sqrt{\xi_{\mathbf{k}}^2 + \Delta^2}} \right) = \frac{1}{2\pi^2} \left(\frac{2m\Delta}{\hbar^2} \right)^{3/2} I_2\left(\frac{\mu}{\Delta}\right) \tag{2.29}$$

where

$$I_2(z) = \int_0^\infty dx\ x^2 \left(1 - \frac{x^2 - z}{\sqrt{(x^2 - z)^2 + 1}}\right) \tag{2.30}$$

which is also finite for any real z. By replacing n with $k_F = (3\pi^2 n)^{1/3}$, we arrive at the following relation

$$1 = \frac{3}{2}\left(\frac{\Delta}{\epsilon_F}\right)^{3/2} I_2\left(\frac{\mu}{\Delta}\right). \tag{2.31}$$

Although the integrals $I_1(x)$ an $I_2(x)$ do not have closed-form expressions, they are related to the complete elliptic integrals and it is possible to find their asymptotic behaviours in various limits [101]. Here, we briefly discuss the behaviours of μ and Δ in both the BCS and BEC limits.

BCS Limit

In the BCS limit with $k_F a_s \to 0^-$ and $z \gg 1$, it is possible to reduce the two integrals in Eq. (2.27) and Eq. (2.31) into much simpler forms. A detailed analysis can be found in Marini *et al.* (1998) [101] and Zwierlein (2006) [174]. Here, we only quote the results we need from their work: as $z \to \infty$, we get

$$I_1(z) = \int_0^\infty dx\ x^2 \left(\frac{1}{\sqrt{(x^2 - z)^2 + 1}} - \frac{1}{x^2}\right) \sim \sqrt{z}[\ln(8z) - 2], \tag{2.32}$$

$$I_2(z) = \int_0^\infty dx\ x^2 \left(1 - \frac{x^2 - z}{\sqrt{(x^2 - z)^2 + 1}}\right) \sim \frac{2}{3} z^{3/2}. \tag{2.33}$$

From the second approximation, we get $\mu \approx \epsilon_F$ from Eq. (2.31), which is what we expect in the weak interaction limit. Hence, the superconducting gap as a function of the scattering length is given by

$$\Delta_{\text{BCS}} = \frac{8\epsilon_F}{e^2} \exp\left[\frac{\pi}{2k_F a_s}\right]. \tag{2.34}$$

While Eq. (2.34) is valid at the BCS level, Gorkov and Melik-Barkhudarov [63] improved this result further by taking the many-body effects into account and renormalising the scattering length $a_s \to a_s + \frac{2}{3\pi}\ln(4e)k_F a_s^2$. This scales the gap by a factor of $\frac{1}{(4e)^{1/3}}$ compared to Eq. (2.34) [3]

$$\Delta_{\text{Gorkov}} = \frac{1}{(4e)^{1/3}} \Delta_{\text{BCS}} = \left(\frac{2}{e}\right)^{7/3} \exp\left[\frac{\pi}{2k_F a_s}\right]. \tag{2.35}$$

[3] A detailed derivation of this factor can be found in Pethick and Smith (2002) [119].

BEC Limit

On the other hand, in the BEC limit where $k_F a_s \to 0^+$ and $|z| \gg 1$ with $z < 0$, the two integrals have another set of approximated values. As $z \to -\infty$, we get [101, 174]

$$I_1(z) = \int_0^\infty dx\ x^2 \left(\frac{1}{\sqrt{(x^2 - z)^2 + 1}} - \frac{1}{x^2} \right) \sim -\frac{\pi}{2}\sqrt{|z|}\left(1 + \frac{1}{16|z|^2}\right) \tag{2.36}$$

$$I_2(z) = \int_0^\infty dx\ x^2 \left(1 - \frac{x^2 - z}{\sqrt{(x^2 - z)^2 + 1}} \right) \sim \frac{\pi}{8\sqrt{|z|}}, \tag{2.37}$$

which gives

$$\mu = -\frac{\epsilon_F}{(k_F a_s)^2} + \frac{2\epsilon_F}{3\pi} k_F a_s = -\frac{\hbar^2}{2ma_s^2} + \frac{\pi\hbar^2 n}{m} a_s, \tag{2.38}$$

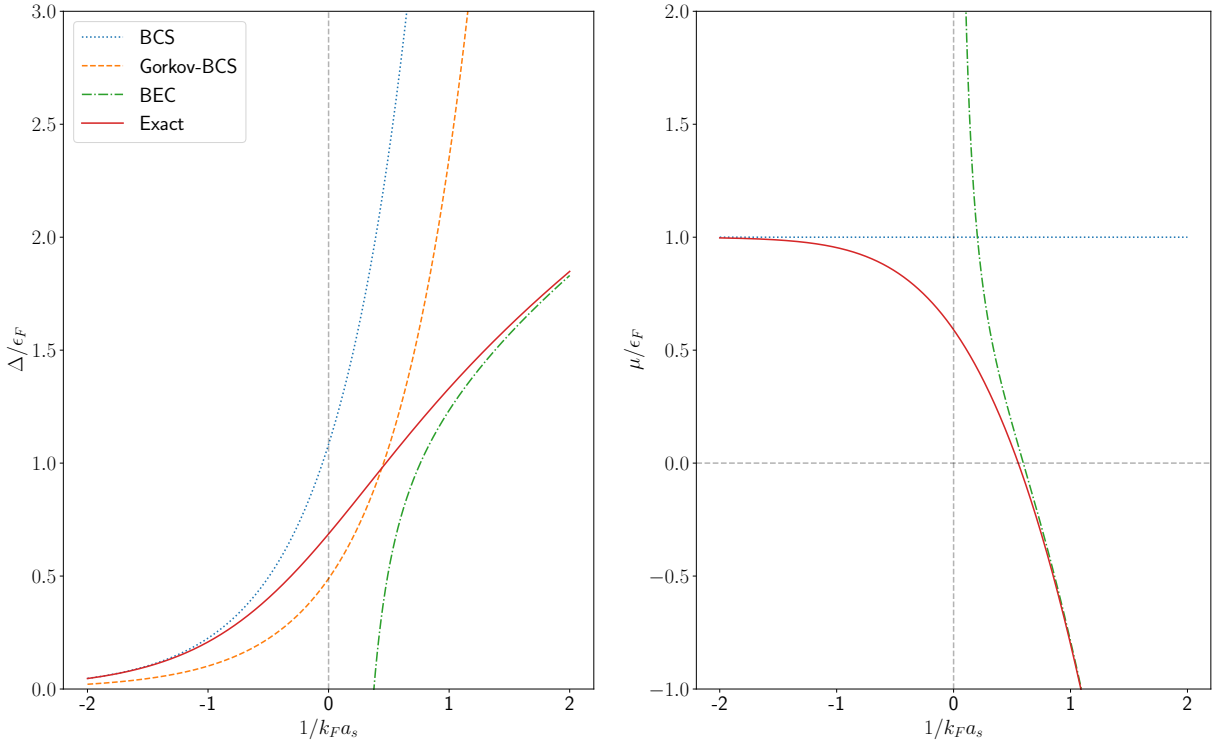
$$\Delta_{\text{BEC}} = \sqrt{\frac{16}{3\pi}} \frac{\epsilon_F}{\sqrt{k_F a_s}} - \frac{2\epsilon_F}{(3\pi)^{3/2}} (k_F a_s)^{5/2} = \frac{2\hbar^2}{m}\sqrt{\frac{\pi n}{a_s}} - \frac{(\pi n)^{3/2}\hbar^2}{m} a_s^{5/2}. \tag{2.39}$$

Note that Eq. (2.38) and Eq. (2.39) both provide two expressions for the quantity on the left-hand side. The first expression contains ϵ_F and k_F describes the result in a fermionic picture, while the second contains the number density n and particle mass m and describes the result in a bosonic picture. In the second expression for the chemical potential in Eq. (2.38), the first term corresponds to the binding energy per particle in a tightly bound pair of fermions, which can be viewed as a composite boson with mass $2m$ and radius a_s [110]. The second term, which coincides with the chemical potential in weakly *repulsive* Bose gas of mass $2m$ and density $n/2$, corresponds to the *effective repulsion* between the composite bosons in the gas. This effective repulsion among the composite bosons, which can be explained by the Popov's composite boson theory [124] with self-energy corrections, arises from the presence of the non-condensed density of the gas [151]. This is similar to the Gorkov and Melik-Barkhudarov corrections in the BCS case.

Plots of the gap and the chemical potential across the crossover can be found in Fig. (2.2). In Fig. (2.2a), the exact values of the gap, obtained by solving Eq. (2.27) and Eq. (2.31) numerically, as well as the asymptotic BCS and BEC expressions in Eq. (2.34), Eq. (2.35) and Eq. (2.39), are plotted for comparison. In Fig. (2.2b), the exact values of the chemical potential are compared with the asymptotic values in the BCS and BEC limits.

Note that although we have referred to Δ as the gap up to this point, the actual energy gap is not equal to Δ in the strongly interacting limit. One should keep in mind that the definition of the energy gap is the minimum energy needed to create a single-particle excitation, given by

$$E_{\text{gap}} = \min_{\epsilon_{\mathbf{k}} \geq 0} E_{\mathbf{k}} = \min_{\epsilon_{\mathbf{k}} \geq 0} (\sqrt{(\epsilon_{\mathbf{k}} - \mu)^2 + \Delta^2}) = \begin{cases} \Delta & \text{for} \quad \mu \geq 0 \\ \sqrt{\mu^2 + \Delta^2} & \text{for} \quad \mu < 0 \end{cases}. \tag{2.40}$$



(a) The gap Δ is plotted as a function of the inverse scattering length $1/k_F a_s$. The BCS (blue dotted) and BEC (green dash-dotted) asymptotes are given by Eq. (2.34) and Eq. (2.39), respectively. The red solid curve, labelled as "Exact", is obtained by solving the coupled nonlinear equations, Eq. (2.27) and Eq. (2.31). The BCS asymptote with Gorkov's correction (orange dashed) in Eq. (2.35) is also plotted.

(b) The chemical potential μ is plotted as a function of the inverse scattering length $1/k_F a_s$. The BCS (blue dotted) and BEC (green dash-dotted) asymptotes are given by $\mu = \epsilon_F$ and Eq. (2.38), respectively. The red solid curve, labelled as "Exact", is obtained by solving the coupled nonlinear equations, Eq. (2.27) and Eq. (2.31).

Figure 2.2: Comparisons various expressions for the gap Δ and the chemical potential μ as functions of the inverse scattering length $1/k_F a_s$ are plotted. The left hand side is the BCS region and the right hand side is the BEC region, with the unitarity limit at $k_F a_s = 0$. Note that while we have plotted the asymptotes for the whole range of $1/k_F a_s$, one should keep in mind that the BCS asymptotes are only valid as $1/k_F a_s \to -\infty$ and the BEC asymptotes are only valid as $1/k_F a_s \to \infty$.

On the weakly interacting BCS side, the energy gap coincides with the gap parameter Δ when the chemical potential is greater than zero, $\mu > 0$. As we increase the strength of the interaction and evolve to the BEC side, at some point the chemical potential changes sign, creating a subtle singularity in the energy gap (change in expression for the gap) near the BCS-BEC crossover. The energy gap in the region of negative chemical potential μ is thus given by the value of E_k when $\epsilon_k = 0$, where $E_{\text{gap}} = \sqrt{\mu^2 + \Delta^2}$ instead of Δ.

2.4.2 The Gap and The Chemical Potential in 2D

For two-dimensional space ($d = 2$), we have a similar relation between the bare interaction and the 2D scattering length (see Appendix A.3 for derivations) [130, 131, 164]

$$\frac{m}{2\pi\hbar^2}\left[\ln\left(\frac{ka_s}{2}\right)\bigg|_{k\to 0} + \gamma - \frac{i\pi}{2}\right] = \frac{1}{U_0} - \frac{1}{\Omega}\sum_{\mathbf{k}''}\frac{1}{2\epsilon_{\mathbf{k}''}}. \tag{2.41}$$

Combining with the BCS gap equation in Eq. (2.18), we have

$$\frac{m}{2\pi\hbar^2}\left[\ln\left(\frac{ka_s}{2}\right)\bigg|_{k\to 0} + \gamma - \frac{i\pi}{2}\right] = \frac{1}{2\Omega}\sum_{\mathbf{k}'}\left(\frac{1}{\sqrt{\xi_{\mathbf{k}'}^2 + \Delta^2}} - \frac{1}{\epsilon_{\mathbf{k}'}}\right). \tag{2.42}$$

Taking the limit of infinite system size, so that we can replace the sum by an integral, $\sum_{\mathbf{k}'} \to \int d\mathbf{k}\frac{\Omega}{(2\pi)^2}$, evaluating that integral and taking the exponential on both sides, we get

$$\sqrt{\mu^2 + \Delta^2} - \mu = \frac{4\hbar^2}{me^{2\gamma}a_s^2} = -\epsilon_b \tag{2.43}$$

where ϵ_b is the binding energy of a fermions pair. Note that the diverging $\ln(k)\big|_{k\to 0}$ term cancels exactly with the infra-red divergent term in the integral, resulting in a finite and well-behaved equation relating the chemical potential μ, the gap Δ and the 2D scattering length a_s. The number density of the Cooper pairs is

$$n = \int \frac{d\mathbf{k}}{(2\pi)^2}\left(1 - \frac{\xi_{\mathbf{k}}}{\sqrt{\xi_{\mathbf{k}}^2 + \Delta^2}}\right) = \frac{m}{2\pi\hbar^2}\left(\sqrt{\mu^2 + \Delta^2} + \mu\right). \tag{2.44}$$

Combining these two equations and solving for μ and Δ, we get [4]

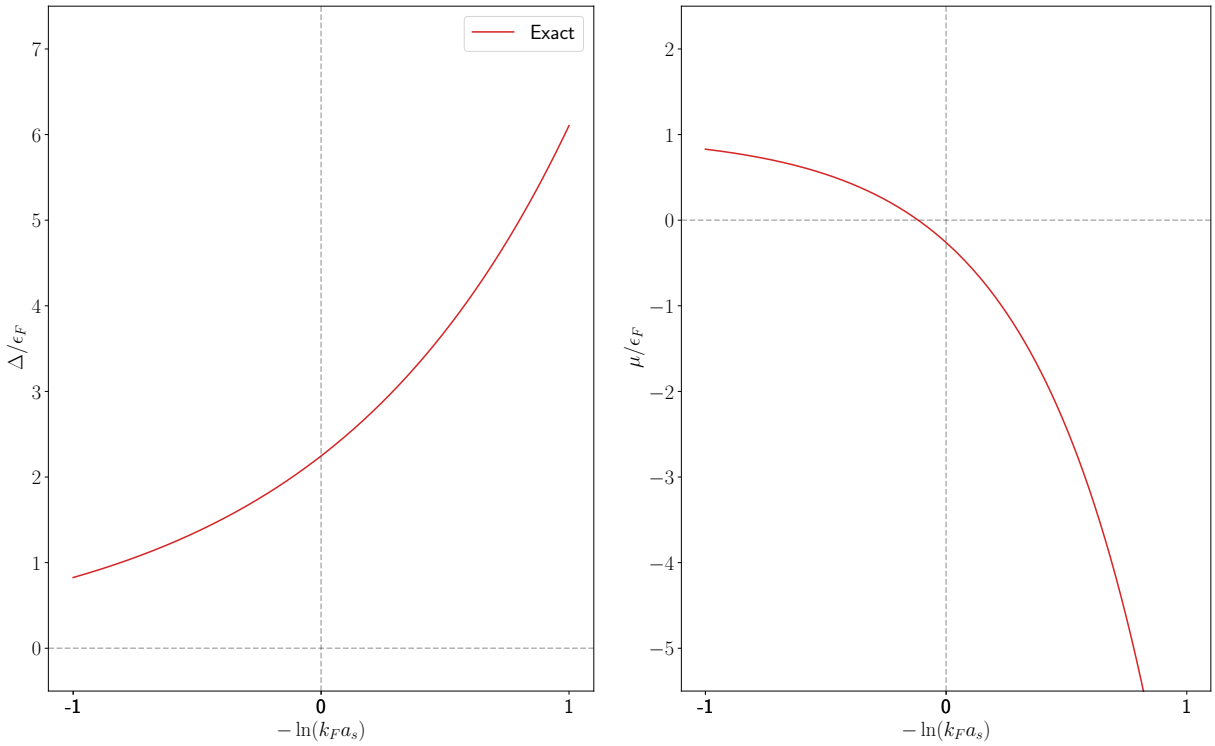
$$\mu = \epsilon_F - \frac{|\epsilon_b|}{2} = \epsilon_F\left[1 - \frac{4}{e^{2\gamma}(k_F a_s)^2}\right] \tag{2.45}$$

$$\Delta = \sqrt{2\epsilon_F|\epsilon_b|} = \frac{4\epsilon_F}{e^{\gamma}k_F a_s} \tag{2.46}$$

where $\epsilon_F = \frac{\hbar^2 k_F^2}{2m}$ is the Fermi energy and $k_F = \sqrt{2\pi n}$ is the Fermi wave vector in 2D. Compared to the 3D case, the 2D equivalent is analytically solvable and surprisingly simple, even though its scattering related quantities such as the scattering amplitude is infra-red divergent.

Plots of the gap and the chemical potential across the BCS-BEC crossover in 2D can be found in Fig. (2.3). The exact values of the gap from Eq. (2.46) and the chemical potential from Eq. (2.45) are plotted in Fig. (2.3a) and Fig. (2.3b) respectively.

[4] In some of the literature, the 2D scattering length is defined as $a_s' = e^{\gamma}a_s/2 \approx 0.89a_s$. In that case, the chemical potential has a simpler form $\mu = \epsilon_F[1 - (k_F a_s')^{-2}]$ and the gap becomes $\Delta = 2\epsilon_F(k_F a_s')^{-1}$.



(a) The gap Δ, as obtained from Eq. (2.46), as a function of the negative natural logarithm of the scattering length, $-\ln(k_F a_s)$.

(b) The chemical potential μ, obtained from Eq. (2.45), as a function of the negative natural logarithm of the scattering length, $-\ln(k_F a_s)$.

Figure 2.3: The gap Δ and the chemical potential μ as functions of the negative natural logarithm of the scattering length $-\ln(k_F a_s)$. The negative logarithm of $k_F a_s$ is a convenient choice of independent variable because it passes through zero in the BCS-BEC crossover regime, aiding characterisation; the negative sign ensures that the interaction strength increases from left to right across the graph (as was also the case in Fig. (2.2)).

Similar to the 3D case, the energy gap in 2D has the same definition as Eq. (2.40), where there is a subtle singularity (change in expression for the gap) when μ changes sign.

2.5 Spin-Imbalanced Fermi Gas

In a *spin-balanced*, *isotropic* Fermi gas with zero-range attractive interactions at $T = 0$, the ground state is always an s-wave superfluid with zero-momentum pairing [128, 142]. Although finite-momentum pairings are, in theory, possible in this system, they are not energetically favourable compared to zero-momentum pairing. On the other hand, in a spin-imbalanced Fermi gas, the Fermi surfaces for opposite spins differ and it is not obvious whether zero-momentum Cooper pairs can exist. Indeed, the Cooper instability does not occur in these systems unless the attractive interaction is sufficiently strong to overcome the normal state energy in the presence

of a spin imbalance. If the polarisation P, defined as

$$P = \frac{N^{\uparrow} - N^{\downarrow}}{N^{\uparrow} + N^{\downarrow}}, \quad N^{\uparrow} \geq N^{\downarrow}, \tag{2.47}$$

is increased further, it will reach a point where Cooper pairing is no longer energetically favourable, and the normal Fermi liquid state is recovered. The polarisation at which the superconductivity disappears is known as the *Clogston-Chandrasekhar limit*. The limiting value, which depends on the superconducting gap at zero polarisation, was first discussed independently and separately by Clogston [37] and Chandrasekhar [29] in the context of superconductors in a magnetic field. Before discussing this limit in the spin-imbalanced Fermi gas, we first present a brief discussion of superconductivity in an external magnetic field.

2.5.1 The Clogston-Chandrasekhar Limit

When an s-wave superconductor is placed in an external magnetic field, the superconductor creates an opposing field that expels the field lines from its interior. This phenomenon, known as the Meissner effect, was first proposed by Meissner and Ochsenfeld [104, 156]. If the field strength is increased further until it reaches some critical value, depending on the type of the superconductor, one of the following will occur. For type-I superconductors, superconductivity breaks down at a critical field strength H_{c1} and a phase transition from the superconducting state to the normal state takes place. For type-II superconductors, the superconductor enters the vortex state at the critical field strength H_{c1}, where it allows part of the external magnetic field to penetrate and couple with the orbital motion of the electrons, creating vortices. As the field strength is increased further, more vortices are created until the upper critical field H_{c2} ($> H_{c1}$) is reached, at which point the pairing fails to maintain the superconducting state and the superconductivity is destroyed [156]. This work will ignore the vortex state and focus on the type-I superconducting-to-normal phase transition.

The breaking of the Meissner effect can be explained by the gain in free energy density from the field, $\Delta u = \frac{B^2}{2\mu_0}$, exceeding the condensation energy in Eq. (2.24). However, this argument ignores the edge effects on the superconductor by the field, which become more apparent for thin superconductors. Indeed, the external magnetic field does not die off suddenly to zero at the surface of the superconductor. Instead, the field penetrates the surface for a very small but finite distance. If the material is sufficiently thin, it is possible for a strong magnetic field to penetrate it without exhibiting *any* Meissner effect. However, there still exists an upper limit on the critical field strength before superconductivity breaks down, which cannot be explained by the breaking of the Meissner effect, as the latter should have happened at a lower critical field.

It was Clogston who realised there is another mechanism that could explain the rise in critical field in the absence of the Meissner effect [37]. If an external magnetic field is applied to a spin-balanced system in the normal state, the system will adjust the population of the spin species to lower its free energy, making it spin-polarised. However, for systems in the superconducting

state, the same cannot be done because of the 2Δ energy cost for breaking a Cooper pair. Hence, if the lowering of the free energy of the normal state due to the Zeeman field exceeds the condensation energy, one would expect a superconducting-to-normal phase transition to occur. This posts another limit on the upper critical field for very thin superconductors that has no Meissner effects, and it is known as the *Clogston-Chandrasekhar limit* [29, 37], given by

$$\mu_B B_c = \frac{\Delta_0}{\sqrt{2}}, \tag{2.48}$$

where μ_B is the Bohr magneton and Δ_0 is the BCS superconducting gap at zero field.

For the scope of our work here, we will not discuss the consequences of applying magnetic fields to superconducting Fermi liquids, but focus instead on a somewhat artificial system: a spin-imbalanced Fermi gas without an applied field, which is analogous to the former. As such, we first define the following quantities

$$h = \frac{1}{2}(\mu_\uparrow - \mu_\downarrow), \qquad \bar{\mu} = \frac{1}{2}(\mu_\uparrow + \mu_\downarrow) \tag{2.49}$$

where h is the effective magnetic field that would in reality be required to create the spin imbalance, and $\bar{\mu}$ is the mean chemical potential. In the weakly interacting BCS limit, the chemical potential at zero polarisation, $P = 0$, roughly equals the Fermi energy, $\epsilon_F = \frac{\hbar^2 k_F^2}{2m}$, for both spin species, as we have seen in the previous section. Recalling that the ground-state energy of the superconducting state is the sum of the normal state energy and the condensation energy in Eq. (2.24). Now, in the presence of a *weak* magnetic field, $h \ll \bar{\mu}$, the chemical potentials μ_σ for the two spin species are shifted by σh, *i.e.*, $\mu_\sigma = \bar{\mu} + \sigma h$. Assuming the system is three dimensional ($d = 3$), using $\mu_\sigma \approx \epsilon_F^\sigma = \frac{\hbar^2 (k_F^\sigma)^2}{2m} = \frac{\hbar^2}{2m}\left(6\pi^2 \frac{N^\sigma}{\Omega}\right)^{2/3}$, the normal state free energy becomes

$$\begin{aligned} E_p &= \frac{3}{5}\mu_\uparrow N^\uparrow + \frac{3}{5}\mu_\downarrow N^\downarrow - \mu_\uparrow N^\uparrow - \mu_\downarrow N^\downarrow \\ &= -\frac{\Omega (2m\bar{\mu})^{3/2}\bar{\mu}}{15\pi^2\hbar^3}\left[\left(1 + \frac{h}{\bar{\mu}}\right)^{5/2} + \left(1 - \frac{h}{\bar{\mu}}\right)^{5/2}\right] \\ &\approx E_n - g_{3\mathrm{D}}(\epsilon_F) h^2 \qquad (h \ll \bar{\mu}) \end{aligned} \tag{2.50}$$

where we have assumed $\bar{\mu} \approx \epsilon_F$, *i.e.* the Fermi energy at zero spin imbalance $h = 0$, and E_n is the normal state free energy also at zero spin imbalance, and $g_{3D}(\epsilon)$ is the density of states of a 3D electron gas with Fermi energy $\bar{\mu}$. Finally, the Clogston-Chandrasekhar limit for the spin-imbalanced Fermi gas in the weakly interacting BCS limit can be obtained by equating the two energies, E_s and E_p , to get [29, 33, 37, 174]:

$$h_c^{\mathrm{CC}} = \frac{\Delta_0}{\sqrt{2}}. \tag{2.51}$$

Note that this expression is very similar to Eq. (2.48). Indeed, Eq. (2.48) can be derived by

setting $h = \mu_B B$ and carrying out the exact same free energy analysis as above. If we take one step further, we can find the critical polarisation P_c of the spin-imbalanced Fermi gas by replacing N^σ with $N^\sigma = \left(\frac{2m\mu_\sigma}{\hbar^2}\right)^{3/2} \frac{\Omega}{6\pi^2}$. This gives [174]:

$$P_c = \frac{\mu_\uparrow^{3/2} - \mu_\downarrow^{3/2}}{\mu_\uparrow^{3/2} + \mu_\downarrow^{3/2}} = \frac{\left(1 + \frac{h_c^{\mathrm{BCS}}}{\bar{\mu}}\right)^{3/2} - \left(1 - \frac{h_c^{\mathrm{BCS}}}{\bar{\mu}}\right)^{3/2}}{\left(1 + \frac{h_c^{\mathrm{BCS}}}{\bar{\mu}}\right)^{3/2} + \left(1 - \frac{h_c^{\mathrm{BCS}}}{\bar{\mu}}\right)^{3/2}} \approx \frac{3}{2}\frac{h_c^{\mathrm{BCS}}}{\bar{\mu}} = \frac{3}{2\sqrt{2}}\frac{\Delta_0^{3\mathrm{D}}}{\epsilon_F}. \tag{2.52}$$

As we will be looking at Fermi gases in two dimensions, we now carry out a similar analysis for $d = 2$. The whole procedure is analogous, with the use of the relation $\mu_\sigma \approx \epsilon_F^\sigma = \frac{\hbar^2 (k_F^\sigma)^2}{2m} = \frac{2\pi\hbar^2 N^\sigma}{m\Omega}$. The Zeeman energy and the condensation energy in 2D have the same form as Eq. (2.50) and Eq. (2.24), respectively, except the $g_{3\mathrm{D}}(\epsilon_F)$ is replaced by the 2D density of states $g_{2\mathrm{D}}(\epsilon_F)$. Hence, the expression for the Clogston-Chandrasekhar limit in Eq. (2.51) remains valid in 2D. The only thing that changes is the prefactor of the critical polarisation, which is now given by:

$$P_c = \frac{\mu_\uparrow - \mu_\downarrow}{\mu_\uparrow + \mu_\downarrow} = \frac{h_c^{\mathrm{BCS}}}{\bar{\mu}} = \frac{1}{\sqrt{2}}\frac{\Delta_0^{2\mathrm{D}}}{\epsilon_F}. \tag{2.53}$$

Note that Eq. (2.53) is only valid on the weakly interacting BCS side of the phase diagram, as we have assumed $\bar{\mu} \approx \epsilon_F$. Nonetheless, it provides a rough estimate of the location of the superfluid-to-normal phase transition in dilute spin-imbalanced Fermi gases.

As mentioned earlier, zero-momentum pairing is energetically favourable compare to finite-momentum pairing in unpolarised Fermi gases. But how about spin-imbalanced Fermi gases? It turns out that by simply adjusting the ratio of the numbers of spin-up and spin-down particles in a two-component Fermi gas is sufficient to drive the system towards exotic phases of matter beyond the standard BCS type s-wave superfluids. In the rest of this section, we provide a brief summary of the possible phases predicted to emerge in the spin-imbalanced Fermi gas.

2.5.2 The FFLO Phase

In this section, we discuss about the idea of finite-momentum pairing in Fermi gases and the possibility of hosting the *Fulde–Ferrell–Larkin–Ovchinnikov (FFLO or LOFF)* phase [50, 87]. First, we present a generalisation of the standard BCS theory to finite momentum pairing.

Consider the Hamiltonian in Eq. (2.3). In our earlier discussion of BCS theory, we assumed pairing has zero momentum and set $\mathbf{q} = 0$. This gives us a homogeneous gap in real space, *i.e.*, $\Delta(\mathbf{r}) = \Delta$. To generalise it to finite-momentum pairing, one can choose an arbitrary inhomogeneous gap function, which can be written in terms of its Fourier components

$$\Delta(\mathbf{r}) = \sum_{\mathbf{k}} \Delta_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{2.54}$$

where $\mathbf{r}$ is the spatial coordinate in real-space. However, the corresponding Schrödinger equation

is generally intractable. Instead, Fulde and Ferrell proposed the following simple Ansatz for the gap in real space that allows for finite momentum pairing within the BCS framework [50]

$$\Delta_{\mathbf{k}} = \delta_{\mathbf{kq}}\Delta_{\mathbf{q}} \quad \Rightarrow \quad \Delta_{\mathrm{FF}}(\mathbf{r}) = \Delta_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}} \tag{2.55}$$

where $\Delta_{\mathbf{q}} \in \mathbb{R}$ is a constant. The mean-field Hamiltonian corresponding to the FF Ansatz is:

$$\begin{aligned}
\hat{H} = \sum_{\mathbf{k}} \Big[& \xi_{\mathbf{k}+\mathbf{q}/2\uparrow}\hat{c}^{\dagger}_{\mathbf{k}+\mathbf{q}/2\uparrow}\hat{c}_{\mathbf{k}+\mathbf{q}/2\uparrow} + \xi_{-\mathbf{k}+\mathbf{q}/2\downarrow}\hat{c}^{\dagger}_{-\mathbf{k}+\mathbf{q}/2\downarrow}\hat{c}_{-\mathbf{k}+\mathbf{q}/2\downarrow} \\
& + \Delta_{\mathbf{q}}\hat{c}^{\dagger}_{\mathbf{k}+\mathbf{q}/2\uparrow}\hat{c}^{\dagger}_{-\mathbf{k}+\mathbf{q}/2\downarrow} + \Delta^{*}_{\mathbf{q}}\hat{c}_{-\mathbf{k}+\mathbf{q}/2\downarrow}\hat{c}_{\mathbf{k}+\mathbf{q}/2\uparrow} - \Delta_{\mathbf{q}}\left\langle \hat{c}^{\dagger}_{\mathbf{k}+\mathbf{q}/2\uparrow}\hat{c}^{\dagger}_{-\mathbf{k}+\mathbf{q}/2\downarrow}\right\rangle \Big] \\
= \sum_{\mathbf{k}} \Bigg[& \begin{pmatrix} \hat{c}^{\dagger}_{\mathbf{k}+\mathbf{q}/2\uparrow} & \hat{c}_{-\mathbf{k}+\mathbf{q}/2\downarrow} \end{pmatrix} \begin{pmatrix} \xi_{\mathbf{k}+\mathbf{q}/2\uparrow} & \Delta_{\mathbf{q}} \\ \Delta^{*}_{\mathbf{q}} & -\xi_{-\mathbf{k}+\mathbf{q}/2\downarrow} \end{pmatrix} \begin{pmatrix} \hat{c}_{\mathbf{k}+\mathbf{q}/2\uparrow} \\ \hat{c}^{\dagger}_{-\mathbf{k}+\mathbf{q}/2\downarrow} \end{pmatrix} \\
& + \left(\xi_{-\mathbf{k}+\mathbf{q}/2\downarrow} - \Delta_{\mathbf{q}}\left\langle \hat{c}^{\dagger}_{\mathbf{k}+\mathbf{q}/2\uparrow}\hat{c}^{\dagger}_{-\mathbf{k}+\mathbf{q}/2\downarrow}\right\rangle\right) \Bigg]
\end{aligned} \tag{2.56}$$

where we have shifted the sum in the kinetic energy by $\frac{\mathbf{q}}{2}$, and ignored any term that is not dependent on the operators. The superconducting gap $\Delta_{\mathbf{q}}$ is defined as

$$\Delta_{\mathbf{q}} = -\frac{U_0}{\Omega}\sum_{\mathbf{k}'}\left\langle \hat{c}_{-\mathbf{k}'+\mathbf{q}/2\downarrow}\hat{c}_{\mathbf{k}'+\mathbf{q}/2\uparrow}\right\rangle \tag{2.57}$$

where, again, we have used the delta attractive potential $U_{\mathbf{kk}'} = -U_0$. After diagonalising the Hamiltonian, one obtains the following dispersion relations:

$$\begin{aligned} E_{\mathbf{kq}\uparrow} &= \sqrt{\xi^{+2}_{\mathbf{kq}} + \Delta^2_{\mathbf{q}}} + \xi^{-}_{\mathbf{kq}} \\ E_{\mathbf{kq}\downarrow} &= \sqrt{\xi^{+2}_{\mathbf{kq}} + \Delta^2_{\mathbf{q}}} - \xi^{-}_{\mathbf{kq}} \end{aligned}, \qquad \xi^{\pm}_{\mathbf{kq}} = \frac{1}{2}(\xi_{\mathbf{k}+\mathbf{q}/2\uparrow} \pm \xi_{-\mathbf{k}+\mathbf{q}/2\downarrow}). \tag{2.58}$$

Note that, due to the existence of a finite momentum $\mathbf{q}$ in the Cooper pair, the single-particle energies of the two electrons in the pair are no longer equal and $\xi^{-}_{\mathbf{kq}} \neq 0$. In the thermodynamic limit, the dispersions are given by

$$\begin{aligned} \xi_{\mathbf{k}+\mathbf{q}/2\uparrow} &= \epsilon_{\mathbf{k}+\mathbf{q}/2\uparrow} - \mu_{\uparrow} = \frac{\hbar^2}{2m}\left(\mathbf{k} + \frac{\mathbf{q}}{2}\right)^2 - \bar{\mu} - h \\ \xi_{-\mathbf{k}+\mathbf{q}/2\downarrow} &= \epsilon_{-\mathbf{k}+\mathbf{q}/2\downarrow} - \mu_{\downarrow} = \frac{\hbar^2}{2m}\left(\mathbf{k} - \frac{\mathbf{q}}{2}\right)^2 - \bar{\mu} + h \end{aligned} \tag{2.59}$$

and the (anti)symmetric components, $\xi^{\pm}_{\mathbf{kq}}$, are given by

$$\xi^{+}_{\mathbf{kq}} = \frac{\hbar^2}{2m}\left(k^2 + \frac{q^2}{4}\right) - \bar{\mu}, \qquad \xi^{-}_{\mathbf{kq}} = \frac{\hbar^2}{2m}(\mathbf{k}\cdot\mathbf{q}) - h. \tag{2.60}$$

It follows that the symmetric component $\xi^{+}_{\mathbf{kq}}$ only depends on the magnitude of the momenta, while the antisymmetric component $\xi^{-}_{\mathbf{kq}}$ depends on both the magnitudes and the angle between momenta $\mathbf{k}$ and $\mathbf{q}$. For completeness, the expression for the finite-momentum BCS ground-state energy is:

$$\begin{aligned}\left\langle \hat{H} \right\rangle_{\mathbf{q}} &= \sum_{\mathbf{k}} \left[\xi_{-\mathbf{k}+\mathbf{q}/2\downarrow} - E_{\mathbf{kq}\downarrow} + \frac{\Delta_{\mathbf{q}}^2}{2\sqrt{\xi^{+2}_{\mathbf{kq}} + \Delta_{\mathbf{q}}^2}} \right] \\ &= \sum_{\mathbf{k}} \left[\xi^{+}_{\mathbf{kq}} - \sqrt{\xi^{+2}_{\mathbf{kq}} + \Delta_{\mathbf{q}}^2} + \frac{\Delta_{\mathbf{q}}^2}{2\sqrt{\xi^{+2}_{\mathbf{kq}} + \Delta_{\mathbf{q}}^2}} \right] .\end{aligned} \tag{2.61}$$

Using Eq. (2.57), the gap equation is found to resemble the BCS case in Eq. (2.18) [5]:

$$\Delta_{\mathbf{q}} = \frac{U_0}{\Omega} \sum_{\mathbf{k}'} \frac{\Delta_{\mathbf{q}}}{2\sqrt{\xi^{+2}_{\mathbf{k}'\mathbf{q}} + \Delta_{\mathbf{q}}^2}} \left[1 - \Theta(-E_{\mathbf{kq}\uparrow}) - \Theta(-E_{\mathbf{kq}\downarrow}) \right] , \tag{2.62}$$

and the number density is given by

$$n_{\mathbf{q}} = \frac{1}{\Omega} \sum_{\mathbf{k}'} \left[1 - \frac{\xi^{+}_{\mathbf{k}'\mathbf{q}}}{\sqrt{\xi^{+2}_{\mathbf{k}'\mathbf{q}} + \Delta_{\mathbf{q}}^2}} \left(1 - \Theta(-E_{\mathbf{kq}\uparrow}) - \Theta(-E_{\mathbf{kq}\downarrow}) \right) \right] , \tag{2.63}$$

where $\Theta(x)$ is the Heaviside step function. Now, because the quasiparticle energies $E_{\mathbf{kq}\sigma}$ can be negative for some combination of $\mathbf{q}$, $\bar{\mu}$ and h, the step functions cannot be set to zero. Hence, we need to consider different cases separately. By comparing the quasiparticle energies, one finds that there are two types of solutions: the BCS-type solutions and the FF-type solutions [154]. The BCS-type solution is obtained when both $E_{\mathbf{kq}\uparrow}$ and $E_{\mathbf{kq}\downarrow}$ are strictly greater than zero, whereas the FF solution corresponds to the case when one of the $E_{\mathbf{kq}\sigma}$ is negative in some regions of momentum space. Note that as $E_{\mathbf{kq}\uparrow} + E_{\mathbf{kq}\downarrow} \geq 0$, the two quasiparticle energies cannot be zero in the same region. The two cases will be discussed separately below.

BCS-type solutions

If $\frac{\hbar^2 \bar{k}_F q}{2m} + h < \Delta_{\mathbf{q}}$ where $\bar{k}_F = \sqrt{\frac{2m\bar{\mu}}{\hbar^2}}$ [6], both of $E_{\mathbf{kq}\sigma}$ are strictly positive, and hence the step functions can be set to zero. The solutions to the number equation and the gap equation are then the same as in the BCS case. In 2D, the solutions are

$$\bar{\mu} = \mu + \frac{\hbar^2 q^2}{8m}, \qquad \Delta_{\mathbf{q}} = \Delta \tag{2.64}$$

[5] Note that in contrast to the spin-balanced BCS case, the terms $\Theta(-E_{\mathbf{kq}\uparrow})$ and $\Theta(-E_{\mathbf{kq}\downarrow})$, which originate from the Fermi-Dirac statistics of the Bogoliubov quasiparticles, can no longer be set to zero as the energies $E_{\mathbf{kq}\sigma}$ of the quasiparticles are no longer strictly positive.

[6] This condition is actually an approximation valid when $\bar{k}_F \gg q$, which holds in the weakly-interacting BCS limit. The exact expression for this condition is much more complicated.

where μ and Δ are given by Eq. (2.45) and Eq. (2.46) respectively, and $q \equiv |\mathbf{q}|$ is the magnitude of the finite pair momentum. The ground-state wave function is given by the following finite momentum BCS-type wave function

$$|\Psi_{\mathbf{q}}\rangle = \prod_{\mathbf{k}} \left(u_{\mathbf{kq}} + v_{\mathbf{kq}} \hat{c}^{\dagger}_{\mathbf{k}+\mathbf{q}/2\uparrow} \hat{c}^{\dagger}_{-\mathbf{k}+\mathbf{q}/2\downarrow} \right) |0\rangle \,. \tag{2.65}$$

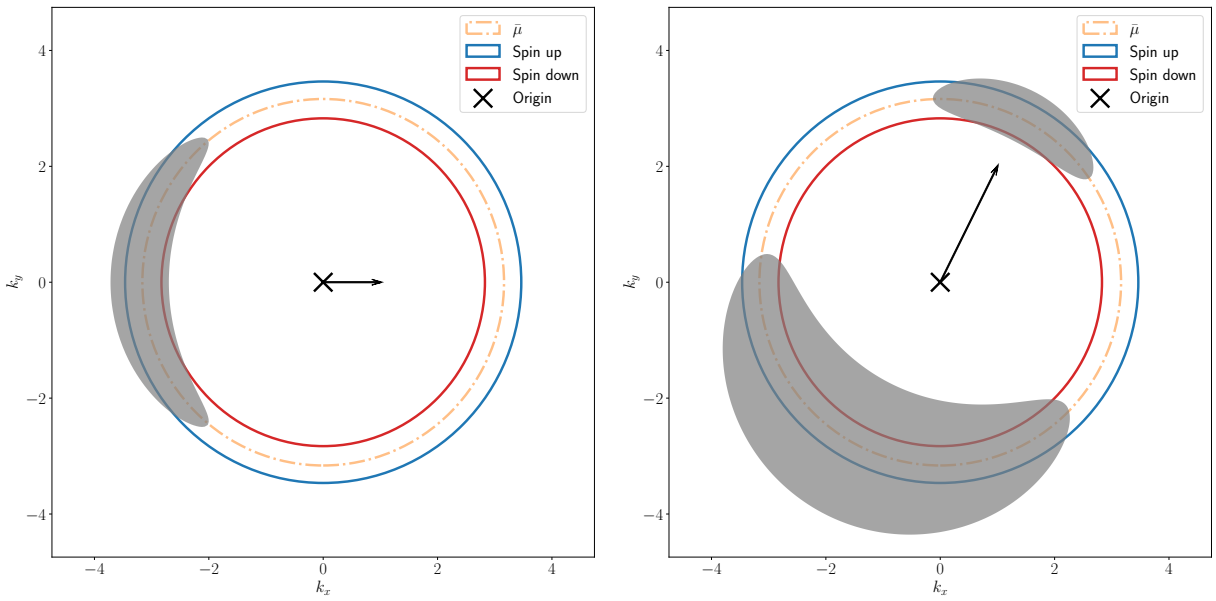
Note that the parameters $u_{\mathbf{kq}}$ and $v_{\mathbf{kq}}$ in Eq. (2.65) depend on the extra momentum $\mathbf{q}$:

$$|u_{\mathbf{kq}}|^2 = \frac{1}{2}\left(1 + \frac{\xi^{+}_{\mathbf{kq}}}{\sqrt{\xi^{+2}_{\mathbf{kq}} + \Delta^2_{\mathbf{q}}}}\right) \qquad |v_{\mathbf{kq}}|^2 = \frac{1}{2}\left(1 - \frac{\xi^{+}_{\mathbf{kq}}}{\sqrt{\xi^{+2}_{\mathbf{kq}} + \Delta^2_{\mathbf{q}}}}\right) . \tag{2.66}$$

If one sets $\mathbf{q} = 0$, the standard BCS solution is recovered provided that $h < \Delta_{\mathbf{q}}$.

FF-type solutions

On the other hand, if $\frac{\hbar^2 \bar{k}_F q}{2m} + h > \Delta_{\mathbf{q}}$, one of $E_{\mathbf{kq}\sigma}$ would be negative in some regions of momentum space. Hence, the summations over momentum space in Eq. (2.62) and Eq. (2.63) need to be split into two or more regions. The boundaries of these regions can be found by solving a fourth-order polynomial in $\mathbf{k}$. Examples can be found in Fig. (2.4).



(a) $q_x = 1$, $q_y = 0$ (b) $q_x = 1$, $q_y = 2$

Figure 2.4: The blocked regions of Cooper pairing in momentum space. The concentric circles, from inner to outer, indicate the Fermi surface for spin-down particles, the mean chemical potential $\bar{\mu}$, and the Fermi surface for spin-up particles, respectively. The shaded areas indicate regions where Cooper pairing is blocked. Here, we set $\bar{\mu} = 10$, $h = 2$ and $\Delta_{\mathbf{q}} = 4$, varying only the momentum $\mathbf{q}$.

In regions where $E_{\mathbf{kq}\sigma} > 0$ for both spins, the gap and the number density are the same as in the BCS case. In the blocked regions, where one of the $E_{\mathbf{kq}\sigma} < 0$, the gap vanishes and the number density is either one or zero, depending on which spin species is being blocked. The ground-state wave function is given by [154]

$$|\Psi_{\mathbf{q}}\rangle = \prod_{E_{\mathbf{kq}\sigma}<0} \hat{c}^{\dagger}_{\sigma\mathbf{k}+\mathbf{q}/2\sigma} \prod_{E_{\mathbf{kq}\uparrow}>0, E_{\mathbf{kq}\downarrow}>0} \left(u_{\mathbf{kq}} + v_{\mathbf{kq}} \hat{c}^{\dagger}_{\mathbf{k}+\mathbf{q}/2\uparrow} \hat{c}^{\dagger}_{-\mathbf{k}+\mathbf{q}/2\downarrow}\right) |0\rangle \,. \tag{2.67}$$

This is known as the *Fulde–Ferrell* state [154]. The exact expression for the superconducting gap under the FF-type solution is analytically intractable. The approximated solution to the gap equation in the BCS limit can be found in Ref. [154]. In 2D, it is possible to show that the critical chemical imbalance h_c is equal to the BCS gap at $h = 0$ [140, 159],

$$h_c^{\mathrm{FF}} = \Delta_0. \tag{2.68}$$

This suggest that the FF phase can survive beyond the Clogston-Chandrasekhar limit h_c^{CC}. The magnitude of the optimal pair momentum $q \equiv |\mathbf{q}|$ is given by $q = \frac{2hm}{\hbar^2 k_F}$, where k_F is the Fermi wave vector at zero imbalance, $h = 0$ [141, 159]. Note that in contrast to the 3D case, where the FF critical field is about $0.754\Delta_0$ [70, 154, 159], the FF region of the 2D phase diagram is much larger. This is due to a better nesting of Fermi surfaces, as well as the possibility of reducing the coupling between the orbital motion of the particles and the magnetic field by adjusting the direction of the external field [140, 159]. Hence, 2D systems serve as a better candidate for hosting phases with finite-momentum Cooper pairs.

In addition to the FF phase, Larkin and Ovchinnikov also proposed the LO Ansatz [87]

$$\Delta_{\mathrm{LO}}(\mathbf{r}) = \Delta_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}} + \Delta_{-\mathbf{q}} e^{i-\mathbf{q}\cdot\mathbf{r}} = 2\Delta_{\mathbf{q}} \cos(\mathbf{q}\cdot\mathbf{r}), \tag{2.69}$$

where we have assumed $\Delta_{\mathbf{q}} = \Delta_{-\mathbf{q}}$. In contrast to the FF phase, the LO phase allows the superconducting gap to be spatially varying in real space, with lines of zero (nodes) that are thought to host localised majority spins [98, 139, 159]. Despite being more energetically favourable than the FF Ansatz [87], the Hamiltonian with the LO Ansatz is not analytically solvable in the thermodynamic limit, as it is no longer possible to write the Hamiltonian as a sum of 2×2 matrices [5, 127]. Instead, the matrix dimension scales with the system size and can only be diagonalised using numerical methods. Additionally, it has been shown [141] that the most favourable FFLO-type state in the spin-imbalanced Fermi gas at $T = 0$ is an infinite linear combination of $\Delta_{\mathbf{q}}$ with equal magnitude, $|\mathbf{q}| = q_c$, *i.e.*, a ring of $\mathbf{q}$s with radius q_c.

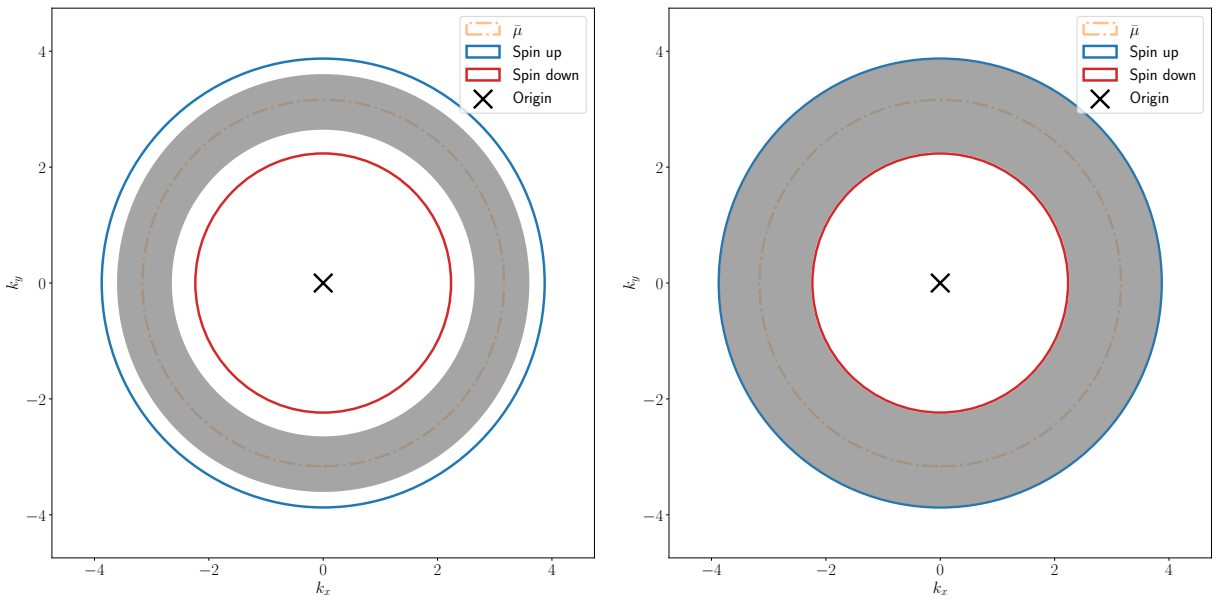
2.5.3 The Sarma Phase

Another possible phase of the spin-imbalanced Fermi gas is the Sarma phase, sometimes known as the breached-pair phase if there is a difference in the mass of the two species [13, 48, 70,

94, 137, 159]. The Sarma phase is a *homogeneous* superfluid with *gapless* fermionic excitations. As illustrated in Fig. (2.5), zero-momentum Cooper pairing occurs near the minority-spin Fermi surface (the red circle). The grey shaded region between the two Fermi surfaces (the red and the blue circles), given by $|\epsilon_{\mathbf{k}} - \bar{\mu}| < \sqrt{h^2 - \Delta_S^2}$, contains only unpaired majority spin fermions (or quasiparticles) with negative energies. The solutions to the number equation in Eq. (2.63) and the gap equation in Eq. (2.62) corresponding to this phase can be obtained by setting $\mathbf{q} = 0$ (homogeneous) and $h > \Delta_S$ where $\Delta_S \equiv \Delta_{\mathbf{q}=0}$. In the weakly interacting (BCS) limit, the gap Δ_S in the Sarma phase is [70, 137, 159]:

$$\Delta_S = \sqrt{\Delta_0(2h - \Delta_0)} \tag{2.70}$$

where Δ_0 is the standard BCS gap at $h = 0$. This equation holds as long as $\frac{\Delta_0}{2} < h < \Delta_0$ and $\Delta_S \ll \bar{\mu}$.



(a) $\Delta_S = 4$ (b) $\Delta_S = 1$

Figure 2.5: The blocked regions of Cooper pairing in momentum space for the Sarma phase. The concentric circles, from inner to outer, indicate the Fermi surface for spin-down particles, the mean chemical potential $\bar{\mu}$, and the Fermi surface for spin-up particles, respectively. The shaded areas indicate the regions of momentum space where Cooper pairing is blocked. Here, we set $\bar{\mu} = 10$, $\mathbf{q} = 0$ and $h = 5$, and only tune the gap $\Delta_{\mathbf{q}}$ to obtain different shaded regions.

While the Sarma state is a possible solution to the gap equation, it is *not energetically favourable* in the weakly interacting BCS limit. This can be seen by the differences between the Sarma state energy E_{Sarma} and the normal E_n, polarised E_p and unpolarised BCS state energies

E_{BCS}, respectively [70, 159]

$$E_{\mathrm{Sarma}} - E_n = g(\bar{\mu}) \left[\frac{1}{2}\Delta_0^2 - 2h\Delta_0 + h^2 \right] \tag{2.71}$$

$$E_{\mathrm{Sarma}} - E_p = \frac{1}{2} g(\bar{\mu}) \left(\Delta_0 - 2h\right)^2 \geq 0 \tag{2.72}$$

$$E_{\mathrm{Sarma}} - E_{\mathrm{BCS}} = g(\bar{\mu}) \left(\Delta_0 - h\right)^2 \geq 0, \tag{2.73}$$

i.e., the Sarma state energy is *always* higher than the polarised state energy and the unpolarised BCS state energy in the weakly interacting BCS limit.

In some cases, it is possible to stabilise the Sarma state. One way, proposed by Forbes *et al.* [48], is to use a finite range interaction in the model. The Sarma phase may then be stabilised by the momentum dependence of the pairing gap [70]. The Sarma phase may also exist in the strongly interacting BEC regime [70, 78, 99, 113, 146], although the picture of two nested Fermi surfaces with unpaired quasiparticles in between is no longer valid, as the chemical potential can be negative. As such, it should be regarded as a Bose-Fermi mixture with gapless fermionic excitations and boson-like dimers [155]. We investigate these possible phases in more detail in Chapter 6.

2.6 Anisotropic Fermi Gas

The spin-imbalanced Fermi gas illustrates one way to achieve spin splitting and allow exotic pairing mechanisms, but it is not the only way. Another approach is based on a novel type of magnetism, called *altermagnetism*.

2.6.1 Altermagnetism

Altermagnetism (AℓM) is a novel class of collinear magnetism that was discovered very recently by four different groups [69, 103, 145, 170]. Although altermagnets share some similarities with ferromagnets and antiferromagnets, they differ from both and their presence in nature was previously unrecognised. Like ferromagnetism, altermagnetism breaks time-reversal symmetry, allowing altermagnets to host the anomalous Hall effect. On the other hand, unlike ferromagnets, altermagnets exhibit *no net magnetisation* – a property shared with antiferromagnets. The name "altermagnet" was coined in 2022 by a group from the Johannes Gutenberg University of Mainz in Germany [144], who also pointed out that altermagnets had already been seen in experiments, although they had been "mislabelled" as antiferromagnets.

Altermagnetic materials are characterised by the combination of zero net magnetisation and a momentum-dependent spin splitting of their electronic band structures [143, 144]. The symmetries that connect the opposite-spin sublattices are non-trivial crystal-rotation symmetries, rather than the simpler symmetries, such as translations, found in antiferromagnets [43]. For

example, on a square lattice, the two spin bands can exhibit a d-wave symmetry. An illustration can be found in Fig. (2.6), where a ferromagnetic Fermi surface on the left is compared with an altermagnetic Fermi surface on the right. Band structures of this altermagnetic type provide a perfect ground provide the perfect ground for hosting exotic phases of matter, and offer promising applications in various research areas, including spintronics, ultra-fast optics, neuromorphics, thermoelectrics, field-effect electronics, multiferroics, high-T_c superconductors, and topological materials [144]. While the research landscape of altermagnets is broad, this thesis will specifically focus on the superconducting phases in altermagnetic metals. Existing mean-field [15, 18, 29, 38, 143] and diagrammatic approaches [171] predict phases with finite-momentum Cooper pairings and pair-density waves (PDW) [7], but there is very little work using *ab-initio* many-body simulation methods and no work employing neural wave function methods to investigate these phases more carefully. Hence, we aim to provide a preliminary study of superconductivity in altermagnet metals, using a simple model of an anisotropic Fermi gas to mimic the spin-split electronic bands of real altermagnets. In this section, we provide a brief introduction to the anisotropic Fermi gas model we will be using. In particular, we focus our attention on 2D systems.

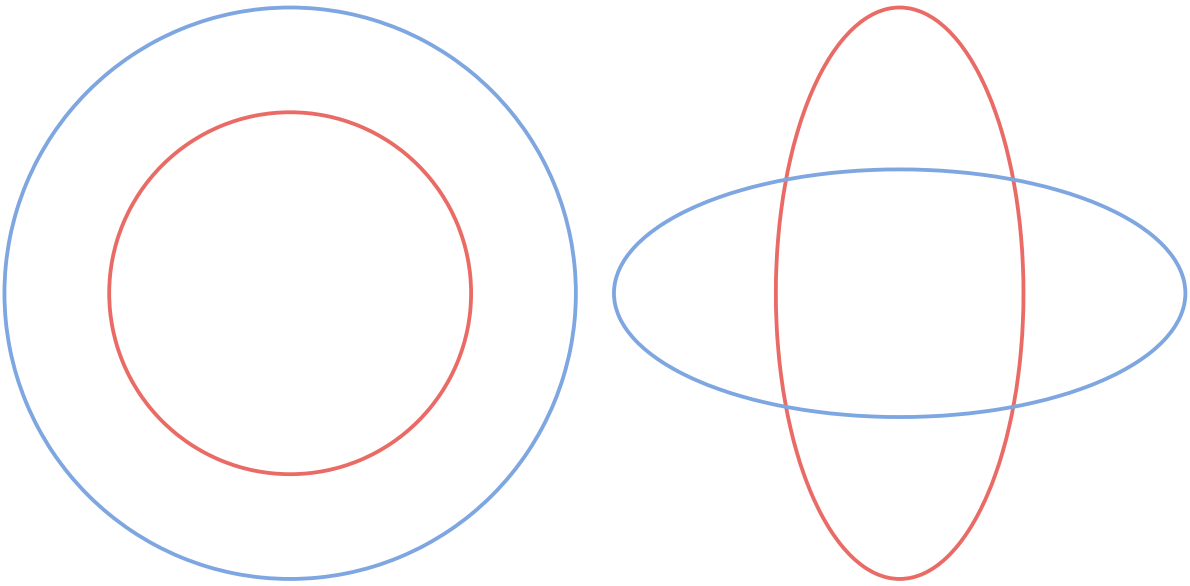


Figure 2.6: Fermi surfaces of the two spin species for ferromagnetism (left) and altermagnetism (right)

To mimic the spin-split bands in an altermagnet metal, the masses for the two spin species must be anisotropic and satisfy d-wave symmetry. For example, the single particle dispersion

[7] The use of the term "pair-density waves" instead of "FFLO phase" to describe finite-momentum Cooper pairing reflects both historical and conceptual considerations. The FFLO phase was originally proposed in the context of *spin-imbalanced systems*, whereas altermagnets are *spin-balanced*. In addition, the term "pair-density waves" more generally describe superconducting states with spatially modulated order parameters and do not rely on spin imbalance. Therefore, the term "pair-density waves" is more appropriate in the present context, even though both phenomena originate from finite-momentum Cooper pairing.

could be given by

$$\xi_{\mathbf{k}\sigma}(\delta) = \frac{\hbar^2}{2m}\left[\frac{k_x^2}{1+\sigma\delta} + \frac{k_y^2}{1-\sigma\delta}\right] - \mu_\sigma = \frac{\hbar^2}{2m(1-\delta^2)}\left[k^2 + \sigma\delta(k_y^2 - k_x^2)\right] - \mu_\sigma \tag{2.74}$$

where $\delta \in [0, 1)$ is the shift in effective mass, and $\sigma \in \{-1, +1\}$ is the spin. The effective masses are given by $m(1 \pm \delta)$, with the sign taken depending on both the spin channel and the direction in momentum space.. The non-interacting ground state for $\delta > 0$ and $h = 0$ is given by two ellipses with orthogonal major axes. A schematic diagram of the two Fermi surfaces can be found in the right-hand panel of Fig. (2.6), where one can see that the two Fermi surfaces are connected by a $\mathcal{C}_4^z$ rotation in momentum space.

At the mean-field level, the presence of an attraction between opposite-spin fermions has been found to induce multiple different phases [18, 42, 143]. For large anisotropies ($\delta \to 1$), the system is in a normal phase if the interaction is weak or a fully gapped BCS state if the interaction is strong. On the other hand, if the anisotropy is small ($\delta \to 0$), a fully gapped BCS phase is expected to exist regardless of the interaction strength. The interesting region where competing phases might occur is at intermediate anisotropies with moderate strength of interaction. Possibilities in this regime include the PDW phase, the gapless superfluid (G-SF) phase, and the Cooper-pair-Bose-metal (CPBM), also known as the d-wave Bose metal (DBM). For the PDW phase, the finite pair-momentum has a fourfold rotational symmetry due to the d-wave symmetry of the system, and mean-field calculations show finite-momentum pairing occurs at the vertices of the elliptical Fermi surfaces [143].

On the other hand, similar to the Sarma phase in the spin-imbalanced Fermi gas, the G-SF phase is characterised by the existence of gapless fermion excitations in the superfluid. At the BCS level, the blocked regions of pairing is given by non-positive regions of quasiparticle energies, $E_{\mathbf{k}\mathbf{q}\sigma}$. If we consider the case $\mathbf{q} = 0$, these regions can be determinated by the following equation [42]

$$\xi_{\mathbf{k}\uparrow}(\delta)\xi_{\mathbf{k}\downarrow}(\delta) = -\Delta^2, \tag{2.75}$$

where the dispersions, $\xi_{\mathbf{k}\sigma}(\delta)$, are defined in Eq. (2.74). Although the previous section argued that the Sarma phase is unstable in the weakly interacting limit of the spin-imbalanced Fermi gas, it is predicted to be stable within a large region of the phase diagram of the anisotropic Fermi gas [10, 16, 42].

The final option, the CPBM phase, hosts Cooper pairs but *no* condensate. In other words, the Cooper pairs behave like a metal of bosons, where gapless excitations exist at the Bose surface. Note that the CPBM cannot be understood within a picture of weakly interacting Fermionic quasiparticles in the Fermi liquid theory [18, 42, 106]. This phase is predicted to exist at moderate strengths of interaction and large anisotropy. We investigate these different possible phases in more details in Chapter 7.

2.7 Observables

In this section, we list and derive expressions for the most important observables computed in our numerical simulations. In particular, we discuss density matrices and how they are related to other relevant quantities such as the superfluid condensate fraction. We end by describing the mean-field results expected according to BCS theory.

2.7.1 Density Matrices and the Condensate Fraction

In addition to the formation of a superconducting gap, another signature of s-wave fermionic superfluidity is the existence of off-diagonal long range order (ODLRO) [168]. This can be seen in the non-vanishing order of the pair-pair terms in two-body density matrix as the pair separation tends to infinity. An analogous kind of ODLRO appears in the one-body density matrix (OBDM) of a superfluid Bose gas. Note that, in contrast to a superfluid Bose gas, there is no ODLRO in the OBDM for a fermionic superfluid. The OBDM and TBDM are defined in terms of the fermionic field operators via:

$$\rho^{(1)}_{\alpha}(\mathbf{r}'_1;\mathbf{r}_1) = \left\langle \hat{\psi}^{\dagger}_{\alpha}(\mathbf{r}'_1)\hat{\psi}_{\alpha}(\mathbf{r}_1)\right\rangle \tag{2.76}$$

$$\rho^{(2)}_{\alpha\beta}(\mathbf{r}'_1,\mathbf{r}'_2;\mathbf{r}_1,\mathbf{r}_2) = \left\langle \hat{\psi}^{\dagger}_{\alpha}(\mathbf{r}'_1)\hat{\psi}^{\dagger}_{\beta}(\mathbf{r}'_2)\hat{\psi}_{\beta}(\mathbf{r}_2)\hat{\psi}_{\alpha}(\mathbf{r}_1)\right\rangle \tag{2.77}$$

where α, β are the spin indices of particles 1 and 2, respectively. The normalisation is chosen such that

$$\sum_{\alpha}\int_{\Omega} d\mathbf{r}_1\ d\mathbf{r}'_1\ \delta(\mathbf{r}'_1-\mathbf{r}_1)\rho^{(1)}_{\alpha}(\mathbf{r}'_1;\mathbf{r}_1) = N \tag{2.78}$$

$$\sum_{\alpha}\sum_{\beta}\int_{\Omega} d\mathbf{r}_1\ d\mathbf{r}'_1\ d\mathbf{r}_2\ d\mathbf{r}'_2\ \delta(\mathbf{r}'_1-\mathbf{r}_1)\delta(\mathbf{r}'_2-\mathbf{r}_2)\rho^{(2)}_{\alpha\beta}(\mathbf{r}'_1,\mathbf{r}'_2;\mathbf{r}_1,\mathbf{r}_2) = N(N-1). \tag{2.79}$$

If the system is homogeneous, one can define the *translationally averaged* OBDM and TBDM as [107]

$$\bar{\rho}^{(1)}_{\alpha}(\mathbf{r}) = \frac{1}{\Omega}\int_{\Omega} d\mathbf{r}_1\ \rho^{(1)}_{\alpha}(\mathbf{r}_1+\mathbf{r};\mathbf{r}_1) \tag{2.80}$$

$$\bar{\rho}^{(2)}_{\alpha\beta}(\mathbf{r}) = \frac{1}{\Omega^2}\int_{\Omega} d\mathbf{r}_1\ d\mathbf{r}_2\ \rho^{(2)}_{\alpha\beta}(\mathbf{r}_1+\mathbf{r},\mathbf{r}_2+\mathbf{r};\mathbf{r}_1,\mathbf{r}_2) \tag{2.81}$$

where Ω is the volume of the system, and the bar at the top of the ρs denotes the translational average. If the system is also isotropic, one can rotationally average the density matrices, giving

[107]

$$\bar{\rho}_{\alpha}^{TR(1)}(r) = \frac{1}{S(r)\Omega} \int d\mathbf{r}' \; d\mathbf{r}_1 \; \delta(|\mathbf{r}'| - r)\rho_{\alpha}^{(1)}(\mathbf{r}_1 + \mathbf{r}'; \mathbf{r}_1) \tag{2.82}$$

$$\bar{\rho}_{\alpha\beta}^{TR(2)}(r) = \frac{1}{S(r)\Omega^2} \int d\mathbf{r}' \; d\mathbf{r}_1 \; d\mathbf{r}_2 \; \delta(|\mathbf{r}'| - r)\rho_{\alpha\beta}^{(2)}(\mathbf{r}_1 + \mathbf{r}', \mathbf{r}_2 + \mathbf{r}'; \mathbf{r}_1, \mathbf{r}_2) \tag{2.83}$$

where $S(r)$ is the surface area of the d-sphere [8], and TR denotes the translational and rotational average.

In an s-wave superfluid, pairing happens between *opposite spins*. Therefore, one needs to look at the opposite-spin TBDM, $\rho_{\uparrow\downarrow}^{(2)}$, the largest eigenvalue of which is proportional to the number of condensed fermion pairs in the system, to see whether the system is in a superfluid state. The condensate fraction, which defines the fraction of condensed pairs in the system, can be found from the translationally and rotationally averaged TBDM in Eq. (2.83) as

$$c_{\uparrow\downarrow} = \frac{\Omega^2}{N_{\text{pair}}} \lim_{r\to\infty} \bar{\rho}_{\uparrow\downarrow}^{TR(2)}(r) \tag{2.84}$$

where N_{pair} is the number of pairs of fermions in the system. In *spin-balanced* systems, $N_{\text{pair}} = N^{\uparrow} = N^{\downarrow}$. For *spin-imbalanced* or *partially spin-polarised* systems where $N^{\uparrow} \neq N^{\downarrow}$, the number of pairs defined by the equation $N_{\text{pair}} = \min(N^{\uparrow}, N^{\downarrow})$. A discussion on how to use QMC to accumulate these observables can be found in Appendix B.

2.7.2 Momentum Densities

When looking for exotic phases, such as the FFLO/PDW phase where spatial modulation is present in the TBDM, it is perhaps more insightful to look at the Fourier transformed density matrices, since the Fourier transformed versions will clearly mark any spatial modulation in real space. The Fourier transform of the translationally-averaged OBDM and TBDM are known as the one- and two-body momentum densities, respectively, and are given by:

$$n_{\mathbf{q}\alpha} = \left\langle \hat{c}_{\mathbf{q}\alpha}^{\dagger} \hat{c}_{\mathbf{q}\alpha} \right\rangle = \frac{1}{\Omega} \int d\mathbf{r} \; e^{i\mathbf{q}\cdot\mathbf{r}} \bar{\rho}_{\alpha}^{(1)}(\mathbf{r}) \tag{2.85}$$

$$f_{\mathbf{q}}^{\alpha\beta} = \sum_{\mathbf{k}} \left\langle \hat{c}_{\mathbf{k}\alpha}^{\dagger} \hat{c}_{\mathbf{q}-\mathbf{k}\beta}^{\dagger} \hat{c}_{\mathbf{q}-\mathbf{k}\beta} \hat{c}_{\mathbf{k}\alpha} \right\rangle = \frac{1}{\Omega} \int d\mathbf{r} \; e^{i\mathbf{q}\cdot\mathbf{r}} \bar{\rho}_{\alpha\beta}^{(2)}(\mathbf{r}). \tag{2.86}$$

Again, for an s-wave superfluid, apart from the definition in Eq. (2.84), we can also write the condensate fraction $c_{\uparrow\downarrow}$ in terms of $f_{\mathbf{q}}^{\uparrow\downarrow}$ as

$$c_{\uparrow\downarrow} = \frac{\Omega^2}{N_{\text{pair}}} \sum_{\mathbf{q}} f_{\mathbf{q}}^{\uparrow\downarrow} \lim_{r\to\infty} \frac{1}{S(r)} \int d\mathbf{r}' \; \delta(|\mathbf{r}'| - r) e^{-i\mathbf{q}\cdot\mathbf{r}'}. \tag{2.87}$$

[8] In $d = 3$, $S(r) = 4\pi r^2$.

By evaluating the integral and taking the limit as $r \to \infty$, we get

$$c_{\uparrow\downarrow} = \frac{\Omega^2}{N_{\text{pair}}} \sum_{\mathbf{q}} f_{\mathbf{q}}^{\uparrow\downarrow} \delta_{\mathbf{q},0} = \frac{\Omega^2}{N_{\text{pair}}} f_0^{\uparrow\downarrow} = \frac{\Omega^2}{N_{\text{pair}}} \sum_{\mathbf{k}} \left\langle \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} \hat{c}_{-\mathbf{k}\downarrow} \hat{c}_{\mathbf{k}\uparrow} \right\rangle, \tag{2.88}$$

which holds true for *any* spatial dimension d [9] . Hence, the quantity $f_0^{\uparrow\downarrow}$ can be interpreted as the number density of condensed Cooper pairs with *zero* centre-of-mass momentum.

While the asymptotic behaviour of the real-space TBDM allows one to identify any existence of ODLRO and the condensation of zero-momentum Cooper pairs, the whole TBDM in Eq. (2.81) can reveal the signs of exotic phases. In the following section, we will use the BCS wave function to compute $c_{\uparrow\downarrow}$.

2.7.3 Mean Field Results

The BCS wave function from Eq. (2.14) can be used to calculate the mean-field values of the observables we will later compute beyond the mean field using neural wave functions. The quantities that are relevant to us are the one- and two-body momentum densities from Eq. (2.85) and Eq. (2.86), and the condensate fraction from Eq. (2.84).

For the *spin-balanced, isotropic* Fermi gas, the BCS condensate fraction is given by

$$c_{\uparrow\downarrow} = \frac{\Delta^2}{2n\Omega} \sum_{\mathbf{k}} \frac{1}{\xi_{\mathbf{k}}^2 + \Delta^2} \tag{2.89}$$

where n is the number density of the system. By evaluating the integral, one can write the condensate fraction in terms of the gap Δ and the chemical potential μ, in both 2D and 3D

$$c_{\uparrow\downarrow}^{\text{2D}} = \frac{\Delta}{4\epsilon_F} \left[\frac{\pi}{2} + \tan^{-1}\left(\frac{\mu}{\Delta}\right) \right], \qquad c_{\uparrow\downarrow}^{\text{3D}} = \frac{3\pi}{8} \left(\frac{\Delta}{\epsilon_F}\right)^{3/2} \sqrt{\frac{1}{2}\left(\frac{\mu}{\Delta} + \sqrt{1 + \left(\frac{\mu}{\Delta}\right)^2}\right)}. \tag{2.90}$$

The pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow}$ in such a system is only non-zero at $\mathbf{q} = 0$, and its value $f_0^{\uparrow\downarrow}$ is

[9] The integral in Eq. (2.87) has a well-known form in terms of Bessel function of the first kind, which can be found, for example, in Ref. [56] and Ref. [150] as

$$\int d\mathbf{r}' \; \delta(|\mathbf{r}'| - r) e^{-i\mathbf{q}\cdot\mathbf{r}'} = (2\pi)^{\nu+1} r^{2\nu+1} \frac{J_\nu(|\mathbf{q}|r)}{(|\mathbf{q}|r)^\nu}$$

where $\nu = \frac{d}{2} - 1$ and d is the spatial dimension. With the following asymptotic behaviours of the Bessel function $J_\nu(x)$ [163]

$$\lim_{x\to\infty} J_\nu(x) = 0, \quad \lim_{x\to 0} \frac{J_\nu(x)}{x^\nu} = \frac{1}{2^\nu \Gamma(\nu+1)}, \quad x \in \mathbb{R}$$

and $S(r) = \frac{2\pi^{\nu+1} r^{2\nu+1}}{\Gamma(\nu+1)}$, we get the following identity

$$\lim_{r\to\infty} \frac{1}{S(r)} \int d\mathbf{r}' \; \delta(|\mathbf{r}'| - r) e^{-i\mathbf{q}\cdot\mathbf{r}'} = \delta_{\mathbf{q},0}$$

which holds true for *any* integer $d \geq 1$.

related to the condensate fraction by $f_0^{\uparrow\downarrow} = \frac{N_{\text{pair}}}{\Omega^2} c_{\uparrow\downarrow}$.

For the spin-imbalanced Fermi gas and the anisotropic Fermi gas, other exotic phases of matter, such as the FFLO phase, may emerge. The condensate fraction becomes ill-defined as there is no zero-momentum pairing in the system, making $f_0^{\uparrow\downarrow} = 0$, even though the system is still a superfluid. As such, we need to look at the pair-momentum densities $f_{\mathbf{q}}^{\uparrow\downarrow}$ at non-zero values of the momentum $\mathbf{q} \neq 0$.

Chapter 3

Neural Network Variational Monte Carlo

In this chapter, we discuss how to solve the many-body Schrödinger equation for fermionic systems using numerical methods. In particular, we focus on one of the quantum Monte Carlo (QMC) methods – the variational Monte Carlo (VMC) method – and discuss how deep neural networks can be used to overcome some of the shortcomings of the conventional VMC method.

3.1 The Many-Body Problem

The ground-state properties of a fermionic many-body system are described by the many-body wave function $\Psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$, governed by the Schrödinger equation, with $\mathbf{x}_i = (\mathbf{r}_i, \sigma_i)$ being the spatial position and the spin of the i-th particle, respectively. If the Hamiltonian is time-independent and we are only concerned about the energy eigenstates, the time-dependent part of the wave function becomes trivial and one only needs to solve the time-independent Schrödinger equation (TISE)

$$\hat{H}\Psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) = E\Psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) \tag{3.1}$$

to get the energy E and the corresponding energy eigenstates. In the non-relativistic limit, the Hamiltonian of a system of N interacting particles often takes the following general form [1]

$$\hat{H} = \sum_{i=1}^{N} \left[-\frac{\nabla_i^2}{2} + V(\mathbf{r}_i) \right] + \sum_{\substack{i=1 \\ i<j}}^{N} \sum_{j=1}^{N} U(\mathbf{r}_i - \mathbf{r}_j) \tag{3.2}$$

where the first term is the kinetic energy of the ith particle, $V(\mathbf{r}_i)$ is the potential acting on the ith particle, and $U(\mathbf{r}_i - \mathbf{r}_j)$ is the interaction between the ith particle and the jth particle. Due to the exponential scaling of Hilbert space with system size, this TISE is analytically intractable in

[1] In numerical simulations in quantum mechanics, it is convenient to use Hartree (atomic) units, where $\hbar = m_e = e = 4\pi\epsilon_0 = 1$.

general, except for a handful of simple systems such as the hydrogen atom. As a result, different approaches with different extents of approximations are often used to determine the qualitative and quantitative physics of interacting systems.

On the other hand, as an alternative route to solve the TISE by pen and paper, numerical methods are also employed to obtain approximations of physical quantities. Perhaps the most widely known and used one is the density-functional theory (DFT) [76, 85, 116], which transforms the problem from solving for the wave function into solving for the one-particle density

$$n(\mathbf{r}) = \int d\sigma_1 d\mathbf{x}_2 \cdots d\mathbf{x}_N |\Psi(\mathbf{x}, \mathbf{x}_2, \ldots, \mathbf{x}_N)|^2, \tag{3.3}$$

hence reducing the $d \times N$-dimensional problem into an d-dimensional problem. Observables, such as the energy, can be expressed as functionals of the one-particle density, and the system's ground state can be obtained by minimising the energy functional $E[n(\mathbf{r})]$. While the DFT is the most widely used numerical method to tackle physics and chemistry problems thanks to its computational efficiency, it is not a perfect method. For example, it is known to be less accurate at systems with weak, long-range interactions, such as the van der Waals force, or systems that are strongly correlated, such as in antiferromagnets and transition metal complexes. As such, other computational methods are usually preferred to study these systems.

Another class of numerical method, known as the quantum Monte Carlo (QMC) method, are also widely used to determine accurate quantitative results, especially at places where DFT are less accurate and reliable. They utilise the Markov-chain Monte Carlo (MCMC) sampling technique to get estimates of the $d \times N$-dimensional integrals that yield the expectation values of operators. In general, QMC methods are considered to be more accurate with fewer approximations, but at the same time much more computationally expensive than the DFT method. Nonetheless, they are still widely used, especially in strongly correlated systems where DFT fails.

3.2 Variational Monte Carlo

The variational Monte Carlo (VMC) method is conceptually the simplest of the many quantum Monte Carlo (QMC) methods. It is based on a well-known principle in quantum mechanics – *the variational principle* - which states that the expectation value of the energy of any wave function, Ψ_T, is always greater than or equal to E_0, the true ground-state energy of the system:

$$\left\langle \hat{H} \right\rangle = \frac{\langle \Psi_T | \hat{H} | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle} \geq E_0. \tag{3.4}$$

The equality holds if and only if the wave function Ψ_T is the true ground-state wave function of the system. This quantity contains $d \times N$-dimensional integrals that cannot be evaluated analytically, but can be estimated using numerical sampling methods.

In the VMC method, one starts with a trial wave function $\Psi_T(\mathbf{X}; \boldsymbol{\theta})$, which depends on

a set of N_p optimisable parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_{N_p})^T$, and the electron coordinates $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$. The energy expectation value of the system is then given by

$$E(\boldsymbol{\theta}) = \frac{\langle \Psi_T^{\boldsymbol{\theta}} | \hat{H} | \Psi_T^{\boldsymbol{\theta}} \rangle}{\langle \Psi_T^{\boldsymbol{\theta}} | \Psi_T^{\boldsymbol{\theta}} \rangle} = \frac{\int d\mathbf{X} \Psi_T^*(\mathbf{X};\boldsymbol{\theta}) \hat{H} \Psi_T(\mathbf{X};\boldsymbol{\theta})}{\int d\mathbf{X} |\Psi_T(\mathbf{X};\boldsymbol{\theta})|^2} \tag{3.5}$$

which can be minimised by optimising the parameters $\boldsymbol{\theta}$ in the trial wave function $\Psi_T(\mathbf{X};\boldsymbol{\theta})$.

The standard method to calculate these integrals is by using the *Metropolis-Hastings* algorithm [2]. It samples the configuration space of the system, $\mathbf{X}$, to generate a set of samples, $\{\mathbf{X}_i \mid i \in \mathbb{N}, i \leq M\}$, drawn from the probability density, $p(\mathbf{X};\boldsymbol{\theta}) = |\Psi_T(\mathbf{X};\boldsymbol{\theta})|^2$. Expectation values such as $E(\boldsymbol{\theta})$ can be approximated using the sampled points $\{\mathbf{X}_i\}$ as follows:

$$E(\boldsymbol{\theta}) = \frac{\int d\mathbf{X} \Psi_T^*(\mathbf{X};\boldsymbol{\theta}) \hat{H} \Psi_T(\mathbf{X};\boldsymbol{\theta})}{\int d\mathbf{X} |\Psi_T(\mathbf{X};\boldsymbol{\theta})|^2} \approx \frac{1}{M} \sum_{i=1}^{M} \frac{\hat{H} \Psi_T(\mathbf{X}_i;\boldsymbol{\theta})}{\Psi_T(\mathbf{X}_i;\boldsymbol{\theta})} \tag{3.6}$$

where $E_L = \frac{\hat{H}\Psi_T(\mathbf{X}_i;\boldsymbol{\theta})}{\Psi_T(\mathbf{X}_i;\boldsymbol{\theta})}$ is usually referred to as the *local energy*.

To optimise the wave function, one needs to optimise the variational parameters $\boldsymbol{\theta}$ to minimise the energy expectation value. This is generally done by gradient descent methods, where at each step, the parameters are updated by the following expression

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n - \alpha \nabla_{\boldsymbol{\theta}} E(\boldsymbol{\theta}) \tag{3.7}$$

where n is the iteration number, α is known as the learning rate, and $\nabla_{\boldsymbol{\theta}} E(\boldsymbol{\theta})$ is the gradient of the energy expectation value with respect to the variational parameters. The gradient $\nabla_{\boldsymbol{\theta}} E(\boldsymbol{\theta})$

[2] A detailed description of the Metropolis-Hastings algorithm will not be presented here, but can be found in various books, such as Ref. [9].

can be expressed as

$$
\begin{aligned}
&\nabla_{\boldsymbol{\theta}} E(\boldsymbol{\theta}) \\
&= \nabla_{\boldsymbol{\theta}} \left[\frac{\int d\mathbf{X} \Psi_T^{\boldsymbol{\theta}*} \hat{H} \Psi_T^{\boldsymbol{\theta}}}{\int d\mathbf{X} \left|\Psi_T^{\boldsymbol{\theta}}\right|^2} \right] \\
&= \frac{\int d\mathbf{X} \left|\Psi_T^{\boldsymbol{\theta}}\right|^2 \left(\frac{\nabla_{\boldsymbol{\theta}} \Psi_T^{\boldsymbol{\theta}*}}{\Psi_T^{\boldsymbol{\theta}*}} \frac{\hat{H}\Psi_T^{\boldsymbol{\theta}}}{\Psi_T^{\boldsymbol{\theta}}} + \frac{\nabla_{\boldsymbol{\theta}} \Psi_T^{\boldsymbol{\theta}}}{\Psi_T^{\boldsymbol{\theta}}} \frac{\hat{H}\Psi_T^{\boldsymbol{\theta}*}}{\Psi_T^{\boldsymbol{\theta}*}} \right)}{\int d\mathbf{X} \left|\Psi_T^{\boldsymbol{\theta}}\right|^2} - \frac{\int d\mathbf{X} \Psi_T^{\boldsymbol{\theta}*} \hat{H} \Psi_T^{\boldsymbol{\theta}}}{\int d\mathbf{X} \left|\Psi_T^{\boldsymbol{\theta}}\right|^2} \frac{\int d\mathbf{X} \left|\Psi_T^{\boldsymbol{\theta}}\right|^2 \left(\frac{\nabla_{\boldsymbol{\theta}} \Psi_T^{\boldsymbol{\theta}*}}{\Psi_T^{\boldsymbol{\theta}*}} + \frac{\nabla_{\boldsymbol{\theta}} \Psi_T^{\boldsymbol{\theta}}}{\Psi_T^{\boldsymbol{\theta}}} \right)}{\int d\mathbf{X} \left|\Psi_T^{\boldsymbol{\theta}}\right|^2} \\
&= \frac{\int d\mathbf{X} \left|\Psi_T^{\boldsymbol{\theta}}\right|^2 2 \operatorname{Re} \left[(E_L - E(\boldsymbol{\theta})) \nabla_{\boldsymbol{\theta}} \log \Psi_T^{\boldsymbol{\theta}} \right]}{\int d\mathbf{X} \left|\Psi_T^{\boldsymbol{\theta}}\right|^2} \\
&= 2 \operatorname{Re} \left[\mathbb{E}_{|\Psi_T|^2} \left[(E_L - E(\boldsymbol{\theta})) \nabla_{\boldsymbol{\theta}} \log \Psi_T^{\boldsymbol{\theta}} \right] \right] \\
&\approx \frac{1}{M} \sum_{i=1}^{M} 2 \operatorname{Re} \left[(E_L(\mathbf{X}_i) - E(\boldsymbol{\theta})) \nabla_{\boldsymbol{\theta}} \log \Psi_T(\mathbf{X}_i; \boldsymbol{\theta}) \right],
\end{aligned}
\tag{3.8}
$$

where $\Psi_T^{\boldsymbol{\theta}} \equiv \Psi_T(\mathbf{X}_i; \boldsymbol{\theta})$, and $\mathbb{E}_{|\Psi_T|^2}[\cdots] = \frac{\int d\mathbf{X} \left|\Psi_T^{\boldsymbol{\theta}}\right|^2 [\cdots]}{\int d\mathbf{X} \left|\Psi_T^{\boldsymbol{\theta}}\right|^2}$ denotes the expectation value of the quantity inside the squared brackets, $[\cdots]$, with respect to the probability density $|\Psi_T|^2$.

3.3 The Slater-Jastrow(-Backflow) Trial Wave Function

To describe a fermionic system, the Pauli exclusion principle requires that the wave function must be antisymmetric under the exchange of any two fermions. The most common and simplest method to achieve this is by using a Slater determinant (SD), which is the determinant of a matrix of single-particle orbitals $\phi(\mathbf{x})$:

$$
\Psi_{\mathrm{SD}}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) = \det \left[\phi_i(\mathbf{x}_j) \right] = \det \begin{pmatrix} \phi_1(\mathbf{x}_1) & \phi_1(\mathbf{x}_2) & \cdots & \phi_1(\mathbf{x}_N) \\ \phi_2(\mathbf{x}_1) & \phi_2(\mathbf{x}_2) & \cdots & \phi_2(\mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_N(\mathbf{x}_1) & \phi_N(\mathbf{x}_2) & \cdots & \phi_N(\mathbf{x}_N) \end{pmatrix}. \tag{3.9}
$$

This satisfies the fermionic exchange symmetry as swapping any pair of coordinates $(\mathbf{x}_i, \mathbf{x}_j) \to (\mathbf{x}_j, \mathbf{x}_i)$ is equivalent to swapping two columns of the matrix, resulting in a sign difference compared to the original wave function. However, a simple SD wave function does not capture the particle-particle correlations in interacting systems — it only gives the *Hartree-Fock* (mean-field) solution. In addition, if there are divergences in the interactions, as occur in the Coulomb interaction whenever charged particles approach each other, the SD wave function cannot satisfy the *cusp conditions*, which is required to avoid the divergences in the local energy [82]. An example of the cusp conditions can be seen in the wave function of a hydrogen atom, which is obtained

by solving the TISE:

$$\hat{H}\psi(\mathbf{r}) = -\frac{1}{2}\nabla^2\psi - \frac{\psi}{r} = -\frac{1}{2}\left[\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) + \frac{\hat{L}^2\psi}{r^2}\right] - \frac{\psi}{r} = E\psi. \tag{3.10}$$

If we consider the ground state, we can set $\psi(\mathbf{r}) \equiv \psi(r)$, *i.e.* the wave function has no angular dependence. This gives

$$\frac{\psi''}{\psi} + \frac{2}{r}\left(\frac{\psi'}{\psi} + 1\right) + 2E_0 = 0 \tag{3.11}$$

where $\psi' = \frac{d\psi}{dr}$ and $\psi'' = \frac{d^2\psi}{dr^2}$ are the first and second derivatives of the wave function with respect to r, respectively. Assuming the derivatives are finite as $r \to 0$, the divergences from the term containing $1/r$ must tends to zero in this limit, giving the hydrogen cusp conditions $\left[\frac{1}{\psi}\frac{d\psi}{dr}\right]_{r\to 0} = -1$. For a general many-body system with multiple electrons and nuclei, there exists both the electron-nucleus and the electron-electron cusp conditions, all of which need to be satisfied by the trial wave function [3].

To satisfy the cusp conditions and capture some of the correlations, one can multiply the SD by a so-called *Jastrow factor*, $e^{J(\mathbf{R})}$, where $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_N)$. The function $J(\mathbf{R})$ depends on both the electron-nucleus and electron-electron distances, and is chosen to satisfy the cusp conditions. The resulting *Slater-Jastrow wave function* takes the form:

$$\Psi_{\mathrm{SJ}}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) = e^{J(\mathbf{R})}\det[\phi_i(\mathbf{x}_j)]. \tag{3.12}$$

Note that the Jastrow factor must be a symmetric function of the electron coordinates, such that the wave function remains antisymmetric under the exchange of any two electrons.

3.3.1 The Backflow Transformation

Although the SJ wave function can capture some of the electron-electron correlations and satisfies the cusp conditions, the quality of the wave function can be further improved by introducing additional electron-electron correlations in the single-particle orbitals. A standard way of achieving this is by a coordinate transformation [4]

$$\mathbf{r}_i \to \mathbf{q}_i = \mathbf{r}_i + \sum_{\substack{j=1\\ j\neq i}}^{N} \eta(r_{ij})(\mathbf{r}_i - \mathbf{r}_j) \tag{3.13}$$

[3] For the scope of our work, we will not provide a thorough discussion of the cusp conditions and how to impose them using a Jastrow factor, as the wave functions of the Fermi gas systems we study do not have any cusps. A more detailed discussion on this matter can be found in Ref. [49].

[4] We have implicitly assumed there is only one nucleus at the origin. As discussed in the CASINO manual (see Ref. [107]), an analogous cusp condition applies to every electron-nucleus separation vector if there are more than one nucleus in the system.

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ is the electron-electron distance and $\eta(r_{ij})$ is a function of the electron-electron distance. This is known as the *backflow transformation* [44, 49]. The single-particle orbitals $\phi_i(\mathbf{x}_j)$ then become many-particle orbitals $\phi_i(\mathbf{q}_j, \sigma_j)$, which are invariant under exchange of any two particle indices excluding particle j. Together with the Jastrow factor, the wave function is then given by

$$\Psi_{\mathrm{SJB}}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) = e^{J(\mathbf{R})} \det[\phi_i(\mathbf{q}_j, \sigma_j)]. \tag{3.14}$$

which is known as a Slater-Jastrow-Backflow (SJB) wave function. This is the standard wave function used in most conventional continuum VMC and diffusion Monte Carlo (DMC) simulations.

3.3.2 Spin Independent Observables

In a many-body system with wave function $\Psi(\mathbf{X})$, the expectation value of any operator, $\hat{O}$, is evaluated as

$$\langle O \rangle = \frac{\langle \Psi | \hat{O} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \frac{\int d\mathbf{X} \Psi^*(\mathbf{X}) \hat{O} \Psi(\mathbf{X})}{\int d\mathbf{X} |\Psi(\mathbf{X})|^2} = \frac{\sum_{\boldsymbol{\sigma}} \int d\mathbf{R} \Psi^*(\mathbf{R}, \boldsymbol{\sigma}) \hat{O} \Psi(\mathbf{R}, \boldsymbol{\sigma})}{\sum_{\boldsymbol{\sigma}} \int d\mathbf{R} |\Psi(\mathbf{R}, \boldsymbol{\sigma})|^2} \tag{3.15}$$

where we have separated the summation over spins and the spatial integral in the last equality. If we assume the operator, $\hat{O}$, to be independent of the spin configuration, $\boldsymbol{\sigma}$, then every spatial integral in the summation over spins gives the same value. Hence, if all of the operators of interest are independent of spin, it is convenient to choose a single spin configuration and omit the sums over $\boldsymbol{\sigma}$. In a system with $N = N^{\uparrow} + N_{\downarrow}$ particles, the standard choice is to let the first $N^{\uparrow}$ particles be spin-up and the remaining $N^{\downarrow} = N - N^{\uparrow}$ particles be spin-down, *i.e.*, $\{(\mathbf{r}_1, \sigma_1), \ldots, (\mathbf{r}_N, \sigma_N)\} \to \{(\mathbf{r}_1, \uparrow), \ldots, (\mathbf{r}_{N^{\uparrow}}, \uparrow), (\mathbf{r}_{N^{\uparrow}+1}, \downarrow), \ldots, (\mathbf{r}_N, \downarrow)\}$.

In terms of the SJ/SJB wave functions, one can adapts this choice of spin configuration and further assumes every spin-orbital, $\phi_i(\mathbf{x}_j)$, is either a spin-up orbital, $\phi_i^{\uparrow}(\mathbf{x}_j)$, or a spin-down orbital, $\phi_i^{\downarrow}(\mathbf{x}_j)$, such that $\phi_i^{\uparrow}(\mathbf{r}_j, \downarrow) = \phi_i^{\downarrow}(\mathbf{r}_j, \uparrow) = 0$. Under these settings, the $N \times N$ Slater determinant can be replaced by a product of two determinants, one with dimension $N^{\uparrow} \times N^{\uparrow}$ and the other one with dimension $N^{\downarrow} \times N^{\downarrow}$ [49, 157] [5]

$$\det[\phi_i(\mathbf{x}_j)] \to \det\Big[\phi_i^{\uparrow}(\mathbf{r}_j, \uparrow)\Big] \det\Big[\phi_i^{\downarrow}(\mathbf{r}_j, \downarrow)\Big]. \tag{3.16}$$

Note that either of the two determinants (the full one and the spin-factored one) will give the same expectation value of any spin-independent operator, but the latter is more computationally efficient, as no spin sampling is required. Hence, for applications of QMC to physical and chemical systems, spin-assigned wave functions are used.

[5] Note that the two notations $(\mathbf{r}_i, \sigma_i)$ and $\mathbf{r}_i^{\sigma_i}$ will be used interchangeably in this thesis.

3.3.3 Multi-determinant Wave Functions

Another way to improve the quality of the wave function is by using a linear combination of Slater determinants, known as a multi-determinant expansion. The exchange (anti)symmetry is not affected, since each Slater determinant is still antisymmetric under the exchange of any two particles. A multi-determinant SJ wave function takes the form

$$\Psi_{SJ}^{D}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) = e^{J(\mathbf{R})} \sum_{k=1}^{D} \omega_k \det\Big[\phi_i^{k\uparrow}(\mathbf{r}_j, \uparrow)\Big] \det\Big[\phi_i^{k\downarrow}(\mathbf{r}_j, \downarrow)\Big], \tag{3.17}$$

where ω_k are the weights of the determinants and D is the number of determinants. Multi-determinant SJ/SJB trial wave functions are often used to achieve higher accuracies in chemical systems as multiple determinants can capture more electron-electron correlation than a single determinant.

3.4 Neural Network Wave Functions

Despite its successes, the flexibility of a SJB wave function is still limited by its functional form, which is in turn restricted by the system-dependent basis set that can be a bottleneck when it comes to the prediction of novel materials and phases of matter. Recently, a novel type of wave function Ansatz, known as the *neural wave function* or *neural network quantum state*, has emerged. The neural wave function is not a choice of functional form but is represented by a parameterised neural network that respects the exchange symmetry of the system. In this section, we first explain the basics of a neural network, and how the idea of a neural wave function originated.

3.4.1 Neural Networks

As the heart of artificial intelligence (AI) and machine learning (ML), an artificial neural network (ANN), also known as a neural network, is a computational model that is inspired by the biological neural networks in animal brains. Similar to an animal brain, the smallest unit in a neural network is a "neuron", also known as a *perceptron*, which is essentially a floating point number. The perceptrons are connected to each other with *edges*, and information is passed via the edges between perceptrons.

A prominent example of a neural network is the feed-forward neural network, or multi-layer perceptron (MLP). Like all ANNs, an MLP is in effect a very general parameterised function that maps vectors of input values to vectors of output values [61]. It usually consists of a number of layers, with the first layer being the *input layer* and the final layer being the *output layer*. The layers between the input and output layers are known as *hidden layers*. Information is passed, layer by layer, from the input layer to the output layer through the hidden layers, hence the

name "feed-forward". A schematic diagram of a simple MLP is shown in Fig. (3.1), where each coloured circle represent a perceptron, connected to perceptrons in different layers.

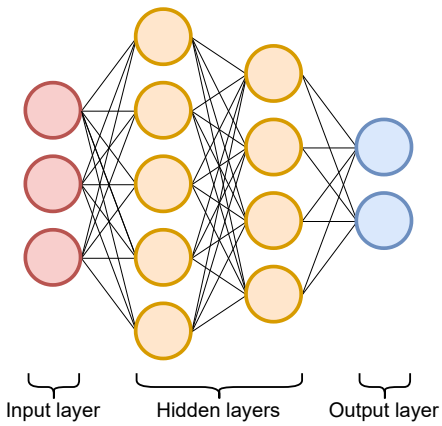


Figure 3.1: A schematic diagram for a simple multi-layer perceptron neural network. It is consisted of an input layer (red), two hidden layers (orange), and an output layer (blue). Each layer contain different number of perceptrons and the neighbouring layers are fully connected by the edges.

Consider a neural network with an input and output layer, and L hidden layers. The sets of neurons in the input and output layers are denoted by $\mathbf{x} \in \mathbb{R}^{n_I}$ and $\mathbf{y} \in \mathbb{R}^{n_O}$, respectively, whereas the neurons in the hidden layers are denoted as $\mathbf{h}^l \in \mathbb{R}^{n_l}$, where n_l is the number of hidden units in that layer and $l \in \{1, \ldots, L\}$. The propagation of the vectors from each layer l to the next layer $l+1$ is given by

$$\mathbf{h}^{l+1} = \sigma\left(W^l \mathbf{h}^l + \mathbf{b}^l\right) \tag{3.18}$$

where $W^l \in \mathbb{R}^{n_{l+1} \times n_l}$ and $\mathbf{b}^l \in \mathbb{R}^{n_l}$ are known as the *weights* and *biases* of the neural network. The function σ is non-linear and acts element-wise on its vector inputs, *e.g.*, $\sigma([a\ b]^T) = [\sigma(a)\ \sigma(b)]^T$. The output is produced in a similar fashion, but without the application of the non-linear function σ after the linear projection,

$$\mathbf{y} = W^L \mathbf{h}^L + \mathbf{b}^L. \tag{3.19}$$

Although MLP networks were invented more than half a century ago, they only became widely used in the last decade. Due to its complex nature, computing the output of an MLP and its gradient for back-propagation for a given input vector requires substantial computational resources. Specifically, it requires fast computation of matrix multiplication and storage of large matrices. Neither was readily available at a large scale until the recent advancements in graphics processing units (GPU). As GPUs were originally designed to carry out large amounts of low-precision parallel computation, as required in video games, they can also be used to compute

and store matrices in a neural network. Hence, the efficiency with which they handle linear algebra, and the development of low-level GPU programming languages such as CUDA, have allowed GPUs to be used to power the usage of ANNs in technologies like speech recognition, image classification, and natural language processing. They are now widely used in areas such as medical imaging, self-driving cars and predictions in stock markets. GPU-based neural-network approaches have also permeated to the different areas in natural sciences, where they are used for everything from discovering new materials to designing new chemical syntheses.

As neural networks are *universal function approximators*, know to be able to represent *any* Lebesgue-integrable function [77], there is no doubt that they can also be used to represent the wave functions of quantum systems. The first application of neural networks in VMC calculations was by Troyer and Carleo [19], who used restricted Boltzmann machine (RBM) – a statistical-mechanics-inspired neural network originally designed to model probability distributions – to study one and two-dimensional spin systems, obtaining state-of-the-art results. This work marks the emergence of an important new computational approach to quantum condensed matter physics.

Neural networks used to represent wave functions are known as "neural wave functions" (NWF) or "neural quantum states" (NQS). There is no rigorous distinction between the two and the names are often used interchangeably in much of the literature. The term NQS (or, equivalently, neural-network quantum states, NNQS) was invented by Troyer and Carleo [19], and is most often used to refer to neural Ansätze for *discrete* systems such as spin and lattice models. The term NWF was proposed later [53, 54] and is mostly used to refer to neural Ansätze for the wave functions of particles moving in continuous Hilbert spaces. Examples are electrons in molecules or Fermi gases. Although the different is subtle, the communities using NQS/NWF to study discrete and continuous systems are largely separate, despite their common roots. Before discussing NWF in continuous system, we first give a brief introduction to NQS.

3.4.2 Neural Network Quantum States

A neural quantum state is a neural network representation of the wave function of a quantum many-body system. The network takes an array of inputs that represent the configuration of the system, such as arrangement of spins, locations of electrons, etc., and returns a real or complex number that gives the corresponding value of the wave function. Consider a quantum many-body system with N degrees of freedom $\mathcal{X} \equiv (\mathcal{X}_1, \ldots, \mathcal{X}_N)$ in a chosen basis set. The wave function of the system can be written in the chosen basis set as

$$|\Psi\rangle = \sum_{\{\mathcal{X}\}} \Psi(\mathcal{X}) \, |\mathcal{X}\rangle \tag{3.20}$$

and the quantity $\Psi(\mathcal{X})$ can be represented with a NQS, denoted as $\Psi_{\text{NQS}}(\mathcal{X}; \mathcal{W})$, where $\mathcal{W} \equiv (\mathcal{W}_1, \ldots, \mathcal{W}_{N_p})$ is a set of N_p variational parameters – the network weights and biases – that can

be optimised under the VMC framework for a given Hamiltonian $\mathcal{H}$.

The actual choice of the neural network can vary depending on the system. Apart from RBM and MLP, convolutional neural network (CNN), recurrent neural networks (RNN), and the transformer networks have been applied to study various systems with great successes [19, 36, 75, 96, 117, 161].

3.4.3 Neural Wave Functions in Continuous Systems

The first fermionic neural wave function for QMC simulations in continuous space, called the DeepWF, was proposed by Han *et al.* in 2018 [68]. The DeepWF uses the Vandermonde-like determinant to construct a fully antisymmetric fermionic wave function instead of using a Slater determinant. While the evaluation of a Vandermonde-like determinant scales as $\mathcal{O}(N^2)$, in contrast to the $\mathcal{O}(N^3)$ scaling for the evaluation of Slater determinants, it is believed to be the reason of getting poor accuracies in certain quantum chemistry tasks [74]. Shortly after the proposal of DeepWF, two different fermionic NWFs were proposed at similar times. One, the "PauliNet", was introduced by Hermann *et al.* [73], and the other, the "FermiNet", by Pfau *et al.* [121]. Although the details of the implementations of the two are different, their main idea is the same, which is to generalise the many-particle functions $\phi_i(\mathbf{x}_j)$ in the SJB wave function, replacing them by neural networks that depend on the positions of all of the electrons, while maintaining the permutation-invariance property of the function. In this section, we discuss the architecture of the latter one (FermiNet) as it is used throughout our work.

The fermionic neural network, or FermiNet, is a NWF that was initially designed for describing electrons in chemical systems under the Born-Oppenheimer (BO) approximation [121, 149], but was later extended to study positron-molecule complexes [23] and circular quantum dots [40], as well as periodic systems such as the homogeneous electron gas (HEG) [24] and the unitary Fermi gas [95]. While each of these works focused on vastly different systems with their own modifications of FermiNet, the basic architecture of the network is unaltered. We will discuss the original architecture below, and then discuss how it can be extended to other systems.

Consider a chemical system of $N = N^{\uparrow} + N^{\downarrow}$ electrons with N_A spatially-fixed nuclei, each with positions $\{\mathbf{R}_I\}$ with $I \in \{1, 2, \ldots, N_A\}$. By taking electron positions relative to the nuclei, we get a set of spatial coordinates of the particles $\{\mathbf{r}_i^{\alpha} - \mathbf{R}_I\}$ with $i \in \{1, 2, \ldots, N^{\alpha}\}$, $I \in \{1, 2, \ldots, N_A\}$, and $\alpha \in \{\uparrow, \downarrow\}$. Similarly, we define a set of inter-electron coordinates, $\{\mathbf{r}_i^{\alpha} - \mathbf{r}_j^{\beta}\}$ with $i \in \{1, 2, \ldots, N^{\alpha}\}$, $j \in \{1, 2, \ldots, N^{\beta}\}$, and $\alpha, \beta \in \{\uparrow, \downarrow\}$. Assuming $d = 3$, for each pair of particles (electron and electron or electron and nucleus), the three components of the relative position vector and the length of that vector provide four inputs to the neural network. Thus, the $4N_A$ electron-nuclear inputs associated with the position of electron i with respect to to the N_A nuclei are

$$\mathbf{h}_i^{0\alpha} = \bigoplus_{I=1}^{N_A} \left(\mathbf{r}_i^{\alpha} - \mathbf{R}_I, |\mathbf{r}_i^{\alpha} - \mathbf{R}_I|\right), \tag{3.21}$$

where the symbol $\bigoplus_{I=1}^{N_A}(\mathbf{a}_I, \mathbf{b}_I)$ denotes the concatenation of the vectors $(\mathbf{a}_I, \mathbf{b}_I)$ from $I = 1$ to $I = N_A$, *i.e.* $\bigoplus_{I=1}^{N_A}(\mathbf{a}_I, \mathbf{b}_I) = (\mathbf{a}_1, \mathbf{b}_1, \mathbf{a}_2, \mathbf{b}_2, \ldots, \mathbf{a}_{N_A}, \mathbf{b}_{N_A})$. The four inputs associated with the relative positions of electrons i and j are

$$\mathbf{h}_{ij}^{0\alpha\beta} = \left(\mathbf{r}_i^\alpha - \mathbf{r}_j^\beta, |\mathbf{r}_i^\alpha - \mathbf{r}_j^\beta|\right). \tag{3.22}$$

[6] The inputs in Eq. (3.21) are being passed to the *one-electron stream* network, while the inputs in Eq. (3.22) are being passed to the *two-electron stream* network, with the superscript 0 denoting the input layer of both networks. At each layer $l \in \{0, \ldots, L-1\}$ of the networks, the outputs $\mathbf{h}_i^{l\alpha}$ and $\mathbf{h}_{ij}^{l\alpha\beta}$ from the two streams are averaged and concatenated in the following way

$$\begin{aligned}
\mathbf{g}^{l\uparrow} &= \frac{1}{N^\uparrow}\sum_{j=1}^{N^\uparrow}\mathbf{h}_j^{l\uparrow}, \quad \mathbf{g}^{l\downarrow} = \frac{1}{N^\downarrow}\sum_{j=1}^{N^\downarrow}\mathbf{h}_j^{l\downarrow}, \\
\mathbf{g}_i^{l\alpha\uparrow} &= \frac{1}{N^\uparrow}\sum_{j=1}^{N^\uparrow}\mathbf{h}_{ij}^{l\alpha\uparrow}, \quad \mathbf{g}_i^{l\alpha\downarrow} = \frac{1}{N^\downarrow}\sum_{j=1}^{N^\downarrow}\mathbf{h}_{ij}^{l\alpha\downarrow}, \\
\mathbf{f}_i^{l\alpha} &= \left(\mathbf{h}_i^{l\alpha}, \mathbf{g}^{l\uparrow}, \mathbf{g}^{l\downarrow}, \mathbf{g}_i^{l\alpha\uparrow}, \mathbf{g}_i^{l\alpha\downarrow}\right)
\end{aligned} \tag{3.23}$$

The concatenated one-electron vectors are then passed into the next layer, as are the two-electron vectors

$$\begin{aligned}
\mathbf{h}_i^{(l+1)\alpha} &= \tanh\left(\mathbf{V}^l\mathbf{f}_i^{l\alpha} + \mathbf{b}^l\right) + \mathbf{h}_i^{l\alpha}, \\
\mathbf{h}_{ij}^{(l+1)\alpha\beta} &= \tanh\left(\mathbf{W}^l\mathbf{h}_{ij}^{l\alpha\beta} + \mathbf{c}^l\right) + \mathbf{h}_{ij}^{l\alpha\beta},
\end{aligned} \tag{3.24}$$

where $\mathbf{V}^l \in \mathbb{R}^{n_l \times (3n_l + 2m_l)}$ and $\mathbf{W}^l \in \mathbb{R}^{m_l \times m_l}$ are matrices, $\mathbf{b}^l \in \mathbb{R}^{n_l}$ and $\mathbf{c}^l \in \mathbb{R}^{m_l}$ are vectors, and all of them are optimisable. We denote the number of hidden units (size of the vectors $\mathbf{h}$) at each layer l in the one- and two-electron streams by n_l and m_l respectively, such that $\mathbf{h}_i^{l\alpha} \in \mathbb{R}^{n_l}$ and $\mathbf{h}_{ij}^{l\alpha\beta} \in \mathbb{R}^{m_l}$ [7]. A schematic diagram of the operations in Eq. (3.23) and Eq. (3.24) can be found in Fig. (3.2). After the final layer of the network, denoting by the letter L, a set of latent space vectors $\mathbf{h}_i^{L\alpha} \equiv \mathbf{h}^{L\alpha}(\mathbf{r}_i^\alpha; \{\mathbf{r}_{/i}^\alpha\}; \{\mathbf{r}^{\bar{\alpha}}\}) \in \mathbb{R}^{n_L}$ with n_L being the size of the final layer of the network are generated, which are then used to construct the many-particle functions

$$\phi_i^{k\alpha}(\mathbf{r}_j^\alpha; \{\mathbf{r}_{/j}^\alpha\}; \{\mathbf{r}^{\bar{\alpha}}\}) = \left[\mathbf{w}_i^{k\alpha} \cdot \mathbf{h}_j^{L\alpha} + g_i^{k\alpha}\right]\chi_i^{k\alpha}(\mathbf{r}_j^\alpha), \tag{3.25}$$

where $\mathbf{w}_i^{k\alpha}$ and $g_i^{k\alpha}$ are both optimisable, and the superscript k denotes the kth term in the multi-determinant expansion. The $\chi_i^{k\alpha}(\mathbf{r}_j^\alpha)$ factor is an envelope function to ensure that the

[6] Since the neural network used in FermiNet is a continuous function, it is very difficult for the network to fit cusps, which are discontinuous in their first derivative. Hence, as discussed in the FermiNet paper in Ref. [121], by including the electron-nucleus and electron-electron distances as inputs, it becomes significantly easier for the network to learn the cusp conditions.

[7] Although n_l can be different in different layers, in practice, we use an equal number of hidden units in each layer except for the input layer, where n_0 is fixed by the number of particles in the system.

wave function satisfies the relevant boundary conditions. For example, in systems of isolated atoms and molecules, one could use the exponential envelope [121, 149]

$$\chi_i^{k\alpha}(\mathbf{r}_j^\alpha) = \sum_m \pi_{im}^{k\alpha} \exp\Big(-\sigma_{im}^{k\alpha}|\mathbf{r}_j^\alpha - \mathbf{R}_m|\Big), \tag{3.26}$$

where $\pi_{im}^{k\alpha}$ and $\sigma_{im}^{k\alpha}$ are optimisable parameters, to ensure the boundary conditions $\Psi(|\mathbf{r}_j^\alpha - \mathbf{R}_m| \to \infty) \to 0$ are satisfied.

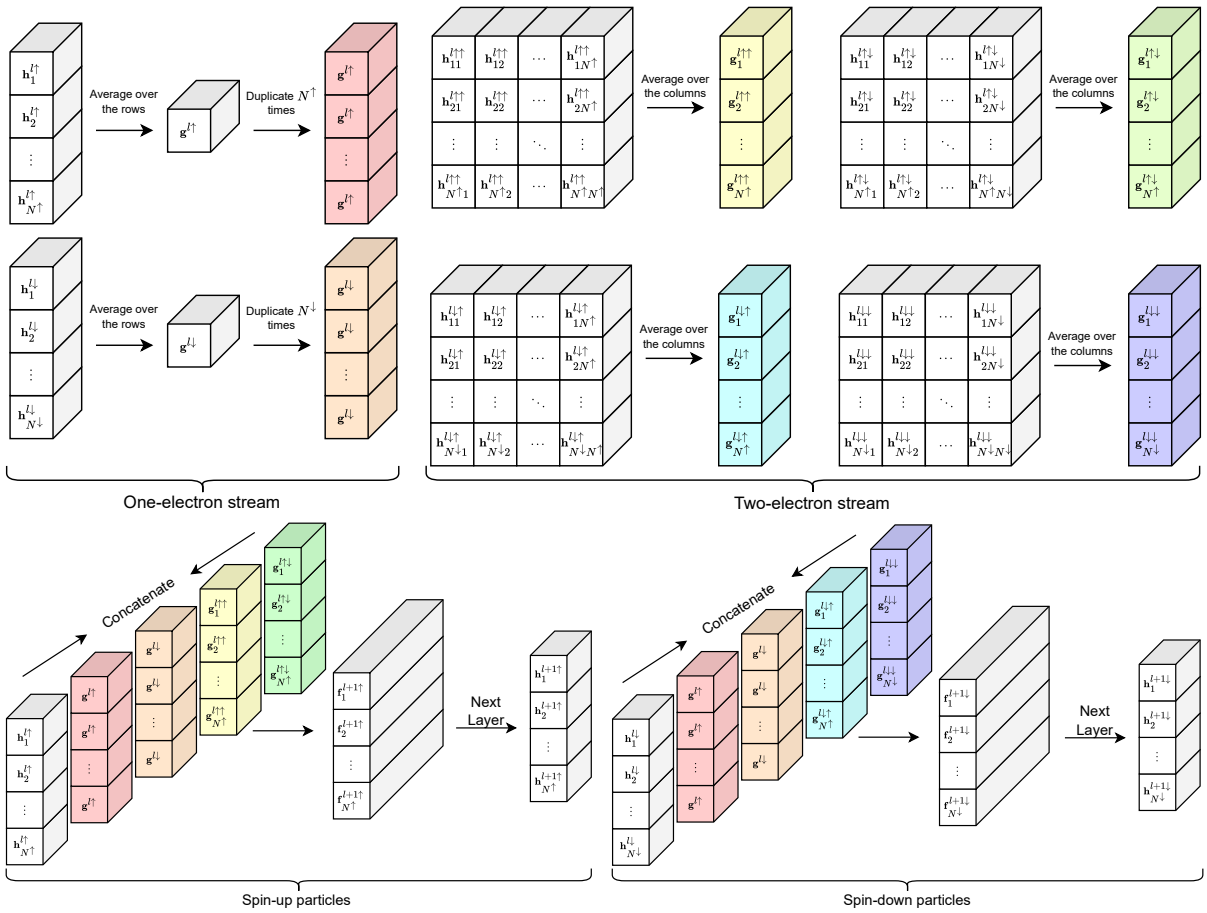


Figure 3.2: Schematic diagram of the operations in the intermediate layers of the FermiNet architecture in Eq. (3.23) and Eq. (3.24). The sets of arrays in the one and two-electron streams are first averaged and concatenated before being passed to the next layer.

The FermiNet NWF is given by [8]

$$\begin{aligned}&\Psi^D_{\text{FermiNet}}(\mathbf{r}_1^\uparrow, \dots, \mathbf{r}_{N^\uparrow}^\uparrow, \mathbf{r}_1^\downarrow, \dots, \mathbf{r}_{N^\downarrow}^\downarrow)\\ &= \sum_{k=1}^{D} \omega_k \det\Big[\phi_i^{k\uparrow}(\mathbf{r}_j^\uparrow; \{\mathbf{r}_{/j}^\uparrow\}; \{\mathbf{r}^\downarrow\})\Big] \det\Big[\phi_i^{k\downarrow}(\mathbf{r}_j^\downarrow; \{\mathbf{r}_{/j}^\downarrow\}; \{\mathbf{r}^\uparrow\})\Big].\end{aligned} \tag{3.27}$$

[8] If the inputs to the network contain electron-nuclear and electron-electron distances, as in the FermiNet [121], a Jastrow factor is not needed as the network is able to capture the cusp conditions.

3.4.4 Dense Determinant

As pointed out by the authors of the FermiNet [62], instead of using a spin-factored block-diagonal (BD) Slater determinant in Eq. (3.28), one can also use a full Slater determinant, such as the one in Eq. (3.29), in FermiNet. This has no effect on the overall exchange symmetry of the wave function.

$$\det \begin{bmatrix} \phi_i^{\uparrow}(\mathbf{r}_j^{\uparrow}; \{\mathbf{r}_{/j}^{\uparrow}\}; \{\mathbf{r}^{\downarrow}\}) & 0 \\ 0 & \phi_i^{\downarrow}(\mathbf{r}_j^{\downarrow}; \{\mathbf{r}_{/j}^{\downarrow}\}; \{\mathbf{r}^{\uparrow}\}) \end{bmatrix} = \begin{matrix} \det\left[\phi_i^{\uparrow}(\mathbf{r}_j^{\uparrow}; \{\mathbf{r}_{/j}^{\uparrow}\}; \{\mathbf{r}^{\downarrow}\})\right] \\ \times \det\left[\phi_i^{\downarrow}(\mathbf{r}_j^{\downarrow}; \{\mathbf{r}_{/j}^{\downarrow}\}; \{\mathbf{r}^{\uparrow}\})\right] \end{matrix} \tag{3.28}$$

$$\det \begin{bmatrix} \phi_i^{\uparrow}(\mathbf{r}_j^{\uparrow}; \{\mathbf{r}_{/j}^{\uparrow}\}; \{\mathbf{r}^{\downarrow}\}) & \phi_i^{\uparrow}(\mathbf{r}_j^{\downarrow}; \{\mathbf{r}_{/j}^{\downarrow}\}; \{\mathbf{r}^{\uparrow}\}) \\ \phi_i^{\downarrow}(\mathbf{r}_j^{\uparrow}; \{\mathbf{r}_{/j}^{\uparrow}\}; \{\mathbf{r}^{\downarrow}\}) & \phi_i^{\downarrow}(\mathbf{r}_j^{\downarrow}; \{\mathbf{r}_{/j}^{\downarrow}\}; \{\mathbf{r}^{\uparrow}\}) \end{bmatrix} = \det\left[\phi_i^{\alpha}(\mathbf{r}_j^{\beta}; \{\mathbf{r}_{/j}^{\beta}\}; \{\mathbf{r}^{\bar{\beta}}\})\right] \tag{3.29}$$

where α, β are the spins and $\bar{\beta}$ denotes the anti-parallel spin to β, e.g. if $\beta = \uparrow$ then $\bar{\beta} = \downarrow$ and vice versa. For some systems, the use of full Slater determinants has been shown to provide improved accuracy with negligible additional cost [24, 55, 58, 93, 95, 132]. We refer to this as a dense determinant (DD).

3.4.5 Periodic Systems in NNVMC

To extend the FermiNet to periodic systems, one can modify the inputs to make the Ansatz periodic. Consider a d-dimensional system with a periodically repeating finite simulation cell. The Bravais lattice vectors of the periodic lattice of copies of the simulation cell are $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_d\}$. Any position vector can be written as

$$\mathbf{r} = \sum_{i=1}^{d} s_i \mathbf{a}_i = \mathbf{A}\mathbf{s} \tag{3.30}$$

where $\mathbf{A} \equiv [\mathbf{a}_1, \ldots, \mathbf{a}_d] \in \mathbb{R}^{d \times d}$ is the change-of-basis matrix. To ensure that the FermiNet represents a periodic function, the position coordinates $\mathbf{s} = [s_1, \ldots, s_d]^T$ are mapped to pairs of periodic functions, $f\colon \mathbf{s} \mapsto (\sin(2\pi\mathbf{s}), \cos(2\pi\mathbf{s}))$. The element-wise function f maps every element s_i of the d-dimensional vector $\mathbf{s}$ to two real numbers, $\sin(2\pi s_i)$ and $\cos(2\pi s_i)$. The periodic input vectors are given by

$$\mathbf{h}_i^{0\alpha} = \left(f(\mathbf{s}_i^{\alpha}), \|f(\mathbf{s}_i^{\alpha})\|_p\right), \tag{3.31}$$

$$\mathbf{h}_{ij}^{0\alpha\beta} = \left(f(\mathbf{s}_{ij}^{\alpha\beta}), \left\|f(\mathbf{s}_{ij}^{\alpha\beta})\right\|_p\right), \tag{3.32}$$

and the periodic norm $\|\cdot\|_p$ is given by [24]

$$\|f(\mathbf{s})\|_p^2 := \sum_{ij} [1 - \cos(2\pi s_i)] S_{ij} [1 - \cos(2\pi s_j)] + \sin(2\pi s_i) S_{ij} \sin(2\pi s_j) \tag{3.33}$$

where $S_{ij} = \mathbf{a}_i \cdot \mathbf{a}_j$ is the metric tensor of the lattice coordinate system. This is to replace the L^2-norm in Euclidean space, and it also reduces to the Euclidean norm as $\mathbf{s}$ becomes small. Thus, if any electron is moved by any simulation-cell Bravais lattice vector, the inputs to the network are unchanged. It follows that the output, the value of the wave function, is also unchanged.

Although it is not strictly required, using a periodic envelope function of the form:

$$\chi_i^{k\alpha}(\mathbf{r}_j^\alpha) = \sum_m \left[\nu_{im}^{k\alpha} \cos\big(\mathbf{k}_m \cdot \mathbf{r}_j^\alpha\big) + \mu_{im}^{k\alpha} \sin\big(\mathbf{k}_m \cdot \mathbf{r}_j^\alpha\big)\right], \tag{3.34}$$

may sometimes increase the rate at which the optimisation of the network parameters converges [24]. Here the $\mathbf{k}_m$ are the simulation-cell reciprocal lattice vectors that is chosen up to a limit in momentum space. In system such as the homogeneous electron gas, this limit is chosen to be the Fermi wavevector of the non-interacting Fermi gas. This specific way of adapting the FermiNet to periodic systems was proposed by Cassella *et al.* [24], although other similar methods exist [91, 165].

3.5 Neural Wave Function Optimisation

In this section, we provide a brief introduction to the different optimisation methods for neural wave functions. As discussed in previous sections, the most straightforward method is to use gradient descent in Eq. (3.7). However, it is rarely used in practice, as it is prone to getting stuck in local minima or plateau regions in complex, non-convex landscapes [9]. In daily machine learning tasks, the most common optimisation algorithm is the ADAM algorithm [84]. This mitigates the issue of being trapped in local minima by introducing an adaptive learning rate and a "momentum" – an exponentially weighted average of past gradients – in the process of gradient descent. However, while sufficiently high accuracies can be achieved with the use of ADAM in machine learning tasks, it is not suitable for NNVMC problems, where both high accuracy and precision are required. As such, more sophisticated optimisation methods are often used, such as the stochastic reconfiguration (SR) method Here, we focus on two methods: the stochastic reconfiguration (SR) method [148] and the Kronecker-factored approximate Curvature (KFAC) method [102].

3.5.1 The Stochastic Reconfiguration Method

Introduced by Sorella [9, 148] in 1998, the stochastic reconfiguration method is widely used in conventional QMC and in some NNVMC problems. In contrast to vanilla gradient descent, which optimises the wave function in parameter space, the SR method aims to optimise the trial wave function in Hilbert space. It does so by utilising the Fubini-Study distance, $d(\psi_1, \psi_2) =$

[9] In mathematics, a convex function is a function whose Hessian is positive semi-definite everywhere. It has no saddle points, and any local minimum must be a global minimum [61]. In the field of machine learning, the loss function in almost all optimisation tasks is non-convex, as is the loss function in NNVMC discussed here.

$\arccos\sqrt{\frac{\langle\psi_1|\psi_2\rangle\langle\psi_2|\psi_1\rangle}{\langle\psi_1|\psi_1\rangle\langle\psi_2|\psi_2\rangle}}$, between two quantum states, ψ_1 and ψ_2, instead of their distance in the (Euclidean) parameter space. If we consider two quantum states, ψ and $\psi + \delta\psi$, where $\delta\psi$ is a infinitesimal variation of the state ψ that can be Taylor expanded as $\delta\psi \approx \delta\boldsymbol{\theta} \cdot \nabla_{\boldsymbol{\theta}}\psi$, the Fubini-Study distance becomes

$$\delta d^2 = \delta\boldsymbol{\theta}^T \mathcal{G} \delta\boldsymbol{\theta} \tag{3.35}$$

where $\mathcal{G}$ is the Fubini-Study metric tensor (FSM), also known as the quantum geometric tensor (QGT). It is given by [10]

$$\mathcal{G} = \mathrm{Re}\left\{\mathbb{E}_{|\Psi_T|^2}\left[\left(\nabla_{\boldsymbol{\theta}} \ln \Psi_T^{\boldsymbol{\theta}} - \mathbb{E}_{|\Psi_T|^2}\left[\nabla_{\boldsymbol{\theta}} \ln \Psi_T^{\boldsymbol{\theta}}\right]\right)^* \left(\nabla_{\boldsymbol{\theta}} \ln \Psi_T^{\boldsymbol{\theta}} - \mathbb{E}_{|\Psi_T|^2}\left[\nabla_{\boldsymbol{\theta}} \ln \Psi_T^{\boldsymbol{\theta}}\right]\right)^T\right]\right\} \tag{3.36}$$

and the SR update of the parameters $\boldsymbol{\theta}$ is given by [11]

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n - \alpha \mathcal{G}^{-1} \nabla_{\boldsymbol{\theta}} E(\boldsymbol{\theta}), \tag{3.37}$$

where n is the iteration number and α is the learning rate. This modifies the gradient term in the usual gradient descent procedure by multiplying it with the inverse of the FSM, a technique often known as preconditioning the gradient [12].

While the SR method is very efficient when used in conventional QMC methods, a problem emerges with the use of neural wave functions when calculating the inverse of the FSM. For conventional VMC methods, the calculation of the FSM is often achievable with a reasonable amount of computational resources, since the number of variational parameters is relatively small. However, this is not the case for neural wave functions, where the number of variational parameters can be of order 10^6. Therefore, reasonable approximations need to be made to make the computation of the inverse of the FSM more feasible with neural wave functions.

3.5.2 The Kronecker-Factored Approximate Curvature (KFAC) Method

The Kronecker-Factored Approximate Curvature (KFAC) algorithm, proposed by J. Martens and R. Grosse in 2015 [102], is designed specifically for optimising neural networks, and hence neural wave functions. It provides a more efficient way to compute the inverse of the Fisher information matrix (FIM)

$$\mathcal{F} = \mathbb{E}_{p(\mathbf{X};\boldsymbol{\theta})}\left[\left(\nabla_{\boldsymbol{\theta}} \log p(\mathbf{X};\boldsymbol{\theta})\right)\left(\nabla_{\boldsymbol{\theta}} \log p(\mathbf{X};\boldsymbol{\theta})\right)^T\right] \tag{3.38}$$

[10] We have assumed that the variational parameter $\boldsymbol{\theta}$ are all real.

[11] A detailed derivation of the SR method can be found in the book by Becca and Sorella [9].

[12] Readers from a machine learning or statistics background might be more familiar with the term "natural gradient descent" and "Fisher information matrix" as opposed to SR and FSM. Indeed, as shown in Ref. [121], the SR method with an unnormalised distribution is equivalent to the natural gradient descent method with the same but normalised distribution. The FSM is also equal to the Fisher information matrix up to a constant prefactor if the wave function is real.

by utilising the structure of a neural network.

Start by considering a feed-forward neural network with L network layers, similar to the one defined in Section 3.4.1, the FIM in Eq. (3.38) can be viewed as a $L \times L$ block matrix,

$$\mathcal{F} = \begin{pmatrix} F_{11} & F_{12} & \dots & F_{1L} \\ F_{21} & F_{22} & \dots & F_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ F_{L1} & F_{L2} & \dots & F_{LL} \end{pmatrix} \tag{3.39}$$

with each block F_{ij} given by

$$F_{ij} = \mathbb{E}_{p(\mathbf{X};\boldsymbol{\theta})} \left[\left(\frac{\partial \log p(\mathbf{X};\boldsymbol{\theta})}{\partial \mathbf{W}^i} \right) \left(\frac{\partial \log p(\mathbf{X};\boldsymbol{\theta})}{\partial \mathbf{W}^j} \right)^T \right] = \mathbb{E}_{p(\mathbf{X};\boldsymbol{\theta})} \left[(\mathbf{a}_i \otimes \mathbf{e}_i) (\mathbf{a}_j \otimes \mathbf{e}_j)^T \right] \tag{3.40}$$

where $\mathbf{W}^i = \mathrm{vec}(W^i) \in \mathbb{R}^{n_i + n_{i+1} + 1}$ are the vectorised weights and biases matrix, $W^i \in \mathbb{R}^{(n_{i+1}) \times (n_i + 1)}$, at layer i. To arrive at the second equality in Eq. (3.40), we have defined $\mathbf{a}_i$ as the forward activations, where the operations at layer i is given by $\mathbf{s}_i = W^i \mathbf{a}_i$, and

$$\mathbf{e}_i = \frac{\partial \log p(\mathbf{X};\boldsymbol{\theta})}{\partial \mathbf{s}^i} \tag{3.41}$$

is known as the backward sensitivities. Hence, from $\frac{\partial \log p(\mathbf{X};\boldsymbol{\theta})}{\partial \mathbf{W}^i} = \mathrm{vec}(\mathbf{e}_i \mathbf{a}_i^T)$ and using the identity $\mathrm{vec}(\mathbf{u}\mathbf{v}^T) = \mathbf{v} \otimes \mathbf{u}$, we arrive at the final step in Eq. (3.40).

The KFAC method mainly utilises two approximations for computing the inverse of the FIM in neural networks. The first approximation simply assumes the covariance matrix between different neural network layers to be zero, *i.e.*, $F_{ij} = 0$ if $i \neq j$, such that the FIM, $\mathcal{F}$, becomes a block-diagonal matrix. The second approximation further assumes the diagonal blocks of the FIM as a Kronecker product of much smaller matrices:

$$F_{ii} = \mathbb{E}_{p(\mathbf{X};\boldsymbol{\theta})} \left[(\mathbf{a}_i \otimes \mathbf{e}_i) (\mathbf{a}_i \otimes \mathbf{e}_i)^T \right] \approx \mathbb{E}_{p(\mathbf{X};\boldsymbol{\theta})} \left[\mathbf{a}_i \mathbf{a}_i^T \right] \otimes \mathbb{E}_{p(\mathbf{X};\boldsymbol{\theta})} \left[\mathbf{e}_i \mathbf{e}_i^T \right], \tag{3.42}$$

such that the inverse of this matrix, F_{ii}^{-1}, is simply

$$F_{ii}^{-1} = \mathbb{E}_{p(\mathbf{X};\boldsymbol{\theta})} \left[(\mathbf{a}_i \otimes \mathbf{e}_i) (\mathbf{a}_i \otimes \mathbf{e}_i)^T \right]^{-1} \approx \mathbb{E}_{p(\mathbf{X};\boldsymbol{\theta})} \left[\mathbf{a}_i \mathbf{a}_i^T \right]^{-1} \otimes \mathbb{E}_{p(\mathbf{X};\boldsymbol{\theta})} \left[\mathbf{e}_i \mathbf{e}_i^T \right]^{-1}, \tag{3.43}$$

from the matrix identity $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$. These two approximations allow the FIM to be computed using reasonable computational resources, without having to compute the exact inverse of the $10^6 \times 10^6$ FIM. In addition to these two approximations, KFAC also implements several numerical tricks to improve optimisations and prevent numerical instabilities. More details on the implementation of the KFAC method can be found in Ref. [102].

Note that the FSM used in the SR method in Eq. (3.36) with an *unnormalised* probability

density, $|\psi(\mathbf{X})|^2$, is equivalent to the FIM in Eq. (3.38) with the same but *normalised* probability density, $p(\mathbf{X};\boldsymbol{\theta}) \propto |\psi(\mathbf{X})|^2$, provided the wave function, $\psi(\mathbf{X})$, is real. Hence, the discussion presented in this section also applies to the FSM and neural wave functions. A more detailed discussion on KFAC with neural wave functions can be found in Ref. [121].

Chapter 4

Neural AGP Wave Functions

Despite the great successes of neural wave functions, the choice of the antisymmetrisation method can have a significant impact on the accuracies of the results [95], as we show in later chapters. Hence, in this chapter, we provide an alternative method of building a neural wave function based on the antisymmetrised geminal power (AGP) Ansatz [6, 7, 14, 57]. Unlike a Slater determinant, which is an antisymmetrised product of one-particle orbitals, an AGP wave function is an antisymmetrised product of two-particle states known as pairing orbitals or geminals. [6, 7, 14, 26, 57, 100]. The AGP wave function may be viewed as a fixed-particle number analogue of the BCS wave function, a form known to be better at capturing pairing correlations between particles [8, 14].

In the following section, we first provide a brief introduction to the AGP wave functions, followed by a discussion of how to build generalisations of them using neural networks. The chapter ends by presenting few proof-of-concept calculations on atoms and molecules.

Before introducing the AGP wave function, we first discuss how it originated from one of the most famous wave functions in condensed matter physics – the BCS wave function.

4.1 Fixed Particle Number BCS Wave Function

The standard BCS ground-state wave function in Eq. (2.14) from Chapter 2 is a Fock-space superposition of paired states with different numbers of pairs, and may be obtained by applying a superposition of products of different numbers of momentum space fermionic creation operators to the vacuum:

$$|\Psi_{\mathrm{BCS}}\rangle = \prod_{\mathbf{k}} \left(u_{\mathbf{k}} + v_{\mathbf{k}} \hat{c}^{\dagger}_{\mathbf{k},\uparrow} \hat{c}^{\dagger}_{-\mathbf{k},\downarrow} \right) |0\rangle \,. \tag{4.1}$$

To write this in first-quantised form for numerical simulations, it needs to be projected onto a fixed-particle number subspace:

$$\begin{aligned}
|\Psi_{\mathrm{BCS}}\rangle &= \prod_{\mathbf{k}} u_{\mathbf{k}} \left(1 + \varphi_{\mathbf{k}} \hat{c}^{\dagger}_{\mathbf{k},\uparrow} \hat{c}^{\dagger}_{-\mathbf{k},\downarrow}\right) |0\rangle \\
&= \left(\prod_{\mathbf{k}} u_{\mathbf{k}}\right) \left(\prod_{\mathbf{k}} e^{\varphi_{\mathbf{k}} \hat{c}^{\dagger}_{\mathbf{k},\uparrow} \hat{c}^{\dagger}_{-\mathbf{k},\downarrow}}\right) |0\rangle \\
&= \left(\prod_{\mathbf{k}} u_{\mathbf{k}}\right) \left(e^{\sum_{\mathbf{k}} \varphi_{\mathbf{k}} \hat{c}^{\dagger}_{\mathbf{k},\uparrow} \hat{c}^{\dagger}_{-\mathbf{k},\downarrow}}\right) |0\rangle \\
&= \left(\prod_{\mathbf{k}} u_{\mathbf{k}}\right) \sum_{p} \frac{1}{p!} \left[\sum_{\mathbf{k}} \varphi_{\mathbf{k}} \hat{c}^{\dagger}_{\mathbf{k},\uparrow} \hat{c}^{\dagger}_{-\mathbf{k},\downarrow}\right]^{p} |0\rangle
\end{aligned} \tag{4.2}$$

where $\varphi_{\mathbf{k}} = \frac{v_{\mathbf{k}}}{u_{\mathbf{k}}}$ is known as the pairing function. By expanding the exponential, we obtain a superposition of states with different numbers of pairs. Ignoring global phase factors and coefficients, the fixed-particle BCS wave function with p pairs of particles can be written as

$$|\Psi_{\mathrm{pBCS}}\rangle_p = \left[\sum_{\mathbf{k}} \varphi_{\mathbf{k}} \hat{c}^{\dagger}_{\mathbf{k},\uparrow} \hat{c}^{\dagger}_{-\mathbf{k},\downarrow}\right]^{p} |0\rangle \,, \tag{4.3}$$

where $p = N/2$ is the number of pairs in the system, N is the desired number of particles, and is the pairing function.

After a further projection of the wave function onto the position space, we get the real space wave function [14]:

$$\Psi_{\mathrm{pBCS}} = \mathcal{A}[\Phi(\mathbf{x}_1, \mathbf{x}_{1'})\Phi(\mathbf{x}_2, \mathbf{x}_{2'}) \cdots \Phi(\mathbf{x}_p, \mathbf{x}_{p'})], \tag{4.4}$$

where $\mathcal{A}$ is the antisymmetriser, and the primed and unprimed suffices index different spin species. The function $\Phi(\mathbf{x}_i, \mathbf{x}_{i'})$ is the pairing orbital between particle $\mathbf{x}_i \equiv (\mathbf{r}_i, \sigma_i)$ and particle $\mathbf{x}_{i'} \equiv (\mathbf{r}_{i'}, \sigma_{i'})$, given by

$$\Phi(\mathbf{x}_i, \mathbf{x}_{i'}) = \varphi_w(\mathbf{r}_i, \mathbf{r}_{i'}) \times \langle \sigma_i \sigma_{i'} | \frac{1}{\sqrt{2}} (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle), \tag{4.5}$$

where $\varphi_w(\mathbf{r}_i, \mathbf{r}_{i'}) = \sum_{\mathbf{k}} \varphi_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_{i'})}$ is the Fourier transform of $\varphi_{\mathbf{k}}$.

After spin assignment, it can be shown that the spatial part of Eq. (4.4) can be written as a determinant of the pairing function $\varphi_w(\mathbf{r}_i, \mathbf{r}_{i'})$ after spin-assignment [14, 26]

$$\Psi^{p}_{\mathrm{pBCS}} = \det \left[\varphi_w(\mathbf{r}^{\uparrow}_i, \mathbf{r}^{\downarrow}_j)\right]. \tag{4.6}$$

Note that this pBCS wave function only contains pairing between opposite spin particles. Hence, it can be used to describe systems with s-wave or d-wave pairing, but not systems with p-wave pairing of like spins.

4.2 AGP Singlet Wave Function

It is straight-forward to generalise Eq. (4.6) beyond BCS. Note that the pairing orbitals $\varphi_w(\mathbf{r}_i^{\uparrow}, \mathbf{r}_j^{\downarrow}) = \sum_{\mathbf{k}} \varphi_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{r}_i^{\uparrow}-\mathbf{r}_j^{\downarrow})}$ can be written as a bilinear product of single-particle plane wave orbitals, *i.e.*,

$$\varphi_w(\mathbf{r}_i^{\uparrow}, \mathbf{r}_j^{\downarrow}) = \sum_{\mathbf{k}} \varphi_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{r}_i^{\uparrow}-\mathbf{r}_j^{\downarrow})} = \sum_{\mathbf{k}} \sum_{\mathbf{q}} [\delta_{\mathbf{kq}} \varphi_{\mathbf{q}}] e^{i\mathbf{k}\cdot\mathbf{r}_i^{\uparrow}} e^{-i\mathbf{q}\cdot\mathbf{r}_j^{\uparrow}} \tag{4.7}$$

where the sums over $\mathbf{k}$ and $\mathbf{q}$ are both over all wave vectors commensurate with the periodically repeated simulation cell. If one replaces the plane wave orbitals with arbitrary single-particle orbitals $\phi_i(\mathbf{r}_j^{\sigma_j})$ and the diagonal matrix $[\delta_{\mathbf{kq}}\varphi_{\mathbf{k}}]$ by an arbitrary matrix, Eq. (4.7) becomes [1]

$$\varphi(\mathbf{r}_i^{\uparrow}, \mathbf{r}_j^{\downarrow}) = \sum_k \sum_l A_{kl}^{\uparrow\downarrow} \phi_k^{\uparrow *}(\mathbf{r}_i^{\uparrow}) \phi_l^{\downarrow}(\mathbf{r}_j^{\downarrow}) \tag{4.8}$$

and the pBCS wave function in Eq. (4.6) becomes the antisymmetric geminal power *singlet* (AGPs) wave function [2] [57].

4.2.1 AGPs Wave Function with Unpaired States

For systems with unpaired particles, such as spin-imbalanced systems, one needs to introduce a set of single-particle orbitals $\phi_i^{\sigma_i}(\mathbf{r}_j^{\sigma_j})$ to accommodate the unpaired particles.

For generality, consider a system with $N = 2p + u + d$ particles, where p is the number of pairs, u is the number of unpaired spin-up particles, and d is the number of unpaired spin-down particles. The total number of spin-up particles is $p + u$ and the total number of spin-down particles is $p + d$. The AGPs wave function can be written as a determinant of pairing functions

[1] Beware that unlike the symbol $\mathbf{k}$ in Eq. (4.7), the index k here is **not** the momentum in k-space – it is just a dummy index to label the single-particle orbitals.

[2] Note that this wave function are also known as the AGP or the BCS wave function in the literature, with the preferred term depending on the field of research.. We hope to clarify the subtle differences between different versions of the AGP wave function in the following sections.

and single-particle orbitals [14]

$$\begin{aligned}\Psi_{\text{AGPs}} &= \mathcal{A}\Big[\Phi(\mathbf{x}_1,\mathbf{x}_{1'})\cdots\Phi(\mathbf{x}_p,\mathbf{x}_{p'}) \\ &\qquad \times \phi_1(\mathbf{x}_{p+1})\cdots\phi_u(\mathbf{x}_{p+u})\phi_{u+1}(\mathbf{x}_{p+u+1})\cdots\phi_{u+d}(\mathbf{x}_{p+u+d})\Big] \qquad (4.9)\\ &= \det\begin{pmatrix} \varphi(\mathbf{r}_1^\uparrow,\mathbf{r}_1^\downarrow) & \cdots & \varphi(\mathbf{r}_1^\uparrow,\mathbf{r}_{p+d}^\downarrow) & \phi_1^\uparrow(\mathbf{r}_1^\uparrow) & \cdots & \phi_u^\uparrow(\mathbf{r}_1^\uparrow) \\ \varphi(\mathbf{r}_2^\uparrow,\mathbf{r}_1^\downarrow) & \cdots & \varphi(\mathbf{r}_2^\uparrow,\mathbf{r}_{p+d}^\downarrow) & \phi_1^\uparrow(\mathbf{r}_2^\uparrow) & \cdots & \phi_u^\uparrow(\mathbf{r}_2^\uparrow) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \varphi(\mathbf{r}_{p+u}^\uparrow,\mathbf{r}_1^\downarrow) & \cdots & \varphi(\mathbf{r}_{p+u}^\uparrow,\mathbf{r}_{p+d}^\downarrow) & \phi_1^\uparrow(\mathbf{r}_{p+u}^\uparrow) & \cdots & \phi_u^\uparrow(\mathbf{r}_{p+u}^\uparrow) \\ \phi_1^\downarrow(\mathbf{r}_1^\downarrow) & \cdots & \phi_1^\downarrow(\mathbf{r}_{p+d}^\downarrow) & 0 & \cdots & 0 \\ \phi_2^\downarrow(\mathbf{r}_1^\downarrow) & \cdots & \phi_2^\downarrow(\mathbf{r}_{p+d}^\downarrow) & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \phi_d^\downarrow(\mathbf{r}_1^\downarrow) & \cdots & \phi_d^\downarrow(\mathbf{r}_{p+d}^\downarrow) & 0 & \cdots & 0 \end{pmatrix} \qquad (4.10)\end{aligned}$$

where the second equation follows after spin-assignment, with the matrix dimension being $(p+u+d)\times(p+u+d)$. In practice, as ground-state systems usually like to pair as many particles as possible, one would set either u or d to be zero. For example, in a spin-imbalanced system with $N = 2p+u$ particles, where there are $p+u$ spin-up particles and p spin-down particles, the AGPs wave function would be

$$\Psi_{\text{AGPs}} = \det\begin{pmatrix} \varphi(\mathbf{r}_1^\uparrow,\mathbf{r}_1^\downarrow) & \cdots & \varphi(\mathbf{r}_1^\uparrow,\mathbf{r}_p^\downarrow) & \phi_1^\uparrow(\mathbf{r}_1^\uparrow) & \cdots & \phi_u^\uparrow(\mathbf{r}_1^\uparrow) \\ \varphi(\mathbf{r}_2^\uparrow,\mathbf{r}_1^\downarrow) & \cdots & \varphi(\mathbf{r}_2^\uparrow,\mathbf{r}_p^\downarrow) & \phi_1^\uparrow(\mathbf{r}_2^\uparrow) & \cdots & \phi_u^\uparrow(\mathbf{r}_2^\uparrow) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \varphi(\mathbf{r}_{p+u}^\uparrow,\mathbf{r}_1^\downarrow) & \cdots & \varphi(\mathbf{r}_{p+u}^\uparrow,\mathbf{r}_p^\downarrow) & \phi_1^\uparrow(\mathbf{r}_{p+u}^\uparrow) & \cdots & \phi_u^\uparrow(\mathbf{r}_{p+u}^\uparrow) \end{pmatrix}. \qquad (4.11)$$

4.3 AGP Wave Function

We now generalise the AGPs wave function with opposite-spin pairing to the AGP wave function with pairing between *all* particles.

Consider an *unpolarised* system of p spin-up fermions and p spin-down fermions, so that the total particle number $N = 2p$. A general pairing function can be written as a sum of bilinear products of arbitrary single-particle orbitals, which can be further separated into the different

spin-channels [6, 7, 57]

$$\begin{aligned}\Phi(\mathbf{x}_i,\mathbf{x}_j)=\sum_{kl}^{M}A_{kl}^{\sigma_i\sigma_j}\phi_k^{\sigma_i}(\mathbf{x}_i)\phi_l^{\sigma_j}(\mathbf{x}_j)=\varphi(\mathbf{r}_i,\mathbf{r}_j)\times\langle\sigma_i\sigma_j|\,\frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle-|\downarrow\uparrow\rangle)\\ +\chi^{+}(\mathbf{r}_i,\mathbf{r}_j)\times\langle\sigma_i\sigma_j|\,\frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle+|\downarrow\uparrow\rangle)\quad,\\ +\chi^{\uparrow\uparrow}(\mathbf{r}_i,\mathbf{r}_j)\times\langle\sigma_i\sigma_j|\uparrow\uparrow\rangle\\ +\chi^{\downarrow\downarrow}(\mathbf{r}_i,\mathbf{r}_j)\times\langle\sigma_i\sigma_j|\downarrow\downarrow\rangle\end{aligned}\tag{4.12}$$

where $\varphi(\mathbf{r}_i,\mathbf{r}_j)$ is a symmetric function while $\chi^{+}(\mathbf{r}_i,\mathbf{r}_j)$, $\chi^{\uparrow\uparrow}(\mathbf{r}_i,\mathbf{r}_j)$ and $\chi^{\downarrow\downarrow}(\mathbf{r}_i,\mathbf{r}_j)$ are antisymmetric functions to preserve exchange symmetry.

The AGP wave function is defined as the total antisymmetrisation of this general pairing function Eq. (4.12)

$$\Psi(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_{2p})=\mathcal{A}\left[\Phi(\mathbf{x}_1,\mathbf{x}_2)\Phi(\mathbf{x}_3,\mathbf{x}_4)\cdots\Phi(\mathbf{x}_{2p-1},\mathbf{x}_{2p})\right],\tag{4.13}$$

where $\mathcal{A}$ is the antisymmetrisation operator and $\Phi(\mathbf{x}_i,\mathbf{x}_j)$ is the pairing function between fermions $\mathbf{x}_i\equiv(\mathbf{r}_i,\sigma_i)$ and $\mathbf{x}_j\equiv(\mathbf{r}_j,\sigma_j)$. This wave function can also be written as the *Pfaffian* of an antisymmetric matrix [6, 7, 57]

$$\Psi(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_{2p})=\mathrm{pf}\begin{pmatrix}0 & \Phi(\mathbf{x}_1,\mathbf{x}_2) & \cdots & \Phi(\mathbf{x}_1,\mathbf{x}_{2p})\\ \Phi(\mathbf{x}_2,\mathbf{x}_1) & 0 & \cdots & \Phi(\mathbf{x}_2,\mathbf{x}_{2p})\\ \vdots & \vdots & \ddots & \vdots\\ \Phi(\mathbf{x}_{2p},\mathbf{x}_1) & \Phi(\mathbf{x}_{2p},\mathbf{x}_2) & \cdots & 0\end{pmatrix}.\tag{4.14}$$

Hence, the AGP wave function is also known as the Pfaffian wave function.

If one assign spins such that the first p particles are spin-up and the rest p particles are spin-down, the matrix of pairing functions can be split into four sub-matrices, each of which is a $p\times p$ matrix,

$$\begin{aligned}\Phi^{\uparrow\uparrow}(\mathbf{r}_i,\uparrow;\mathbf{r}_j,\uparrow)&=\chi^{\uparrow\uparrow}(\mathbf{r}_i,\mathbf{r}_j)\\ \Phi^{\downarrow\downarrow}(\mathbf{r}_i,\downarrow;\mathbf{r}_j,\downarrow)&=\chi^{\downarrow\downarrow}(\mathbf{r}_i,\mathbf{r}_j)\\ \Phi^{\uparrow\downarrow}(\mathbf{r}_i,\uparrow;\mathbf{r}_j,\downarrow)&=\frac{1}{\sqrt{2}}\left[\chi^{+}(\mathbf{r}_i,\mathbf{r}_j)+\varphi(\mathbf{r}_i,\mathbf{r}_j)\right]\\ \Phi^{\downarrow\uparrow}(\mathbf{r}_i,\downarrow;\mathbf{r}_j,\uparrow)&=\frac{1}{\sqrt{2}}\left[\chi^{+}(\mathbf{r}_i,\mathbf{r}_j)-\varphi(\mathbf{r}_i,\mathbf{r}_j)\right],\end{aligned}\tag{4.15}$$

and the AGP wave function of this unpolarised system can then be written as

$$\Psi(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_{2p})=\mathrm{pf}\begin{pmatrix}\Phi^{\uparrow\uparrow}(\mathbf{r}_i,\uparrow;\mathbf{r}_j,\uparrow) & \Phi^{\uparrow\downarrow}(\mathbf{r}_i,\uparrow;\mathbf{r}_j,\downarrow)\\ \Phi^{\downarrow\uparrow}(\mathbf{r}_i,\downarrow;\mathbf{r}_j,\uparrow) & \Phi^{\downarrow\downarrow}(\mathbf{r}_i,\downarrow;\mathbf{r}_j,\downarrow)\end{pmatrix}.\tag{4.16}$$

4.3.1 AGP Wave Function with Unpaired States

To allow the AGP wave function for unpaired states, one needs to introduce a set of single-particle orbitals $\{\phi(\mathbf{x}_i)\}$ to accommodate the unpaired particles.

Let us consider a system of $N = 2p + u + d$ spin-$\frac{1}{2}$ fermions with $p + u$ spin-up and $p + d$ spin-down particles, where u is the number of unpaired spin-up particles, d is the number of unpaired spin-down particles and p is the number of spin-up-spin-down paired particles in the system. By expanding the matrix of pairing orbitals with single-particle orbitals, the AGP wave function of this system can be written as

$$
\begin{aligned}
&\Psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) \\
&= \mathcal{A}\left[\Phi(\mathbf{x}_1, \mathbf{x}_2) \cdots \Phi(\mathbf{x}_{2p-1}, \mathbf{x}_{2p}) \phi_1(\mathbf{x}_{2p+1}) \cdots \phi_u(\mathbf{x}_{2p+u}) \phi_{u+1}(\mathbf{x}_{2p+u+1}) \cdots \phi_{u+d}(\mathbf{x}_N)\right] \\
&= \mathrm{pf} \begin{pmatrix}
0 & \Phi(\mathbf{x}_1, \mathbf{x}_2) & \cdots & \Phi(\mathbf{x}_1, \mathbf{x}_N) & \phi_1(\mathbf{x}_1) & \cdots & \phi_{u+d}(\mathbf{x}_1) \\
\Phi(\mathbf{x}_2, \mathbf{x}_1) & 0 & \cdots & \Phi(\mathbf{x}_2, \mathbf{x}_N) & \phi_1(\mathbf{x}_2) & \cdots & \phi_{u+d}(\mathbf{x}_2) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\Phi(\mathbf{x}_N, \mathbf{x}_1) & \Phi(\mathbf{x}_N, \mathbf{x}_2) & \cdots & 0 & \phi_1(\mathbf{x}_N) & \cdots & \phi_{u+d}(\mathbf{x}_N) \\
-\phi_1(\mathbf{x}_1) & -\phi_1(\mathbf{x}_2) & \cdots & -\phi_1(\mathbf{x}_N) & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
-\phi_{u+d}(\mathbf{x}_1) & -\phi_{u+d}(\mathbf{x}_2) & \cdots & -\phi_{u+d}(\mathbf{x}_N) & 0 & \cdots & 0
\end{pmatrix}
\end{aligned}
\tag{4.17}
$$

and the matrix becomes a $(N + u + d) \times (N + u + d) = 2(p + u + d) \times 2(p + u + d)$-dimensional matrix, which has a well-defined Pfaffian as the matrix dimension is now even.

After having assigned spins, setting the first $p + u$ particles to be spin-up and the remaining $p + d$ particles to be spin-down, the matrix can be split into different sub-matrices [6, 7, 57]:

$$
\Psi(\mathbf{r}_1^{\uparrow}, \mathbf{r}_2^{\uparrow}, \ldots, \mathbf{r}_{p+u}^{\uparrow}, \mathbf{r}_1^{\downarrow}, \mathbf{r}_2^{\uparrow}, \ldots, \mathbf{r}_{p+d}^{\downarrow}) = \mathrm{pf} \left(\begin{array}{c|c|c}
\boldsymbol{\Phi}^{\uparrow\uparrow} & \boldsymbol{\Phi}^{\uparrow\downarrow} & \\
\hline
\boldsymbol{\Phi}^{\downarrow\uparrow} & \boldsymbol{\Phi}^{\downarrow\downarrow} & \boldsymbol{\phi} \\
\hline
\multicolumn{2}{c|}{-\boldsymbol{\phi}^T} & 0
\end{array}\right)
\tag{4.18}
$$

where

$$
\begin{aligned}
&\left[\boldsymbol{\Phi}^{\uparrow\uparrow}\right]_{ij} = \Phi^{\uparrow\uparrow}(\mathbf{r}_i, \uparrow; \mathbf{r}_j, \uparrow) \qquad \left[\boldsymbol{\Phi}^{\uparrow\downarrow}\right]_{ij} = \Phi^{\uparrow\downarrow}(\mathbf{r}_i, \uparrow; \mathbf{r}_j, \downarrow) \qquad [\boldsymbol{\phi}]_{ij} = \phi_j(\mathbf{r}_i, \sigma_i). \\
&\left[\boldsymbol{\Phi}^{\downarrow\uparrow}\right]_{ij} = \Phi^{\downarrow\uparrow}(\mathbf{r}_i, \downarrow; \mathbf{r}_j, \uparrow) \qquad \left[\boldsymbol{\Phi}^{\downarrow\downarrow}\right]_{ij} = \Phi^{\downarrow\downarrow}(\mathbf{r}_i, \downarrow; \mathbf{r}_j, \downarrow)
\end{aligned}
\tag{4.19}
$$

In practice, following the same argument as before, we can set either u or d to be zero if we wish to describe the ground state of a system.

4.3.2 Relation to the AGPs Wave Function with Unpaired States

It is possible to show that the AGP wave function in Eq. (4.18) reduces to Eq. (4.10) if the pairing terms between same-spin particles are zero, *i.e.*, $\Phi^{\uparrow\uparrow}(\mathbf{r}_i,\uparrow;\mathbf{r}_j,\uparrow) = \Phi^{\downarrow\downarrow}(\mathbf{r}_i,\downarrow;\mathbf{r}_j,\downarrow) = 0$. In addition, we require the orbitals to be spin-dependent, such that the first u orbitals are labelled as $\phi_i^{\uparrow}, i \in \{1,\dots,u\}$ and the other d orbitals are labelled as $\phi_i^{\downarrow}, i \in \{1,\dots,d\}$. We also set $\phi_j^{\uparrow}(\mathbf{r}_i,\downarrow) = \phi_j^{\downarrow}(\mathbf{r}_i,\uparrow) = 0$, Eq. (4.18) becomes

$$\Psi(\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_{2p+u+d}) = \mathrm{pf}\begin{pmatrix} 0 & \boldsymbol{\Phi}^{\uparrow\downarrow} & \boldsymbol{\phi}^{\uparrow} & 0 \\ \boldsymbol{\Phi}^{\downarrow\uparrow} & 0 & 0 & \boldsymbol{\phi}^{\downarrow} \\ -\boldsymbol{\phi}^{\uparrow T} & 0 & 0 & 0 \\ 0 & -\boldsymbol{\phi}^{\downarrow T} & 0 & 0 \end{pmatrix}. \tag{4.20}$$

Note that according to Eq. (4.15), the two blocks of pairing orbitals, $\boldsymbol{\Phi}^{\uparrow\downarrow}$ and $\boldsymbol{\Phi}^{\downarrow\uparrow}$, are related to each other by

$$\Phi^{\downarrow\uparrow}(\mathbf{r}_i,\mathbf{r}_j) = \frac{1}{\sqrt{2}}\left[\chi^{+}(\mathbf{r}_i,\mathbf{r}_j) - \varphi(\mathbf{r}_i,\mathbf{r}_j)\right] = -\frac{1}{\sqrt{2}}\left[\chi^{+}(\mathbf{r}_j,\mathbf{r}_i) + \varphi(\mathbf{r}_j,\mathbf{r}_i)\right] = -\Phi^{\uparrow\downarrow}(\mathbf{r}_j,\mathbf{r}_i) \tag{4.21}$$

where we have used the fact that $\chi^{+}(\mathbf{r}_i,\mathbf{r}_j) = -\chi^{+}(\mathbf{r}_j,\mathbf{r}_i)$ and $\varphi(\mathbf{r}_i,\mathbf{r}_j) = \varphi(\mathbf{r}_j,\mathbf{r}_i)$ in the second equality. By swapping some columns and rows, and using the Pfaffian identity,

$$\mathrm{pf}\begin{pmatrix} 0 & M \\ -M^T & 0 \end{pmatrix} = (-1)^{n(n-1)/2}\det(M) \qquad M \in \mathbb{R}^{n\times n},$$

we get

$$\begin{aligned}\Psi(\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_{2p+u+d}) &= \mathrm{pf}\begin{pmatrix} 0 & \boldsymbol{\Phi}^{\uparrow\downarrow} & \boldsymbol{\phi}^{\uparrow} & 0 \\ -\boldsymbol{\Phi}^{\uparrow\downarrow T} & 0 & 0 & \boldsymbol{\phi}^{\downarrow} \\ -\boldsymbol{\phi}^{\uparrow T} & 0 & 0 & 0 \\ 0 & -\boldsymbol{\phi}^{\downarrow T} & 0 & 0 \end{pmatrix} \\ &= (-1)^{(p+u+d)d}\mathrm{pf}\begin{pmatrix} 0 & 0 & \boldsymbol{\Phi}^{\uparrow\downarrow} & \boldsymbol{\phi}^{\uparrow} \\ 0 & 0 & -\boldsymbol{\phi}^{\downarrow T} & 0 \\ -\boldsymbol{\Phi}^{\uparrow\downarrow T} & \boldsymbol{\phi}^{\downarrow} & 0 & 0 \\ -\boldsymbol{\phi}^{\uparrow T} & 0 & 0 & 0 \end{pmatrix} \\ &= (-1)^{(p+u+d)(p+u+3d-1)/2}\det\begin{pmatrix} \boldsymbol{\Phi}^{\uparrow\downarrow} & \boldsymbol{\phi}^{\uparrow} \\ -\boldsymbol{\phi}^{\downarrow T} & 0 \end{pmatrix}\end{aligned} \tag{4.22}$$

which is exactly Eq. (4.10) up to a sign difference.

4.4 Neural Network Pairing Orbitals

Having discussed the AGP wave function in general, we now turn our attention to building AGP wave functions using neural networks. First, we need to generalise the pairing orbitals to the many-body geminals, similar to how a single particle orbital is generalised to many-body orbitals. As we mainly uses the FermiNet NWF in our simulations, our discussion here will rely solely on the architecture of FermiNet. However, we also stress that all of the methods mentioned in this section can be readily applied to other NWF as well. For the rest of the thesis, in order to distinguish the original FermiNet with the new AGP-based FermiNet, we will call the former the *Slater FermiNet* as it is using Slater determinants. The latter, depending on which AGP wave function it is built upon, we will be called either as the *AGPs FermiNet* or the *Pfaffian FermiNet* (*PfaffNet*).

Consider the Slater FermiNet NWF, where the wave function is constructed by summing up determinants of orbital functions. The orbital functions are typically constructed by taking a scalar product of a latent space vector $\mathbf{h}_j^{L\alpha} \equiv \mathbf{h}^{L\alpha}(\mathbf{r}_j^\alpha, \{\mathbf{r}_{/j}^\alpha\}) \in \mathbb{R}^{n_L}$ from the neural network with another vector of variational parameters $\mathbf{w}_i^\alpha \in \mathbb{R}^{n_L}$

$$\phi_i^\alpha(\mathbf{r}_j^\alpha, \{\mathbf{r}_{/j}^\alpha\}) = \left[\mathbf{w}_i^\alpha \cdot \mathbf{h}_j^{L\alpha} + g_i^\alpha\right] \chi_i(\mathbf{r}_j^\alpha), \tag{4.23}$$

where $\alpha \in \{\uparrow, \downarrow\}$ denote the spin, n_L is the latent space dimension of the final layer of the neural network, g_i is a bias term, and $\chi_i(\mathbf{r}_j)$ is an envelope function, which is is frequently used to ensure that the wave function tends to zero as any $\mathbf{r}_i^\alpha$ tends to infinity in simulations of molecules or to enforce periodic or twisted boundary conditions in simulations of solids [121, 149].

To construct an AGP wave function using the same neural network, one needs to construct a pairing function, or geminal. One straightforward way is to take two latent space vectors $\mathbf{h}_j$ and $\mathbf{h}_k$ from the neural network and use them to build a geminal as follows:

$$\Phi^{\alpha\beta}(\mathbf{r}_j^\alpha, \mathbf{r}_k^\beta, \{\mathbf{r}_{/j}^\alpha, \mathbf{r}_{/k}^\beta\}) = \left[\left(\mathbf{h}_j^{L\alpha}\right)^T \mathbf{W}^{\alpha\beta} \mathbf{h}_k^{L\beta} + g^{\alpha\beta}\right] \chi(\mathbf{r}_j^\alpha, \mathbf{r}_k^\beta), \tag{4.24}$$

where $\mathbf{W}^{\alpha\beta}$ is an $n_L \times n_L$ matrix of variational parameters, $g^{\alpha\beta}$ is a bias term, and $\chi(\mathbf{r}_j^\alpha, \mathbf{r}_k^\beta)$ is an envelope function, which should respect the exchange symmetry of the geminal.

In general, geminals need to be antisymmetric under exchange of the two particle coordinates, which can be achieved by enforcing the same symmetry on the matrix $\mathbf{W}^{\alpha\beta}$. However, after spin assignment, only geminals that contain particle coordinates from the same spin channel need to follow this constraint; geminals with particle coordinates from different spin channels are not required to have a definite exchange symmetry, as particles with different spins become distinguishable after spin assignment. Therefore, the matrix $\mathbf{W}^{\alpha\beta}$ must be skew-symmetric for $\Phi^{\uparrow\uparrow}(\mathbf{r}_i^\uparrow; \mathbf{r}_j^\uparrow)$ and $\Phi^{\downarrow\downarrow}(\mathbf{r}_i^\downarrow; \mathbf{r}_j^\downarrow)$. It can be arbitrary for $\Phi^{\uparrow\downarrow}(\mathbf{r}_i^\uparrow; \mathbf{r}_j^\downarrow)$ and $\Phi^{\downarrow\uparrow}(\mathbf{r}_i^\downarrow; \mathbf{r}_j^\uparrow)$.

4.4.1 More Compact Versions of Geminals

As pointed out previously [6, 7, 57], geminals can be written in a more compact form by decomposing the matrix $\mathbf{W}^{\alpha\beta}$.

Same-spin geminals

For same-spin geminals, in order to respect the exchange symmetry, the bias term g must be zero, and the matrix $\mathbf{W}^{\alpha\beta}$ is required to be skew-symmetric. According to Zumino *et al.* [172] and Youla *et al.* [169], any skew-symmetric matrix W can be decomposed as

$$W = U^T \Lambda U, \tag{4.25}$$

where U is orthogonal if W real, or unitary if W is complex [3]. The matrix Λ is a direct-sum of 2×2 block matrices with real elements

$$\Lambda = \begin{pmatrix} 0 & \lambda_1 \\ -\lambda_1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & \lambda_2 \\ -\lambda_2 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & \lambda_M \\ -\lambda_M & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \tag{4.26}$$

where λ_i is the real part of the eigenvalues of W and the rank of matrix W is $2M$, so that there are M 2×2 matrices in the direct sum. This is sometimes known as the Youla decomposition [169]. The matrix U can be absorbed into the latent space vectors $\mathbf{h}_j$ given that the neural network is capable of representing both $\mathbf{h}_j$ and $\tilde{\mathbf{h}}_j = U\mathbf{h}_j$, which differ only by a rotation in latent space [95].

Under this transformation, the same-spin geminal can be written as

$$\Phi^{\alpha\alpha}(\mathbf{r}_j^\alpha, \mathbf{r}_k^\alpha, \{\mathbf{r}_{/j}^\alpha, \mathbf{r}_{/k}^\alpha\}) = \left[\left(\tilde{\mathbf{h}}_j^{L\alpha} \right)^T \mathbf{\Lambda}^{\alpha\beta} \tilde{\mathbf{h}}_k^{L\alpha} \right] \chi(\mathbf{r}_j^\alpha, \mathbf{r}_k^\alpha). \tag{4.27}$$

Given that the latent space dimension is often very large $n_L \sim 10^2$, this form of the geminal has significantly fewer variational parameters than the original form in Eq. 4.24, *i.e.* $n_L/2 \sim n_L$ compared to $n_L(n_L - 1)/2 \sim n_L^2$.

Opposite-spins geminals

For different-spin geminals, the real matrix W is not required to contain any symmetry after spin-assignment, as the opposite-spin particles are considered to be distinguishable. Thus, a singular value decomposition (SVD) can be used to decompose it into

$$W = UDV^T \tag{4.28}$$

[3] In practice, we always use real-valued matrices in the network. If a complex-valued wave function is required, a second set of real-valued parameters at the orbital-shaping layer are generated to produce the imaginary parts of the orbitals/geminals.

where U and V are orthogonal given W is real. Matrix D is a diagonal matrix with singular values of W on the diagonal.

$$D = \mathrm{diag}\left[\omega_1, \omega_2, \ldots, \omega_M, 0, \ldots, 0\right] \tag{4.29}$$

where ω_i are the singular values of W and M is the rank of the matrix W. As for the matrix U in a same-spin geminal, the matrices U and V in Eq. (4.28) can be absorbed into the latent space vectors $\mathbf{h}_j^{L\alpha}$ and the geminal can be written as

$$\begin{aligned}\Phi^{\alpha\bar{\alpha}}(\mathbf{r}_j^\alpha, \mathbf{r}_k^{\bar{\alpha}}, \{\mathbf{r}_{/j}^\alpha, \mathbf{r}_{/k}^{\bar{\alpha}}\}) &= \left[\left(\tilde{\mathbf{h}}_j^{L\alpha}\right)^T D^{\alpha\bar{\alpha}} \tilde{\mathbf{h}}_k^{L\bar{\alpha}} + g^{\alpha\bar{\alpha}}\right] \chi(\mathbf{r}_j^\alpha, \mathbf{r}_k^{\bar{\alpha}}) \\ &= \left[\mathbf{w}^{\alpha\bar{\alpha}} \cdot \left(\tilde{\mathbf{h}}_j^{L\alpha} \odot \tilde{\mathbf{h}}_k^{L\bar{\alpha}}\right) + g^{\alpha\bar{\alpha}}\right] \chi(\mathbf{r}_j^\alpha, \mathbf{r}_k^{\bar{\alpha}})\end{aligned} \tag{4.30}$$

where $\mathbf{w} = [\omega_1, \omega_2, \ldots, \omega_M, 0, \ldots, 0]^T$ and $\odot$ denotes the element-wise (Hadamard) product.

4.4.2 Other Forms of Geminals

So far our discussion has been focused on geminals constructed from the latent space vectors of the neural network, where the latter are usually constructed from the input coordinates of the electrons relative to the nuclei positions [54, 121, 138, 162]. However, some NWF Ansätze, one of which is the FermiNet, also contain networks that have particle-particle separations $\mathbf{r}_i^\alpha - \mathbf{r}_j^\beta$ as input, and output latent space vectors that have two particle indices $\mathbf{h}_{ij}^{L\alpha\beta} \equiv \mathbf{h}^{L\alpha\beta}(\mathbf{r}_i^\alpha, \mathbf{r}_j^\beta, \{\mathbf{r}_{/i}^\alpha, \mathbf{r}_{/j}^\beta\}) \in \mathbb{R}^{m_L}$, where m_L is the network size of the two-electron stream [121, 149]. This can also be added to the geminal in Eq. 4.24 to get

$$\Phi^{\alpha\beta}(\mathbf{r}_j^\alpha, \mathbf{r}_k^\beta, \{\mathbf{r}_{/j}^\alpha, \mathbf{r}_{/k}^\beta\}) = \left[\left(\mathbf{h}_j^{L\alpha}\right)^T \mathbf{W}^{\alpha\beta} \mathbf{h}_k^{L\beta} + \left(\mathbf{h}_{jk}^{L\alpha\beta}\right)^T \mathbf{Q}^{\alpha\beta} \mathbf{h}_{kj}^{L\beta\alpha} + g^{\alpha\beta}\right] \chi(\mathbf{r}_j^\alpha, \mathbf{r}_k^\beta). \tag{4.31}$$

where $\mathbf{Q}^{\alpha\beta} \in \mathbb{R}^{m_L \times m_L}$ is a matrix of variational parameters, which can either be antisymmetric or has no symmetry, depending on the the required exchange symmetry of particles i and j. Furthermore, we can (block) diagonalise the geminal matrices $\mathbf{W}^{\alpha\beta}$ and $\mathbf{Q}^{\alpha\beta}$ such that the many-body geminals become

$$\Phi^{\alpha\alpha}(\mathbf{r}_j^\alpha, \mathbf{r}_k^\alpha, \{\mathbf{r}_{/j}^\alpha, \mathbf{r}_{/k}^\alpha\}) = \left[\left(\tilde{\mathbf{h}}_j^{L\alpha}\right)^T \mathbf{\Lambda}_1^{\alpha\alpha} \tilde{\mathbf{h}}_k^{L\alpha} + \left(\mathbf{h}_{jk}^{L\alpha\alpha}\right)^T \mathbf{\Lambda}_2^{\alpha\alpha} \mathbf{h}_{kj}^{L\alpha\alpha}\right] \chi(\mathbf{r}_j^\alpha, \mathbf{r}_k^\alpha) \tag{4.32}$$

$$\Phi^{\uparrow\downarrow}(\mathbf{r}_j^\uparrow, \mathbf{r}_k^\downarrow, \{\mathbf{r}_{/j}^\uparrow, \mathbf{r}_{/k}^\downarrow\}) = \left[\mathbf{w}^{\uparrow\downarrow} \cdot \left(\tilde{\mathbf{h}}_j^{L\uparrow} \odot \tilde{\mathbf{h}}_k^{L\downarrow}\right) + \mathbf{q}^{\uparrow\downarrow} \cdot \left(\mathbf{h}_{jk}^{L\uparrow\downarrow} \odot \mathbf{h}_{kj}^{L\downarrow\uparrow}\right) + g^{\uparrow\downarrow}\right] \chi(\mathbf{r}_j^\uparrow, \mathbf{r}_k^\downarrow) \tag{4.33}$$

where $\alpha \in \{\uparrow, \downarrow\}$.

From our previous work [95], we found that this form of geminal gives higher accuracy and faster convergence in the unitary Fermi gas, which possesses a superfluid ground-state. We will present more details and in-depth discussion in Chapter (5).

4.5 The Neural AGP Wave Functions

Having discussed the construction of the geminals using neural networks, we now construct an AGP wave function using the geminals.

As mentioned in the previous section, the AGP wave function can be written as the Pfaffian of the matrix of geminals. For *spin-balanced systems* with $N^{\uparrow} = N^{\downarrow} = p$, a general neural AGP wave function with an arbitrary number of Pfaffians can be written as [4]

$$\Psi^{D}_{\text{PfaffNet}}(\mathbf{r}^{\uparrow}_1, \ldots, \mathbf{r}^{\uparrow}_p, \mathbf{r}^{\downarrow}_1, \ldots, \mathbf{r}^{\downarrow}_p) = \sum_{k}^{D} \omega_k \text{pf} \begin{pmatrix} \boldsymbol{\Phi}^{\uparrow\uparrow k} & \boldsymbol{\Phi}^{\uparrow\downarrow k} \\ -\left[\boldsymbol{\Phi}^{\uparrow\downarrow k}\right]^T & \boldsymbol{\Phi}^{\downarrow\downarrow k} \end{pmatrix} \tag{4.34}$$

where D is the number of Pfaffians, ω_k are the coefficients of the Pfaffians, which are usually taken to be one since they can be absorbed into the Pfaffians themselves, and the sub-matrices are

$$\begin{aligned} \left[\boldsymbol{\Phi}^{\uparrow\uparrow}\right]_{ij} &= \Phi^{\uparrow\uparrow}(\mathbf{r}^{\uparrow}_i, \mathbf{r}^{\uparrow}_j, \{\mathbf{r}^{\uparrow}_{/i,j}, \mathbf{r}^{\downarrow}\}), \\ \left[\boldsymbol{\Phi}^{\uparrow\downarrow}\right]_{ij} &= \Phi^{\uparrow\downarrow}(\mathbf{r}^{\uparrow}_i, \mathbf{r}^{\downarrow}_j, \{\mathbf{r}^{\uparrow}_{/i}, \mathbf{r}^{\downarrow}_{/j}\}), , \\ \left[\boldsymbol{\Phi}^{\downarrow\downarrow}\right]_{ij} &= \Phi^{\downarrow\downarrow}(\mathbf{r}^{\downarrow}_i, \mathbf{r}^{\downarrow}_j, \{\mathbf{r}^{\uparrow}, \mathbf{r}^{\downarrow}_{/i,j}\}) \end{aligned} \tag{4.35}$$

where the notation $\{\mathbf{r}^{\alpha}_{/i,j}\}$ means the set of all spatial coordinates of the spin-α particles except particle i and j.

Similarly, for *spin-imbalanced* systems with $(N^{\uparrow}, N^{\downarrow}) = (p+u, p)$, the neural AGP wave function can be written as

$$\Psi^{D}_{\text{PfaffNet}}(\mathbf{r}^{\uparrow}_1, \ldots, \mathbf{r}^{\uparrow}_{p+u}, \mathbf{r}^{\downarrow}_1, \ldots, \mathbf{r}^{\downarrow}_p) = \sum_{k}^{D} \omega_k \text{pf} \begin{pmatrix} \boldsymbol{\Phi}^{\uparrow\uparrow k} & \boldsymbol{\Phi}^{\uparrow\downarrow k} & \boldsymbol{\phi}^{\uparrow k} \\ -\left[\boldsymbol{\Phi}^{\uparrow\downarrow k}\right]^T & \boldsymbol{\Phi}^{\downarrow\downarrow k} & \boldsymbol{\phi}^{\downarrow k} \\ -\left[\boldsymbol{\phi}^{\uparrow k}\right]^T & -\left[\boldsymbol{\phi}^{\downarrow k}\right]^T & 0 \end{pmatrix} \tag{4.36}$$

where the extra sub-matrices are

$$[\boldsymbol{\phi}^{\alpha}]_{ij} = \phi_j(\mathbf{r}^{\alpha}_i; \{\mathbf{r}^{\alpha}_{/i}\}, \{\mathbf{r}^{\bar{\alpha}}\}), \quad \alpha \in \{\uparrow, \downarrow\}. \tag{4.37}$$

These are just standard FermiNet orbitals from the Slater FermiNet Ansatz.

4.5.1 Relationship between the AGPs FermiNet and the block-diagonal Slater FermiNet

Under certain circumstances, it is possible to relate the AGPs FermiNet NWF to the block-diagonal determinant Slater FermiNet NWF.

Consider a spin-balanced ($N^{\uparrow} = N^{\downarrow} = p$) periodic system, the full geminal in Eq. (4.24) for

[4] We refer to this Ansatz as the *Pfaffian FermiNet*, or *PfaffNet*, which is a generalisation of the Slater FermiNet Ansatz.

opposite spin particles, *i.e.*, $\alpha = \uparrow, \beta = \downarrow$, without the envelope function and the bias term can be written as [5]

$$\varphi_{\text{full}}(\mathbf{r}_i^{\uparrow}, \mathbf{r}_j^{\downarrow}; \{\mathbf{r}_{/i}^{\uparrow}\}; \{\mathbf{r}_{/j}^{\downarrow}\}) = \left(\mathbf{h}_j^{L\uparrow}\right)^T \mathbf{W} \mathbf{h}_k^{L\downarrow}, \tag{4.38}$$

where we have omitted the spin indices of the matrix for simplicity, *i.e.*, $\mathbf{W}^{\uparrow\downarrow} \equiv \mathbf{W}$. As we explain below, Eq. (4.38) is equivalent to the simpler FermiNet geminal given by

$$\varphi_{\text{diag}}(\mathbf{r}_i^{\uparrow}, \mathbf{r}_j^{\downarrow}; \{\mathbf{r}_{/i}^{\uparrow}\}; \{\mathbf{r}_{/j}^{\downarrow}\}) = \mathbf{w} \cdot \left(\tilde{\mathbf{h}}_j^{L\uparrow} \odot \tilde{\mathbf{h}}_k^{L\downarrow}\right), \tag{4.39}$$

which is the same as the diagonal geminal in Eq. (4.30) without the envelope function and the bias term. We choose to start with the many-particle geminal in the form of Eq. (4.38) only because this makes it easier to relate to FermiNet orbitals. Since Eq. (4.38) and Eq. (4.39) are equivalent as we will show later, the choice does not affect the conclusions of the argument. [6]

To find the relationship between Eq. (4.38) and Eq. (4.39), it is helpful to represent the matrix $\mathbf{W}$ as its singular-value decomposition (SVD):

$$[\mathbf{W}]_{kl} = W_{kl} = \sum_{\gamma=1}^{n_L} \sigma_\gamma U_{\gamma k} V_{\gamma l}^{\beta}, \tag{4.40}$$

where $U \in \mathbb{R}^{n_L \times n_L}$ and $V \in \mathbb{R}^{n_L \times n_L}$ are orthogonal matrices and n_L is the size of the vectors $\mathbf{h}_i^{L\alpha}$ output by the final layer L of the one-electron stream. This is also known as the number of hidden units in layer L. The many-particle geminal in Eq. (4.38) becomes

$$\begin{aligned}
\varphi_{\text{full}}(\mathbf{r}_i^{\uparrow}, \mathbf{r}_j^{\downarrow}) &= \sum_{kl} W_{kl} h_i^{L\uparrow(k)} h_j^{L\downarrow(l)} && (4.41)\\
&= \sum_{\gamma} \sigma_\gamma \left(\sum_k U_{\gamma k} h_i^{L\uparrow(k)}\right)\left(\sum_l V_{\gamma l} h_j^{L\uparrow(l)}\right) && (4.42)\\
&= \sum_{\gamma} \sigma_\gamma (U\mathbf{h}_i^{L\uparrow})_\gamma (V\mathbf{h}_j^{L\downarrow})_\gamma. && (4.43)
\end{aligned}$$

Given the universal approximation theorem [77], and the fact that every layer of the network contains an arbitrary linear transformation, it is reasonable to assume that the functions $\mathbf{h}_i^{L\alpha}$ and $O\mathbf{h}_i^{L\alpha}$, where O is either U or V, have the same variational freedom and information content. In other words, we assume that any network capable of representing $\mathbf{h}_i^{L\alpha}$ can also represent $O\mathbf{h}_i^{L\alpha}$, since this is merely a rotation of the vectors $\mathbf{h}_i^{L\alpha}$ in the last layer. We thus define $\tilde{\mathbf{h}}_i^{L\uparrow} = U\mathbf{h}_i^{L\uparrow}$

[5] While the use of an envelope function is essential in infinite systems to satisfy the relevant boundary conditions, it can be omitted in periodic systems. As stated in the work by Cassella *et al.* [24], the use of an envelope function can circumvent numerical difficulties and speed up the network optimisation, but it is not always required in actual simulations. The bias term can also be neglected as we usually set it to zero in our simulations without affecting the accuracy of the results.

[6] In the rest of this section, for the sake of simplicity, we omit the sets $\{\mathbf{r}_{/i}^{\uparrow}\}$ and $\{\mathbf{r}_{/j}^{\downarrow}\}$ from the arguments of the many-particle geminals and orbitals.

and $\tilde{\mathbf{h}}_i^{L\downarrow} = V\mathbf{h}_i^{L\downarrow}$, such that the many-particle geminal becomes

$$\varphi_{\text{full}}(\mathbf{r}_i^\uparrow, \mathbf{r}_j^\downarrow) = \sum_\gamma \sigma_\gamma (U\mathbf{h}_i^{L\uparrow})_\gamma (V\mathbf{h}_j^{L\downarrow})_\gamma = \sum_\gamma \sigma_\gamma (\tilde{\mathbf{h}}_i^{L\uparrow} \odot \tilde{\mathbf{h}}_j^{L\downarrow})_\gamma, \tag{4.44}$$

which is equivalent to Eq. (4.39).

To relate the AGPs FermiNet and the Slater FermiNet, we expand an AGPs determinant constructed using the many-particle geminal from Eq. (4.38) as a sum of block-diagonal determinants of FermiNet orbitals. It will be sufficient to consider matrices $W \in \mathbb{R}^{n_L \times n_L}$ of rank M, with $p \leq M \leq n_L$. We can decompose any such matrix using rank factorisation,

$$W_{kl} = \sum_{\gamma=1}^{M} F_{\gamma k} G_{\gamma l}, \tag{4.45}$$

where F and G are matrices in $\mathbb{R}^{M \times n_L}$. Eq. (4.38) then becomes

$$\varphi(\mathbf{r}_i^\uparrow, \mathbf{r}_j^\downarrow) = \sum_{\gamma=1}^{M} \left(\sum_{k=1}^{n_L} F_{\gamma k} h_i^{L\uparrow(k)} \right) \left(\sum_{l=1}^{n_L} G_{\gamma l} h_j^{L\downarrow(l)} \right) = \sum_{\gamma=1}^{M} \phi_\gamma^\uparrow(\mathbf{r}_i^\uparrow) \phi_\gamma^\downarrow(\mathbf{r}_j^\downarrow), \tag{4.46}$$

where the last line defines the FermiNet orbitals $\phi_\gamma^\uparrow$ and $\phi_\gamma^\downarrow$. In the case when $M = p$, where $p = N^\uparrow = N^\downarrow$ is the number of pairs in the system, the determinant of the many-particle geminal can be written as a block-diagonal determinant of FermiNet orbitals:

$$\begin{aligned} \det\left[\varphi(\mathbf{r}_i^\uparrow, \mathbf{r}_j^\downarrow)\right] &= \det\left[\sum_{\gamma=1}^{p} \phi_\gamma^\uparrow(\mathbf{r}_i^\uparrow) \phi_\gamma^\downarrow(\mathbf{r}_j^\downarrow)\right] \\ &= \det\left[\left(\phi^{\uparrow T} \phi^\downarrow\right)_{ij}\right] \\ &= \det\left[\phi_i^\uparrow(\mathbf{r}_j^\uparrow)\right] \det\left[\phi_i^\downarrow(\mathbf{r}_j^\downarrow)\right] \end{aligned} \tag{4.47}$$

where $[\phi^\alpha]_{ij} = \phi_i^\alpha(\mathbf{r}_j^\alpha)$ are matrices in $\mathbb{R}^{M \times p}$ with $M = p$. The product of two $p \times p$ determinants can be written as the determinant of a single $2p \times 2p$ matrix, with the $p \times p$ spin-up and spin-down blocks on the diagonal. Therefore, a single-geminal AGPs FermiNet wave function constructed using the many-particle geminal from Eq. (4.38) with a rank-p matrix W_{kl} is equivalent to a $2p \times 2p$ block-diagonal determinant of FermiNet orbitals. The equivalence is already well known [100] for AGPs wave functions constructed using conventional two-particle orbitals.

Now consider the more general case where $p \leq M \leq n_L$. The Cauchy-Binet formula states

that

$$\det\Big[\varphi(\mathbf{r}_i^{\uparrow},\mathbf{r}_j^{\downarrow})\Big] = \det\left[\sum_{\gamma=1}^{M}\phi_\gamma^{\uparrow}(\mathbf{r}_i^{\uparrow})\phi_\gamma^{\downarrow}(\mathbf{r}_j^{\downarrow})\right]$$
$$= \sum_{1\le j_1<j_2<\cdots<j_p\le M} \det\begin{pmatrix} \phi_{j_1}^{\uparrow}(\mathbf{r}_1^{\uparrow}) & \phi_{j_1}^{\uparrow}(\mathbf{r}_2^{\uparrow}) & \cdots & \phi_{j_1}^{\uparrow}(\mathbf{r}_p^{\uparrow}) \\ \phi_{j_2}^{\uparrow}(\mathbf{r}_1^{\uparrow}) & \phi_{j_2}^{\uparrow}(\mathbf{r}_2^{\uparrow}) & \cdots & \phi_{j_2}^{\uparrow}(\mathbf{r}_p^{\uparrow}) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{j_p}^{\uparrow}(\mathbf{r}_1^{\uparrow}) & \phi_{j_p}^{\uparrow}(\mathbf{r}_2^{\uparrow}) & \cdots & \phi_{j_p}^{\uparrow}(\mathbf{r}_p^{\uparrow}) \end{pmatrix}$$
$$\times \det\begin{pmatrix} \phi_{j_1}^{\downarrow}(\mathbf{r}_1^{\downarrow}) & \phi_{j_1}^{\downarrow}(\mathbf{r}_2^{\downarrow}) & \cdots & \phi_{j_1}^{\downarrow}(\mathbf{r}_p^{\downarrow}) \\ \phi_{j_2}^{\downarrow}(\mathbf{r}_1^{\downarrow}) & \phi_{j_2}^{\downarrow}(\mathbf{r}_2^{\downarrow}) & \cdots & \phi_{j_2}^{\downarrow}(\mathbf{r}_p^{\downarrow}) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{j_p}^{\downarrow}(\mathbf{r}_1^{\downarrow}) & \phi_{j_p}^{\downarrow}(\mathbf{r}_2^{\downarrow}) & \cdots & \phi_{j_p}^{\downarrow}(\mathbf{r}_p^{\downarrow}) \end{pmatrix}, \tag{4.48}$$

where the sum is over all $\binom{M}{p}$ distinct choices of p rows from the two $M \times p$ matrices $\phi_\gamma^{\uparrow}(\mathbf{r}_i^{\uparrow})$ and $\phi_\gamma^{\downarrow}(\mathbf{r}_i^{\downarrow})$. The products of the determinants of the two $p \times p$ matrices associated with each such choice are summed to reproduce the AGPs. This is similar to the linear combination of multiple block-diagonal-determinants of FermiNet orbitals without weights given by Eq. (3.27) and in the original FermiNet paper [121] [7]. Note that the intermediate layers, *i.e.*, the one and two-electron streams, are identical in the Slater FermiNet and the AGPs FermiNet. The only modifications are made at the orbital shaping layer, or, equivalently, the method of antisymmetrisation has changed. Thus, the representational power of the intermediate layers of the AGPs FermiNet remains the same as for the Slater FermiNet.

We have shown that a single AGPs determinant constructed using the many-particle geminal from Eq. (4.38) with a matrix W_{kl} of rank greater than p contains multiple block-diagonal determinants of FermiNet orbitals. If the rank of W_{kl} is equal to p, the AGPs is equivalent to a single block-diagonal FermiNet determinant. Conversely, any single-determinant FermiNet wave function can be written as an AGPs of rank p. Therefore, the AGPs FermiNet provides a more powerful Ansatz with fewer variational parameters than the Slater FermiNet, since the former contains the latter.

It is worth mentioning another advantage of using the FermiNet to build geminals. By generating more sets of independent parameters $\mathbf{w}^k$ in Eq. (4.39), one can easily construct an arbitrary number D of FermiNet geminals $\varphi^k(\mathbf{r}_i^{\alpha},\mathbf{r}_j^{\bar{\alpha}};\{\mathbf{r}_{/i}^{\alpha}\};\{\mathbf{r}_{/j}^{\bar{\alpha}}\})$ with $k \in \{1,2,\ldots,D\}$, all without the use of a basis set. This allows one to use weighted sums of AGPs determinants as trial wave functions, similar to the weighted sum of conventional FermiNet determinants seen in Eq. (3.27).

[7] To get a rough estimate of the number of terms in the sum, we can take the number of hidden units in the one-electron stream to be $M = n_L \sim 200$ and the number of pairs to be $p \sim 20$, this number is approximately 10^{27}.

4.5.2 Relationship between the Pfaffian FermiNet and the dense Slater FermiNet

It is also possible to relate the PfaffNet NWF to the dense determinant Slater FermiNet NWF, similar to the relationship between the AGPs FermiNet and the block-diagonal determinant Slater FermiNet.

Consider the same spin-balanced ($N^{\uparrow} = N^{\downarrow} = p$) periodic system as before, the full geminal for spin-α and spin-β particles, *i.e.*, $\alpha, \beta \in \{\uparrow, \downarrow\}$, without the envelope function and the bias term can be written as

$$\Phi^{\alpha\beta}(\mathbf{r}_j^{\alpha}, \mathbf{r}_k^{\beta}, \{\mathbf{r}_{/j}^{\alpha}, \mathbf{r}_{/k}^{\beta}\}) = \Phi_{jk}^{\alpha\beta} = \left(\mathbf{h}_j^{L\alpha}\right)^T \mathbf{W}^{\alpha\beta} \mathbf{h}_k^{L\beta}, \tag{4.49}$$

where the matrix $\mathbf{W}^{\alpha\beta}$ is antisymmetric if $\alpha = \beta$ and has no symmetry if $\alpha \neq \beta$. [8] The matrix

$$\Phi = \begin{pmatrix} \mathbf{\Phi}^{\uparrow\uparrow} & \mathbf{\Phi}^{\uparrow\downarrow} \\ -\left[\mathbf{\Phi}^{\uparrow\downarrow}\right]^T & \mathbf{\Phi}^{\downarrow\downarrow} \end{pmatrix} = \begin{pmatrix} \left[\mathbf{h}^{L\uparrow}\right]^T W^{\uparrow\uparrow} \mathbf{h}^{L\uparrow} & \left[\mathbf{h}^{L\uparrow}\right]^T W^{\uparrow\downarrow} \mathbf{h}^{L\downarrow} \\ -\left[\mathbf{h}^{L\downarrow}\right]^T \left[W^{\uparrow\downarrow}\right]^T \mathbf{h}^{L\uparrow} & \left[\mathbf{h}^{L\downarrow}\right]^T W^{\downarrow\downarrow} \left[\mathbf{h}^{L\downarrow}\right]^T \end{pmatrix} \tag{4.50}$$

can be written as

$$\Phi = \begin{pmatrix} \left[\mathbf{h}^{L\uparrow}\right]^T & 0 \\ 0 & \left[\mathbf{h}^{L\downarrow}\right]^T \end{pmatrix} \begin{pmatrix} W^{\uparrow\uparrow} & W^{\uparrow\downarrow} \\ -\left[W^{\uparrow\downarrow}\right]^T & W^{\downarrow\downarrow} \end{pmatrix} \begin{pmatrix} \mathbf{h}^{L\uparrow} & 0 \\ 0 & \mathbf{h}^{L\downarrow} \end{pmatrix} \tag{4.51}$$

where $\mathbf{h}^{L\alpha} = \left(\mathbf{h}_1^{L\alpha}, \mathbf{h}_2^{L\alpha}, \ldots, \mathbf{h}_p^{L\alpha}\right) \in \mathbb{R}^{n_L \times p}$ is an array of latent space vectors $\mathbf{h}_j^{L\alpha} \in \mathbb{R}^{n_L}$ of spin-α particles. The matrix in the middle is skew-symmetric by construction, which, similar to Eq. (4.25), can be decomposed as

$$\mathbb{W} = \begin{pmatrix} W^{\uparrow\uparrow} & W^{\uparrow\downarrow} \\ -\left[W^{\uparrow\downarrow}\right]^T & W^{\downarrow\downarrow} \end{pmatrix} = \mathbb{U}^T \Lambda \mathbb{U} \tag{4.52}$$

where $\mathbb{U}$ is orthogonal if $\mathbb{W}$ real, or unitary if $\mathbb{W}$ is complex, and Λ has the same block-diagonal form as Eq. (4.26). Note that if the rank of the matrix $\mathbb{W}$ is $2M$ where $p \leq M \leq n_L$, the matrix Λ will only have M non-zero block-diagonal 2×2 matrices. In this case, the rows and columns of zeros in matrix Λ and the corresponding rows in matrix $\mathbb{U}$ can be dropped, such that the matrix $\mathbb{U} \in \mathbb{R}^{2M \times 2n_L}$ is now rectangular and $\Lambda \in \mathbb{R}^{2M \times 2M}$ becomes a smaller square matrix. Hence,

[8] Again, in the rest of this section, for the sake of simplicity, we omit the sets $\{\mathbf{r}_{/i}^{\uparrow}\}$ and $\{\mathbf{r}_{/j}^{\downarrow}\}$ from the arguments of the many-particle geminals and orbitals.

the geminal matrix in Eq. (4.50) with $\text{rank}(\mathbb{W}) = 2M$ can be written as

$$\begin{aligned}\Phi &= \left[\begin{pmatrix} U^{\uparrow\uparrow} & U^{\uparrow\downarrow} \\ U^{\downarrow\uparrow} & U^{\downarrow\downarrow} \end{pmatrix}\begin{pmatrix} \mathbf{h}^{L\uparrow} & 0 \\ 0 & \mathbf{h}^{L\downarrow} \end{pmatrix}\right]^T \Lambda \begin{pmatrix} U^{\uparrow\uparrow} & U^{\uparrow\downarrow} \\ U^{\downarrow\uparrow} & U^{\downarrow\downarrow} \end{pmatrix}\begin{pmatrix} \mathbf{h}^{L\uparrow} & 0 \\ 0 & \mathbf{h}^{L\downarrow} \end{pmatrix} \\ &= \begin{pmatrix} U^{\uparrow\uparrow}\mathbf{h}^{L\uparrow} & U^{\uparrow\downarrow}\mathbf{h}^{L\downarrow} \\ U^{\downarrow\uparrow}\mathbf{h}^{L\uparrow} & U^{\downarrow\downarrow}\mathbf{h}^{L\downarrow} \end{pmatrix}^T \Lambda \begin{pmatrix} U^{\uparrow\uparrow}\mathbf{h}^{L\uparrow} & U^{\uparrow\downarrow}\mathbf{h}^{L\downarrow} \\ U^{\downarrow\uparrow}\mathbf{h}^{L\uparrow} & U^{\downarrow\downarrow}\mathbf{h}^{L\downarrow} \end{pmatrix}\end{aligned} \tag{4.53}$$

where we have split the $2M \times 2M$ matrix $\mathbb{U}$ into four $M \times M$ blocks as

$$\mathbb{U} = \begin{pmatrix} U^{\uparrow\uparrow} & U^{\uparrow\downarrow} \\ U^{\downarrow\uparrow} & U^{\downarrow\downarrow} \end{pmatrix}, \qquad U^{\alpha\beta} \in \mathbb{R}^{M\times M}, \qquad \alpha, \beta \in \{\uparrow, \downarrow\}, \tag{4.54}$$

Note that under this arrangement for $\mathbb{U}$, some rows and columns of the matrix Λ need to be rearranged such that the form of the matrix becomes

$$\Lambda = \text{diag}\left[\begin{pmatrix} 0 & \lambda_1 \\ -\lambda_1 & 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 & \lambda_M \\ -\lambda_M & 0 \end{pmatrix}\right] \Rightarrow \begin{pmatrix} 0 & D \\ -D & 0 \end{pmatrix} \tag{4.55}$$

where $D = \text{diag}(\lambda_1, \ldots, \lambda_M)$ is a diagonal matrix. This is allowed because the overall product, $\mathbb{U}^T\Lambda\mathbb{U}$, is invariant under orthogonal transformations, *i.e.*, $\Phi = \mathbb{U}^T\Lambda\mathbb{U} = \mathbb{U}^T P^T P\Lambda P^T P\mathbb{U} = (P\mathbb{U})^T(P\Lambda P^T)(P\mathbb{U}) = \tilde{\mathbb{U}}^T\tilde{\Lambda}\tilde{\mathbb{U}}$ with $\tilde{\mathbb{U}} = P\mathbb{U}$ and $\tilde{\Lambda} = P\Lambda P^T$. P is an arbitrary orthogonal transformation matrix, which could act as a transformation matrix that swaps the selected rows when acting on another matrix. We can then define a set of FermiNet orbitals as

$$\left[U^{\alpha\beta}\mathbf{h}^{L\beta}\right]_{ij} = \sum_k U^{\alpha\beta}_{ik} h^{L\beta(k)}_j = \phi^\alpha_i(\mathbf{r}^\beta_j; \{\mathbf{r}^\beta_{/j}\}; \{\mathbf{r}^{\bar{\beta}}\}) \tag{4.56}$$

where $i = 1, \ldots, M$ is the orbital index, and $j = 1, \ldots, p$ is the particle index of spin-β. Finally, if we take the Pfaffian of the geminal matrix and use the minor summation formula [79, 80, 111], we get a factorial sum of dense determinants of selected columns and rows (also known as minors) of the FermiNet orbital matrix, similar to the Cauchy-Binet formula we discussed in the previous section

$$
\begin{aligned}
&\mathrm{pf}(\Phi) \\
&= \mathrm{pf}(\boldsymbol{\phi}^T \Lambda \boldsymbol{\phi}) \\
&= \mathrm{pf}\left[\begin{pmatrix}
\phi_1^\uparrow(\mathbf{r}_1^\uparrow) & \cdots & \phi_M^\uparrow(\mathbf{r}_1^\uparrow) & \phi_1^\downarrow(\mathbf{r}_1^\uparrow) & \cdots & \phi_M^\downarrow(\mathbf{r}_1^\uparrow) \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\phi_1^\uparrow(\mathbf{r}_p^\uparrow) & \cdots & \phi_M^\uparrow(\mathbf{r}_p^\uparrow) & \phi_1^\downarrow(\mathbf{r}_p^\uparrow) & \cdots & \phi_M^\downarrow(\mathbf{r}_p^\uparrow) \\
\phi_1^\uparrow(\mathbf{r}_1^\downarrow) & \cdots & \phi_M^\uparrow(\mathbf{r}_1^\downarrow) & \phi_1^\downarrow(\mathbf{r}_1^\downarrow) & \cdots & \phi_M^\downarrow(\mathbf{r}_1^\downarrow) \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\phi_1^\uparrow(\mathbf{r}_p^\downarrow) & \cdots & \phi_M^\uparrow(\mathbf{r}_p^\downarrow) & \phi_1^\downarrow(\mathbf{r}_p^\downarrow) & \cdots & \phi_M^\downarrow(\mathbf{r}_p^\downarrow)
\end{pmatrix}
\begin{pmatrix} 0 & D \\ -D^T & 0 \end{pmatrix}
\begin{pmatrix}
\phi_1^\uparrow(\mathbf{r}_1^\uparrow) & \cdots & \phi_1^\uparrow(\mathbf{r}_p^\uparrow) & \phi_1^\uparrow(\mathbf{r}_1^\downarrow) & \cdots & \phi_1^\uparrow(\mathbf{r}_p^\downarrow) \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\phi_M^\uparrow(\mathbf{r}_1^\uparrow) & \cdots & \phi_M^\uparrow(\mathbf{r}_p^\uparrow) & \phi_M^\uparrow(\mathbf{r}_1^\downarrow) & \cdots & \phi_M^\uparrow(\mathbf{r}_p^\downarrow) \\
\phi_1^\downarrow(\mathbf{r}_1^\uparrow) & \cdots & \phi_1^\downarrow(\mathbf{r}_p^\uparrow) & \phi_1^\downarrow(\mathbf{r}_1^\downarrow) & \cdots & \phi_1^\downarrow(\mathbf{r}_p^\downarrow) \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\phi_M^\downarrow(\mathbf{r}_1^\uparrow) & \cdots & \phi_M^\downarrow(\mathbf{r}_p^\uparrow) & \phi_M^\downarrow(\mathbf{r}_1^\downarrow) & \cdots & \phi_M^\downarrow(\mathbf{r}_p^\downarrow)
\end{pmatrix}\right] \\
&= (-1)^{p(p-1)/2} \sum_{1 \le j_1 < j_2 < \cdots < j_p \le M} \left(\prod_k \lambda_{j_k}\right) \det \begin{pmatrix}
\phi_{j_1}^\uparrow(\mathbf{r}_1^\uparrow) & \cdots & \phi_{j_1}^\uparrow(\mathbf{r}_p^\uparrow) & \phi_{j_1}^\uparrow(\mathbf{r}_1^\downarrow) & \cdots & \phi_{j_1}^\uparrow(\mathbf{r}_p^\downarrow) \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\phi_{j_p}^\uparrow(\mathbf{r}_1^\uparrow) & \cdots & \phi_{j_p}^\uparrow(\mathbf{r}_p^\uparrow) & \phi_{j_p}^\uparrow(\mathbf{r}_1^\downarrow) & \cdots & \phi_{j_p}^\uparrow(\mathbf{r}_p^\downarrow) \\
\phi_{j_1}^\downarrow(\mathbf{r}_1^\uparrow) & \cdots & \phi_{j_1}^\downarrow(\mathbf{r}_p^\uparrow) & \phi_{j_1}^\downarrow(\mathbf{r}_1^\downarrow) & \cdots & \phi_{j_1}^\downarrow(\mathbf{r}_p^\downarrow) \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\phi_{j_p}^\downarrow(\mathbf{r}_1^\uparrow) & \cdots & \phi_{j_p}^\downarrow(\mathbf{r}_p^\uparrow) & \phi_{j_p}^\downarrow(\mathbf{r}_1^\downarrow) & \cdots & \phi_{j_p}^\downarrow(\mathbf{r}_p^\downarrow)
\end{pmatrix}. \qquad (4.57)
\end{aligned}
$$

The factor of $(-1)^{p(p-1)/2}$ comes from the following Pfaffian identity

$$\mathrm{pf}\begin{pmatrix} 0 & Q \\ -Q^T & 0 \end{pmatrix} = (-1)^{M(M-1)/2}\det(Q)$$

for any matrix $Q \in \mathbb{R}^{M\times M}$. The product of λ_is comes from the determinant of a diagonal matrix, whose elements are selected from the matrix D.

In the limit of $M = p$, all matrices in the Pfaffian become square matrices and the Pfaffian of the geminal matrix becomes

$$\mathrm{pf}(\Phi) = (-1)^{p(p-1)/2}\left(\prod_i \lambda_i\right)\det\left[\phi_i^\alpha(\mathbf{r}_j^\beta; \{\mathbf{r}_{/j}^\beta\}; \{\mathbf{r}^{\bar{\beta}}\})\right] \tag{4.58}$$

which is exactly the dense determinant FermiNet given in Eq. (3.29) up to a prefactor. Hence, the PfaffNet wave function can be written as the Slater FermiNet wave function with one or more dense determinants under certain conditions.

While our discussion so far has been limited to spin-balanced cases, where $N^\uparrow = N^\downarrow = p$, this argument could be applied to spin-imbalanced systems as well. For the scope of our work, we will not discuss it here, but it is straightforward to turn the spin-imbalanced AGP FermiNet wave functions like the one in Eq. (4.17) into FermiNet dense determinants.

4.6 Benchmarking Neural AGP Wave Functions on Beryllium and Nitrogen

In this section, we provide a small set of benchmark results to demonstrate the performance of the Pfaffian FermiNet (PfaffNet) Ansatz with either the full geminal in Eq. (4.24), or the diagonal geminal in Eq. (4.27) and Eq. (4.30). In particular, we benchmark the PfaffNet Ansatz against the Slater FermiNet with one or multiple dense determinants on beryllium and nitrogen.

4.6.1 Beryllium

Firstly, we study the beryllium atom and its dimer molecule. The beryllium dimer, despite having only eight electrons, is notoriously difficult to describe with *ab initio* quantum chemistry methods [12]. One of the difficulties originates from the very close energy levels between the $2s$ and $2p$ orbitals in the beryllium atom, which results in a multi-configurational ground state [67]. Hence, a multi-reference wave function is needed to accurately describe the dissociation of the beryllium dimer. We first use the PfaffNet with one or multiple Pfaffian and compare its results with the Slater FermiNet with various number of dense determinants. The results are shown in Fig. (4.1). The baseline reference energy from Ref. [28] is obtained by combining experimental data and calculations from the configuration interaction (CI) method [153]. From Fig. (4.1),

one can see that the PfaffNet with a diagonal geminal (one Pfaffian) is comparable to the Slater FermiNet with two dense determinants, while the PfaffNet with a full geminal (one Pfaffian) is comparable to the Slater FermiNet with three dense determinants. However, the PfaffNet with both full and diagonal geminals are not able to reach the same level of accuracy as the Slater FermiNet with four or more dense determinants. This is also seen in other systems, where the PfaffNet with a single geminal is able to outperform the Slater FermiNet with one to two dense determinants, but the PfaffNet with multiple geminals is not able to reach the same level of accuracy as the Slater FermiNet with more determinants.

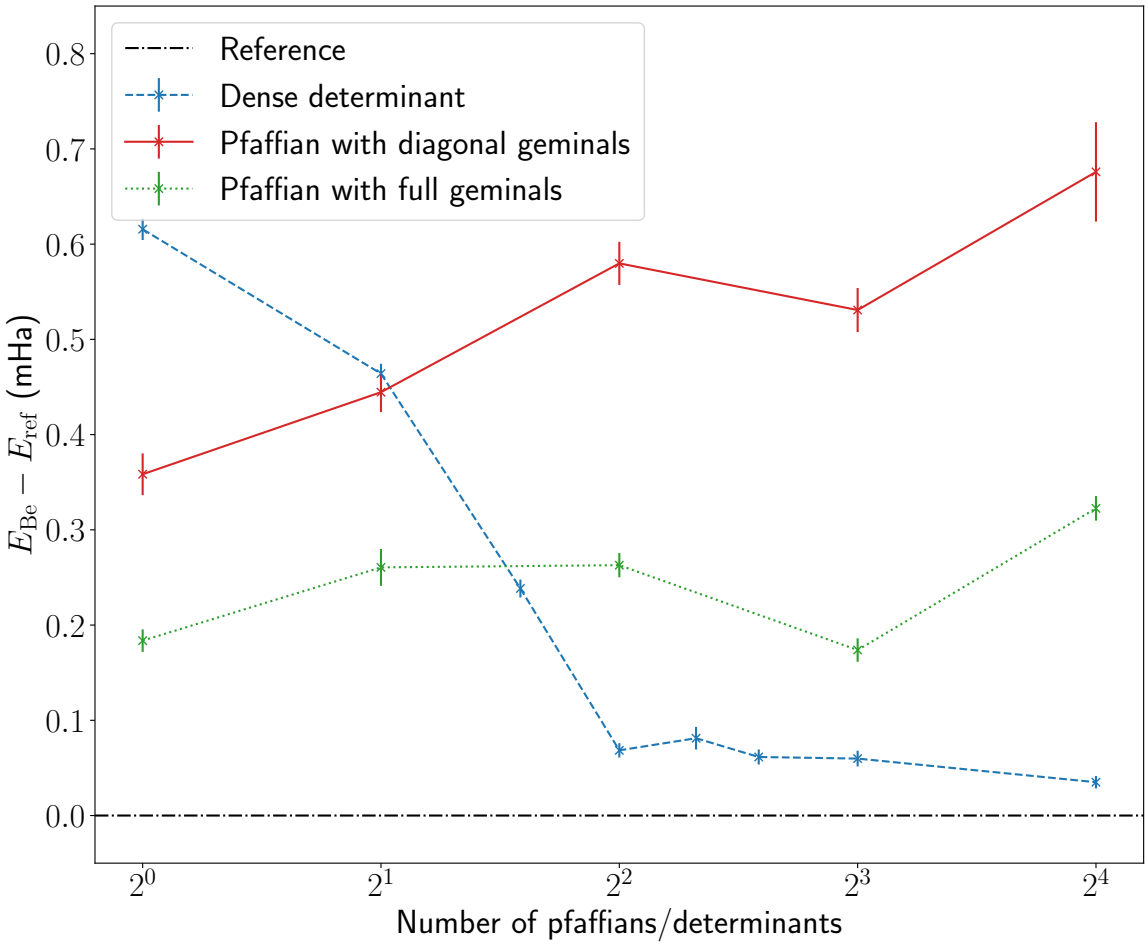


Figure 4.1: Comparison of beryllium atom energy between the PfaffNet with diagonal geminals and full geminals, and the Slater FermiNet with dense determinants. The reference energy of the isolated Be atom is taken from Ref. [28].

Now we compare the beryllium dimer molecule. The results are shown in the Fig. (4.2), together with the full configuration interaction quantum Monte Carlo (FCIQMC) results from

Guther *et al.* [67] [9]. A similar conclusion can be drawn from the dissociation curves of the beryllium dimer, where the two PfaffNet results are between one and 16 dense determinants of the Slater FermiNet.

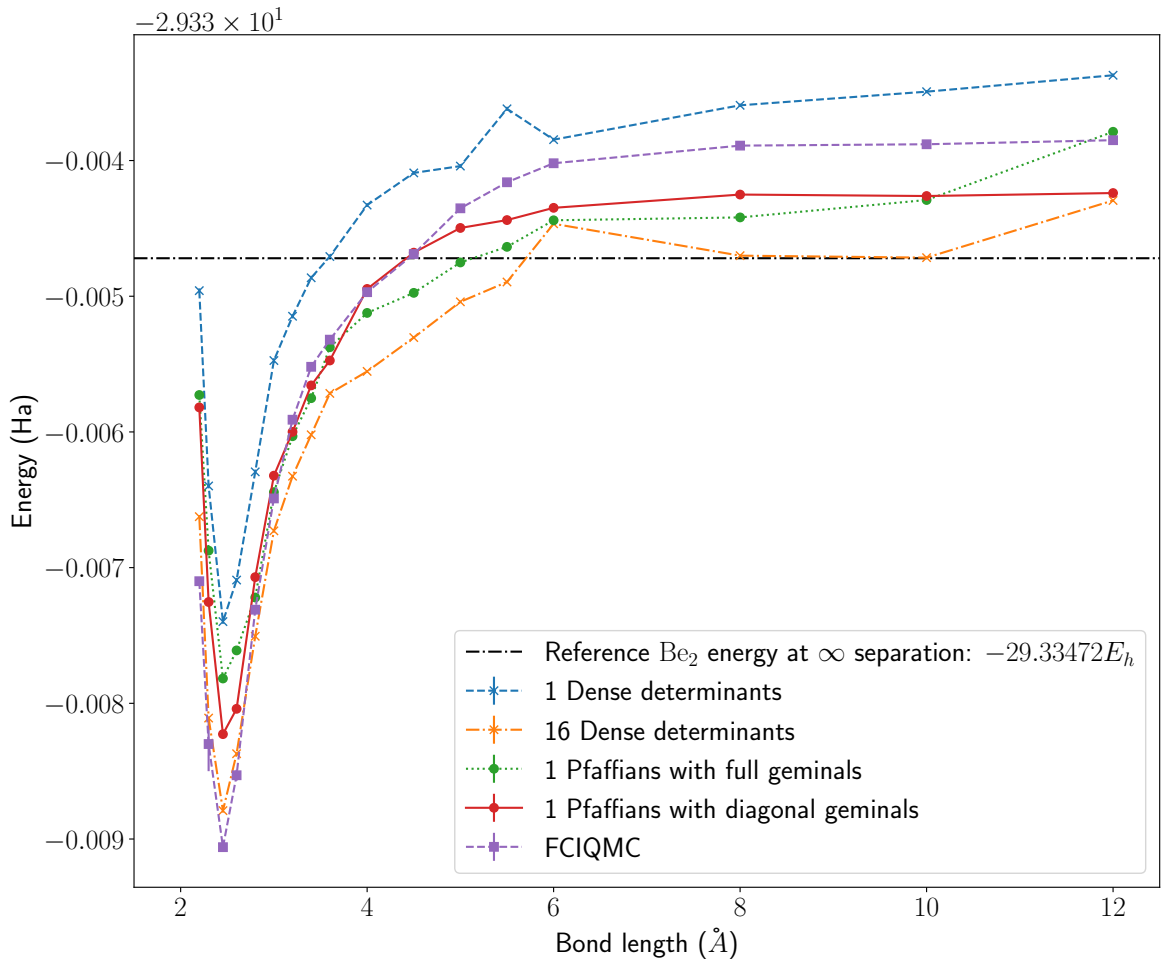
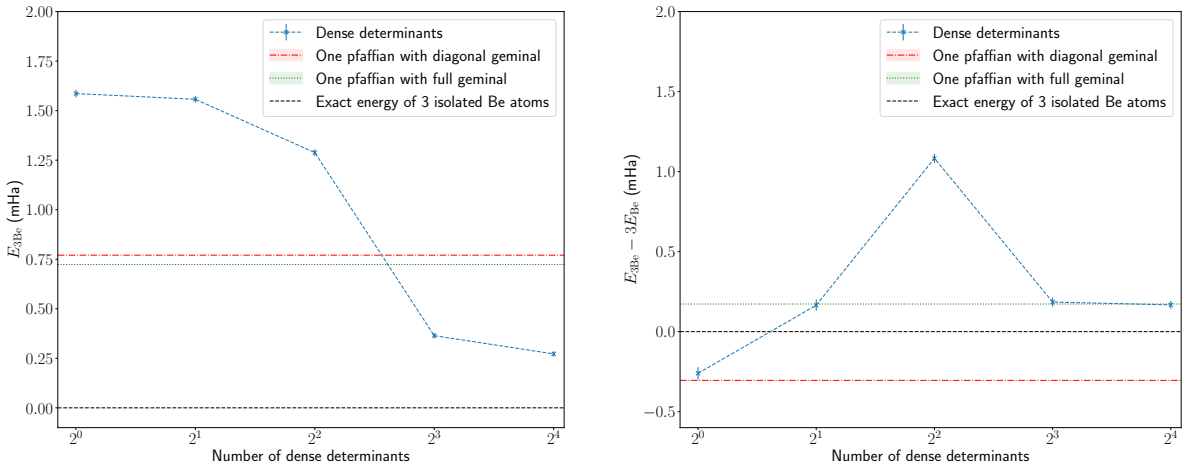


Figure 4.2: Comparison of beryllium dimer energy between the PfaffNet with either one full or one diagonal geminal, and the Slater FermiNet with both 1 and 16 dense determinants. The FCIQMC results are taken from Ref. [67].

Lastly, we test the size-consistency of the PfaffNet Ansatz by comparing the energy of three widely-separated beryllium atoms with three times the energy of one beryllium atom. In quantum chemistry, a method is size-consistent if the calculated total energy of two non-interacting subsystems is equal to the sum of the energy of each of the subsystems, *e.g.*, $E(A+B) = E(A)+E(B)$, where A and B are the subsystems. It is an important property of any quantum chemistry methods if one wish to describe the dissociation of molecules correctly with those methods. We present our results in Fig. (4.3). Firstly, we compare the absolute energy with respect to the

[9] For the FCIQMC results, we have taken the one with the aug-cc-pVTZ basis sets and 68-term Jastrow factor in Ref. [67].



(a) Energy difference with respect to the exact energy of three isolated Be atoms from Ref. [28].

(b) Energy difference between three beryllium atoms and three times the energy of an isolated beryllium atom under the same Ansatz.

Figure 4.3: Comparison of beryllium chain energy between the PfaffNet with either one full or one diagonal geminal, and the Slater FermiNet with 1, 2, 4, 8 and 16 dense determinants. The reference energy of the isolated Be atom is taken from Ref. [28].

exact energy of three isolated beryllium atom in Fig. (4.3a), we see that the PfaffNet with both a full and a diagonal geminal are able to outperform the Slater FermiNet up to 4 dense determinants, but slightly higher than the Slater FermiNet with 8 and 16 dense determinants. This is, again, consistent with the results of the beryllium atom and dimer molecule.

Secondly, to study the size-consistency of the Ansätze, we compare the energy difference between three beryllium atoms and three times the energy of an isolated beryllium atom under the same Ansatz in Fig. (4.3b). If the Ansatz is size-consistent, one would expect the quantity to be zero. Note that because we are computing energy difference here, the results are no longer variational. As seen on Fig. (4.3b), all Ansätze are less than 0.5mHa from zero except for the Slater FermiNet with 4 dense determinants, which could be caused by optimisation issues. This suggested that the PfaffNet Ansatz has a similar size-consistency error as the Slater FermiNet with dense determinants.

4.6.2 Nitrogen

Next, we study the nitrogen atom and its dimer molecule, where the latter is also a textbook example of a multi-configurational system. We carried out a similar comparison as in the beryllium atom, where we compared the energies of the nitrogen atom obtained from PfaffNet with both the full and the diagonal geminals, as well as the Slater FermiNet with dense determinants. The results are shown in the Fig. (4.4), plotted against the number of determinants/Pfaffians used. It can be seen that the difference between PfaffNet and the Slater FermiNet at one and

two determinants/Pfaffians is more pronounced in the nitrogen atom than in the beryllium atom. While the PfaffNet with both geminals has a similar accuracy across the whole range of determinants/Pfaffians, the Slater FermiNet is not able to reach the same level of accuracy as the PfaffNet until four or more dense determinants are used.

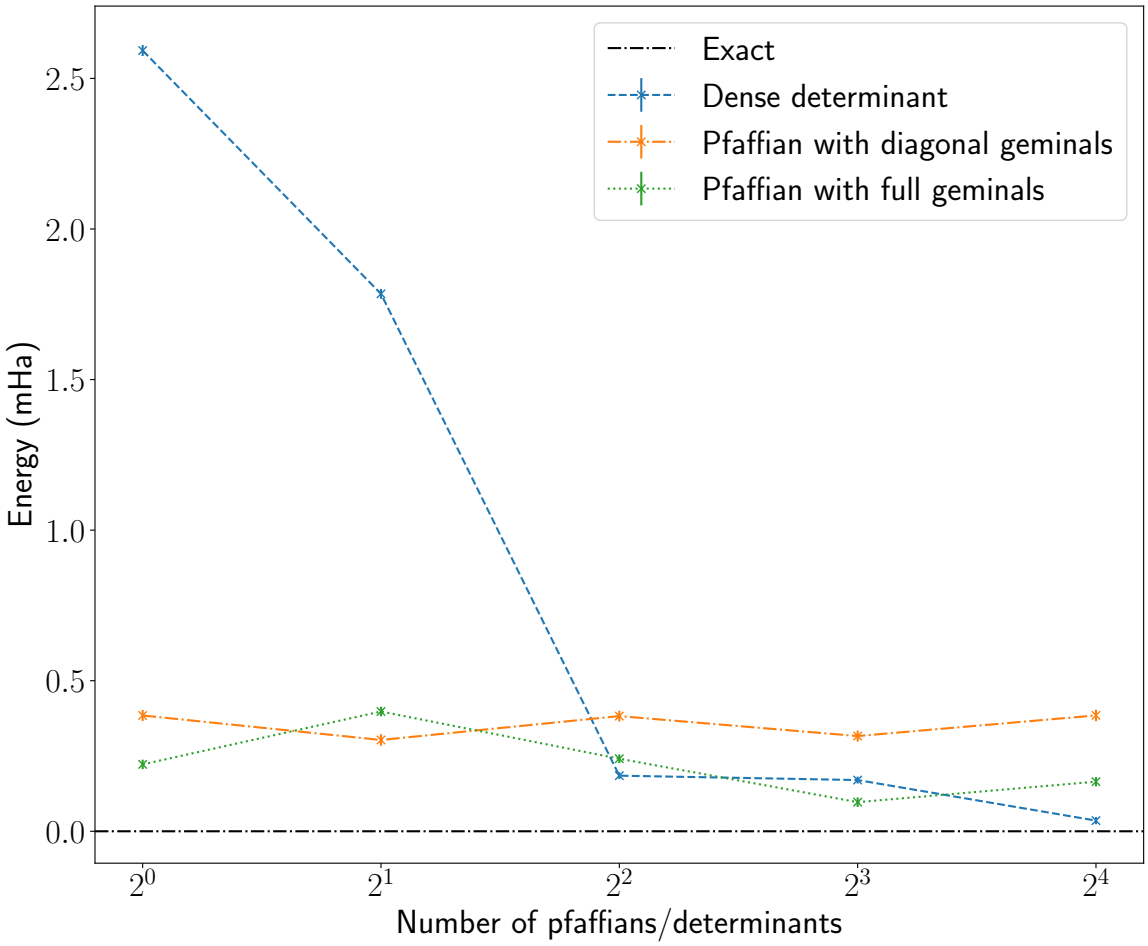


Figure 4.4: Comparison of nitrogen atom energy between the PfaffNet with either one full or one diagonal geminal, and the Slater FermiNet with different number of dense determinants.

Next, we look at the nitrogen dimer molecule, which is also notoriously difficult to describe as it requires multi-reference wave functions. The results are shown in the Fig. (4.5). A similar conclusion can be drawn from the dissociation curves of the nitrogen dimer, where the two PfaffNet results are either better or comparable to the Slater FermiNet with one dense determinant, but significantly higher than the best FermiNet energy, which is obtained with 32 dense determinants.

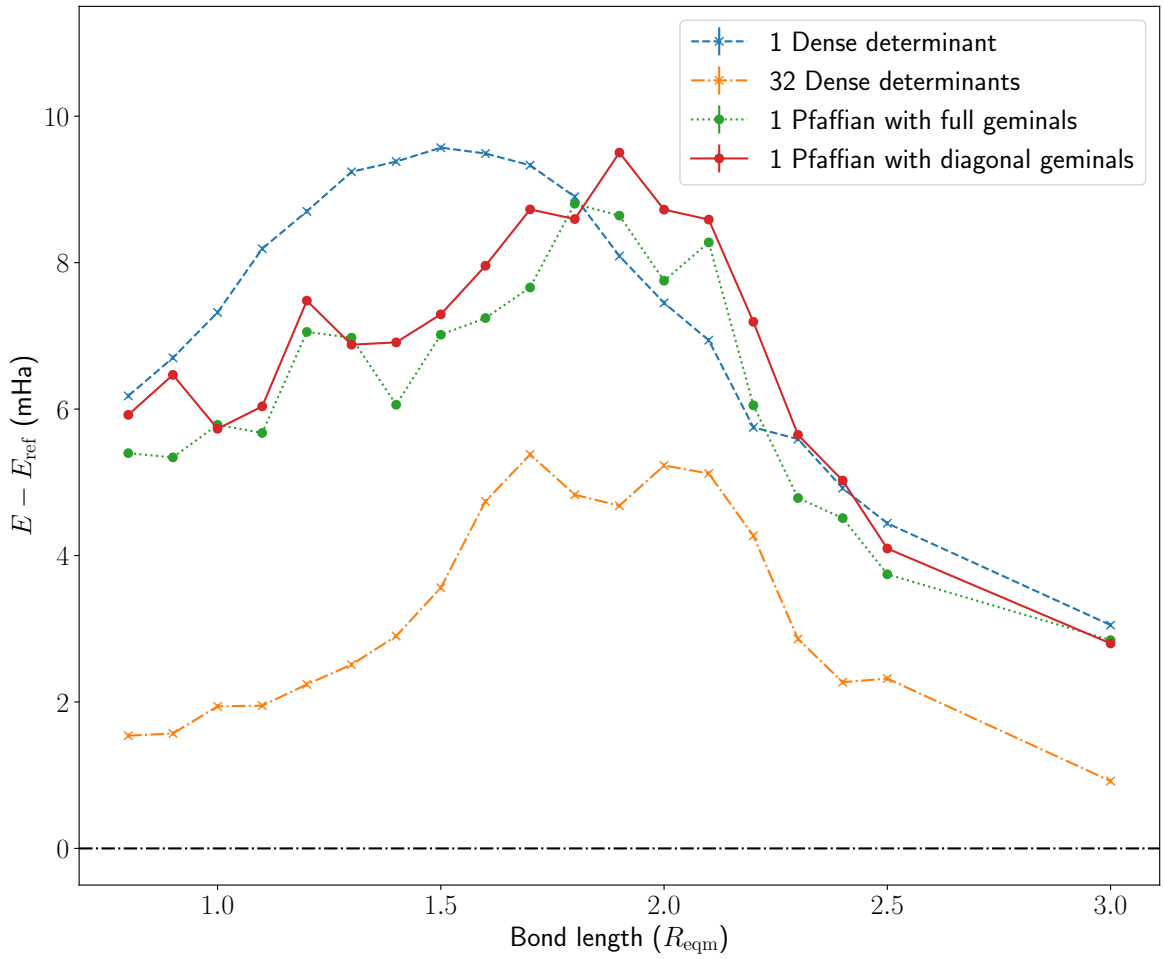


Figure 4.5: Full dissociation curve of the nitrogen dimer molecule. Energy is displayed relative to the spectroscopy value.

4.6.3 Discussion

Overall, our investigation of the PfaffNet Ansätze on the beryllium atom and its dimer showed promising results, where a single-geminal PfaffNet often matched or outperformed the Slater FermiNet using one to three dense determinants. The size-consistency error of the PfaffNet also appeared comparable to the FermiNet with dense determinants, exhibiting small errors across separated beryllium atoms. Similar trends were observed for the nitrogen atom, where the PfaffNet shows higher accuracy until the FermiNet used four or more dense determinants. However, while the PfaffNet with one Pfaffian shows a clear advantage in accuracy compared to the Slater FermiNet with one dense determinant, the performance of the former failed to scale effectively as the number of Pfaffians in the multi-Pfaffian expansion increases, consistently falling short of the best results achieved by the Slater FermiNet with multiple dense determinants. Hence, we conclude that the current PfaffNet Ansätze does not show any clear, scalable

advantages against the Slater FermiNet on isolated atoms and molecules.

Chapter 5

Unitary Fermi Gas

The unitary Fermi gas (UFG) is a paradigmatic example of a strongly interacting system of two-component fermions that possesses a superfluid ground state and lies in the crossover region between a Bardeen-Cooper-Schrieffer (BCS) superconductor and a Bose-Einstein condensate [88, 110]. The effective range of the interaction is zero and the s-wave scattering length diverges (the "unitarity limit"), so the UFG has no intrinsic length scale. The only remaining length is the inverse of the Fermi wavevector, $1/k_F$, on which all thermodynamic quantities depend. For example, regardless of the particle density, the ground-state energy per particle of a unitary Fermi gas can be written as

$$E = \xi E_{\mathrm{FG}} = \xi \frac{3}{5} \frac{\hbar^2 k_F^2}{2m}, \tag{5.1}$$

where E_{FG} is the energy per particle of a non-interacting Fermi gas of the same density. The dimensionless constant ξ is known as the Bertsch parameter [114].

Because of the universality of the UFG model, it can be used to describe many real physical systems at different scales, such as the neutron matter in the inner crust of a neutron star [52] or the quantum criticality of an s-wave atomic superfluid [108, 109]. The size of the pairs in the UFG is comparable to the inter-particle spacing, which is also a feature of many high-T_c superconductors [129, 131, 152]. As a result, the UFG has been studied extensively [60]. Although the UFG is an idealised model, it can be accurately realised in the laboratory using ultracold atomic gases in which the interactions have been tuned by using an external magnetic field to drive the system across a Feshbach resonance [34].

The UFG has been studied for decades, but it remains difficult to calculate its ground-state properties accurately using analytic methods. Mean-field treatments such as BCS theory [8] give good results for systems with weak interactions, but there is no guarantee of success in the strongly interacting regime. As a result, various quantum Monte Carlo (QMC) methods [21, 49] have been used to simulate the properties of the UFG to high accuracy at zero and finite temperature. Methods used include variational Monte Carlo (VMC), fixed-node diffusion Monte Carlo (FN-DMC), fixed-node Green function Monte Carlo, auxiliary field Monte Carlo

and diagrammatic Monte Carlo [22, 30, 31, 47, 59, 64, 72, 81, 105, 126, 147]. However, a full quantitative description remains an open and challenging problem.

Recent advances in machine learning algorithms and the growing availability of inexpensive GPU-based computational resources have allowed neural-network-based approaches to permeate many areas of computational physics, including lattice [19, 35, 75, 96] and continuum [58, 73, 117, 121] QMC simulations. Here we employ a neural network Ansatz within a VMC approach to study the unitary Fermi gas. The Ansatz we use, the Fermionic Neural Network (FermiNet) [121], gives very accurate results for atoms and molecules [91, 121, 149, 166] and has recently been applied to periodic solids and the homogeneous electron gas (HEG) with comparable success [90]. In the case of the HEG, the variational optimisation of the FermiNet Ansatz discovered the quantum phase transition between the Fermi liquid and Wigner crystal ground states without external guidance [24]. In contrast, previous approaches required different Ansätze to be used for the two different phases. The FermiNet has not previously been applied to fermionic superfluids such as the UFG.

We find that the original FermiNet Ansatz is insufficient to capture the two-particle correlations of superfluids. Although a FermiNet wave function with one determinant and a sufficiently large neural network is in principle able to represent any fermionic state [121], it is often advantageous to use a network of a fixed size and a small linear combination of FermiNet determinants. In the case of the unitary Fermi gas, however, we find that the number of block-diagonal (BD) determinants required to describe the ground state accurately scales exponentially with the system size. This is the first example in which the FermiNet has been seen to fail both quantitatively and qualitatively, and suggests that the FermiNet wave function may not be able to represent arbitrary fermionic wave functions in practice. To remedy the problem, we utilise the neural-network part of the FermiNet architecture to build a different type of wave function based on the idea of an antisymmetric geminal power singlet wave function (AGPs) [7, 14, 26, 57, 96], which we discussed in detail in Chapter 3. This leads to substantial improvements, even though the neural-network part of the wave function remains unchanged.

5.1 Results

The power of the AGPs FermiNet Ansatz may be demonstrated by studying the UFG. The Hamiltonian is

$$\hat{H} = -\frac{1}{2}\sum_{i}^{N} \nabla_i^2 + \sum_{ij}^{N/2} U(\mathbf{r}_i^{\uparrow} - \mathbf{r}_j^{\downarrow}), \tag{5.2}$$

where

$$U(\mathbf{r}) = -\frac{2v_0\mu^2}{\cosh^2(\mu r)} \tag{5.3}$$

is the modified Pöschl-Teller potential, which is widely used in variational and diffusion QMC simulations [22, 30, 31, 47, 59, 72, 105, 147]. It would be preferable to use a delta function

interaction with an infinite s-wave scattering length, but it is difficult to simulate systems with delta-like potentials using QMC methods. Thus, a finite but short-ranged interaction is typically used. The s-wave scattering length of the Pöschl-Teller potential diverges when $v_0 = 1$. By changing the value of μ at fixed $v_0 = 1$, it is possible to vary the effective range of the interaction, $r_e = 2/\mu$, whilst holding the s-wave scattering length infinite.

We choose to study a system with density parameter $r_s = 1$, where r_s, the radius of a sphere that contains one particle on average, provides a convenient measure of the inter-particle distance. Throughout this work, we employ the dimensionless system based on Hartree atomic units: the unit of length is the Bohr radius, a_0, and the unit of energy is the Hartree. To ensure that the range of the interaction is small compared with the inter-particle separation, we set $\mu = 12$ ($r_e = 1/6$), keeping $v_0 = 1$ to ensure that the scattering length remains infinite [1]. We have also simulated the system with $k_F = 1$ (equivalent to $r_s = (9\pi/4)^{1/3} \approx 1.92$) and $\mu = 12$ to compare with the fixed-node diffusion Monte Carlo (FN-DMC) result from Forbes *et al.* [47].

5.1.1 Slater FermiNet and AGPs FermiNet

We use both the Slater FermiNet with multiple block-diagonal determinants and the AGPs FermiNet with multiple geminals to study the unitary Fermi gas from $N = 4$ to $N = 38$ particles in a cubic box subject to periodic boundary conditions. In addition, we use the AGPs FermiNet to study a larger system containing $N = 66$ particles. [2]. The same network size, number of determinants, and number of training iterations are used for both Ansätze. Unless otherwise stated, all calculations used a linear combination of 32 determinants or 32 geminals. The FermiNet orbitals are given by Eq. (3.25) without the bias term. The FermiNet geminal used for systems containing from 4 to 28 particles is the one defined in Eq. (4.30). Including contributions from the two-electron stream improves the optimisation rate and can achieve a slightly lower variational energy in larger systems, so Eq. (4.31) was used for systems with $N \geq 29$. The inclusion of plane-wave envelopes as defined in Eq. (3.34) also improves the optimisation rate. For molecular and electron gas systems, we have found that the bias term in the FermiNet orbital projection (Eq. 3.25) does not affect the accuracy or optimisation of the model. We hence set the bias term, $g_i^{k\alpha}$ and g^k as appropriate, to zero for all calculations presented here.

A comparison of the ground-state energy expectation values given by the two Ansätze is shown in Fig. 5.1a. The Slater FermiNet, which consists of a linear combination of block-diagonal determinants of FermiNet orbitals, performs well when the number of particles N is smaller than

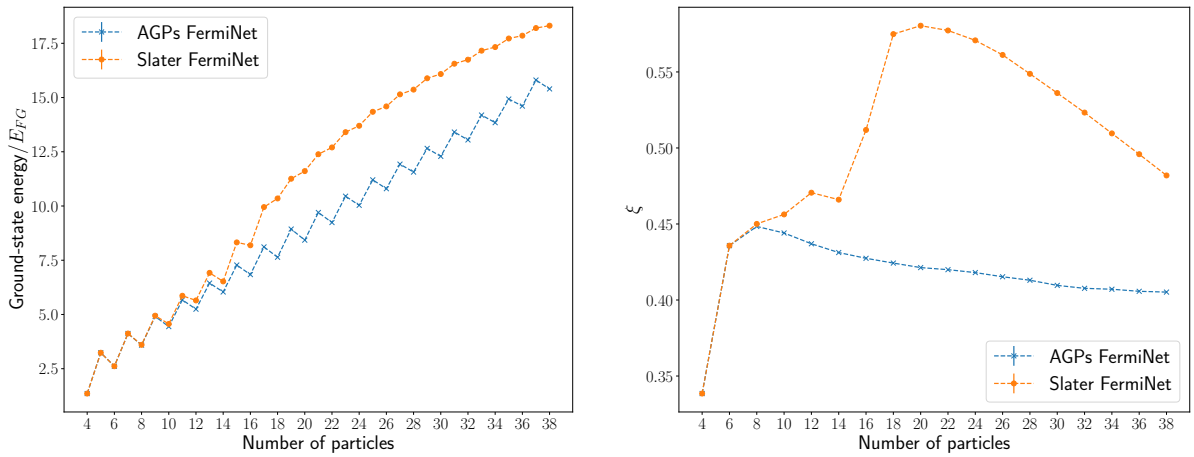
[1] Although the range of the potential is finite, which introduces a second length scale even at unitarity, most of the results in the literature have been obtained with the same value of μ. This allows us to compare our results with theirs. To obtain exact results, one has to study the limit as $\mu \to \infty$, but converging the simulations becomes more difficult as μ increases.

[2] As previous works [20, 46, 72, 105] have shown, the ground-state energy of the $N = 66$ particle UFG is the closest to the ground-state energy in the thermodynamic limit even among larger systems. This can also be seen by comparing the ratio of E_{FG} and the energy per particle from summing over the energies occupied by the particles in the discretised k-space, where this ratio is just 0.5% away from unity for $N = 66$.

around 10, but the AGPs FermiNet is much superior in larger systems. It is clear that the Slater FermiNet Ansatz has difficulties learning the ground states of large paired systems.

In systems containing an odd number of particles, one must be left unpaired. This raises the energy a little and explains the zigzag shape of Fig. 5.1a. The odd-even staggering is lost for larger systems with the Slater FermiNet Ansatz, indicating the absence of pair formation [22, 112]. The Slater FermiNet fails to learn the superfluid state. For the AGPs FermiNet, by contrast, the amplitude of the odd-even zigzag remains constant, superposed on the linear increase with N expected of any extensive quantity.

Another comparison between the two Ansätze is shown in Fig. 5.1b, which depicts the ratio of the interacting and non-interacting energies per particle, known as the Bertsch parameter [114] and defined in Eq. (5.1), as a function of N. All FermiNet energies are variational and the non-interacting energies are exact, so the AGPs FermiNet, for which the Bertsch parameter is lower by up to around 30%, is much better of the two Ansätze.



(a) The total energy of the UFG simulation cell, measured in units of the free Fermi gas energy E_{FG}. The Slater FermiNet Ansatz begins to fail when $N \gtrapprox 10$.

(b) The Bertsch parameter ξ (the ratio of the interacting and non-interacting ground-state energies per particle) as a function of the number of particles N.

Figure 5.1: Comparison between results obtained using the AGPs FermiNet and the Slater FermiNet for different numbers of particles, N, with $r_s = 1$ and $\mu = 12$. All simulations used 32 determinants, 300,000 optimisation steps, and the same hyperparameters.

Dense Determinants

Previous works [24, 53, 58, 93, 132, 162] have suggested that the use of dense determinants in Eq. (3.29) in contrast to block-diagonal determinants with FermiNet in Eq. 3.28, provide a slight gain in accuracy in various infinite and periodic systems. In this subsection, we present a small set of calculations of the UFG with dense determinants and compare the results with the block-diagonal FermiNet and the AGPs FermiNet. The results are presented in Table 5.1. Although

Type of Wave Function	Total Energy [E_{FG}] $N = 36$	$N = 37$	$N = 38$	Pairing Gap [E_{FG}]
Block (BD)	17.8523(5)	18.2032(5)	18.3131(6)	0.121(2)
Dense (DD)	17.7027(5)	18.1601(5)	18.2570(5)	0.180(2)
AGPs	14.6059(4)	15.8060(5)	15.3975(5)	0.804(1)

Table 5.1: Total energy of the UFG with 36 to 38 particles and their corresponding pairing gap using different wave functions.

the use of dense determinants in the UFG does provide lower energies comparing to the block-diagonal determinants, they are still significantly higher than the AGPs FermiNet energies. In addition, the qualitative behaviours, such as the absence of odd-even staggering, are still similar to the block-diagonal FermiNet, which are qualitatively different from the AGPs FermiNet.

5.1.2 AGPs FermiNet and FN-DMC

We next compare our results with the state-of-the-art FN-DMC results of Forbes *et al.* [47], shown in Fig. 5.2 for the case $k_F = 1$ and $\mu = 12$. The AGPs FermiNet achieves a lower energy per particle than FN-DMC for all system sizes except for $N = 4$ and $N = 6$. The dependence of the Bertsch parameter on system size is also smoother when calculated with the AGPs FermiNet [3].

The pairing gap may be found using the approximation formula for finite N [21, 112]:

$$\Delta = (-1)^N \left[E(N+1) - \frac{1}{2}\left[E(N) + E(N+2)\right] \right], \tag{5.4}$$

where N is the total number of particles in the box. The results from $N = 4$ to $N = 36$ are shown in Fig. 5.3. Also shown is the thermodynamic ($N \to \infty$) limit of the BCS pairing gap including Gorkov's polarisation correction [63]:

$$\Delta_{\mathrm{BCS}} = \frac{8}{e^2}\frac{\hbar^2 k_F^2}{2m}\exp\left(\frac{\pi}{2k_F a}\right), \tag{5.5}$$

$$\Delta_{\mathrm{Gorkov}} = \frac{1}{(4e)^{1/3}}\Delta_{\mathrm{BCS}}. \tag{5.6}$$

Here a is the scattering length of the interaction, which is infinite in the UFG. In this limit, $\Delta_{\mathrm{BCS}} = 1.804 E_{\mathrm{FG}}$ and $\Delta_{\mathrm{Gorkov}} = 0.815 E_{\mathrm{FG}}$, where $E_{\mathrm{FG}} = \frac{3}{5}\frac{\hbar^2 k_F^2}{2m}$ is the average energy per particle of an unpolarised non-interacting Fermi gas and e is Euler's number [4]. The UFG is a

[3] Limited by the scope of this work, we did not proceed to carry out a similar set of calculations done by Forbes et al. [47] for extrapolating the results to $\mu \to \infty$.

[4] Although writing the pairing gap in units of the Fermi energy would be more suitable as it is a one-body quantity, we choose to express it in units of E_{FG} instead, because:

(i) we want to be consistent with the units used in the paper by J. Carlson *et al.* [22], which also used E_{FG} as

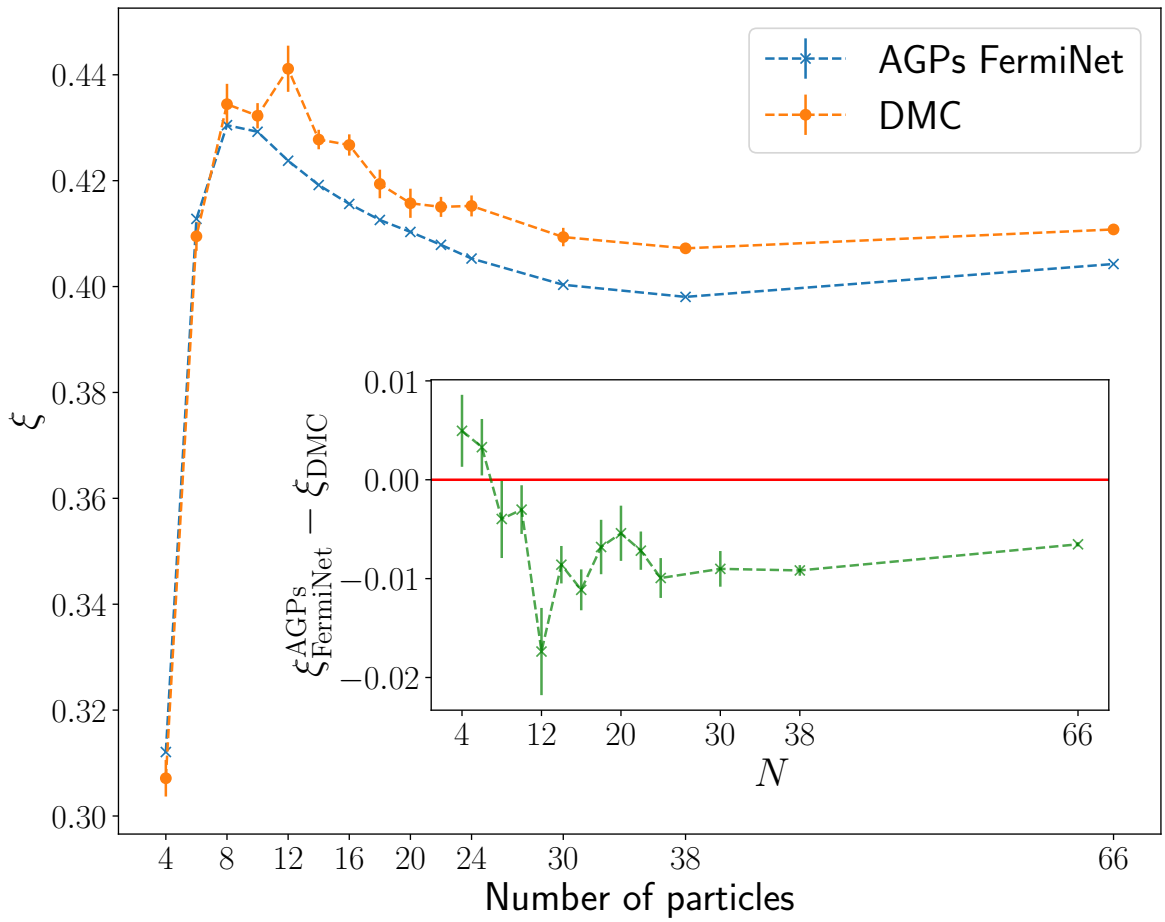


Figure 5.2: Comparison of the system-size dependent values of the Bertsch parameter, ξ, as calculated using the AGPs FermiNet and FN-DMC, with $k_F = 1$ and $\mu = 12$. According to the variational principle, lower values are better. The error bars on the AGPs FermiNet results are smaller than the sizes of the crosses. Inset: difference between the AGPs and FN-DMC values of the Bertsch parameter. The errors in the inset are obtained by adding the standard errors of the FN-DMC and AGPs FermiNet results in quadrature. The latter are obtained by computing the standard error of the MCMC-averaged Bertsch parameter accumulated over 50,000 inference steps.

strongly coupled system, so the BCS and Gorkov estimates of the gap need not be accurate.

The striking collapse of the pairing gap with increasing system size shows that the Slater FermiNet Ansatz struggles to describe paired states in systems of more than 10 particles. The AGPs FermiNet Ansatz behaves much better, although the oscillations with system size suggest that significant finite-size errors remain even for the largest systems simulated.

the energy unit for the pairing gap, and
(*ii*) to keep the units consistent with other quantities in this work.

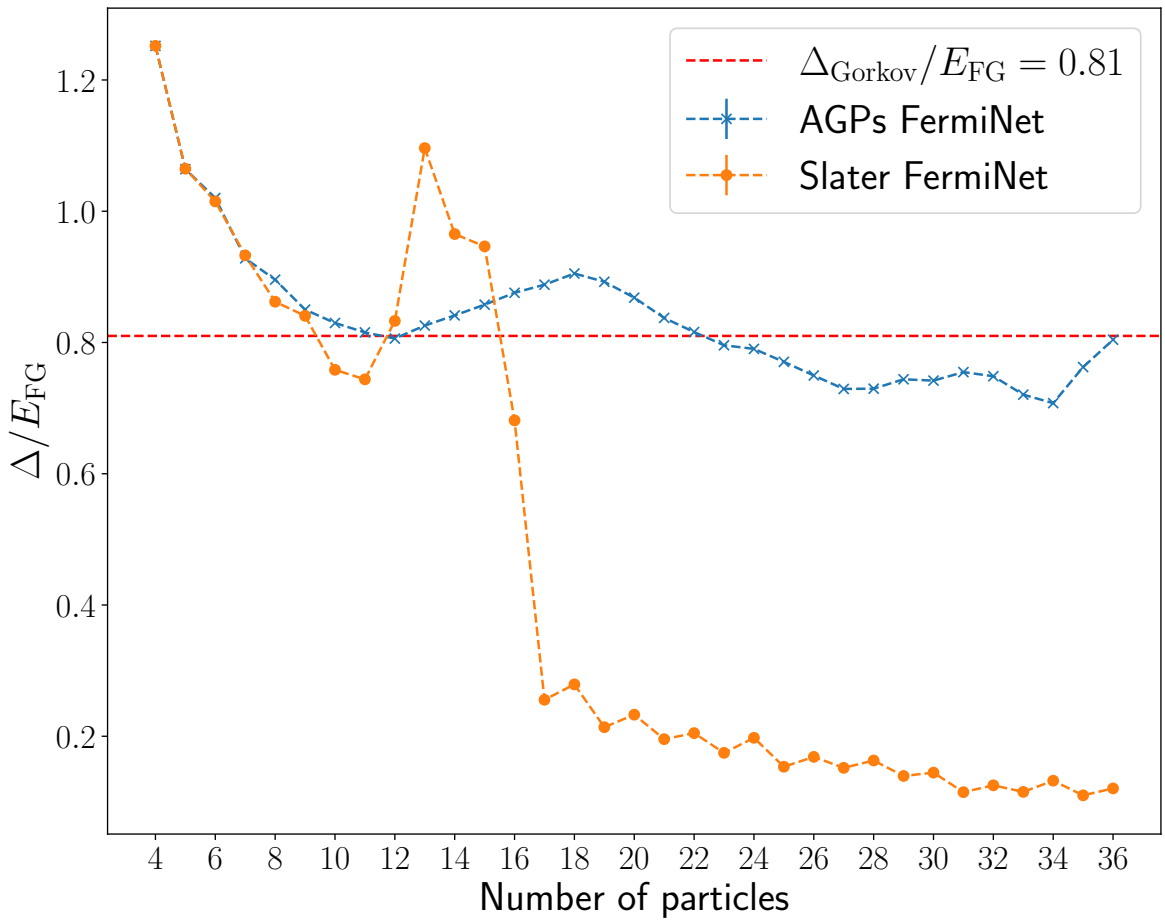


Figure 5.3: Pairing gaps calculated with the Slater FermiNet and the AGPs FermiNet for different numbers of particles N, with $r_s = 1$ and $\mu = 12$.

5.1.3 Condensate Fraction

Another signature of fermionic superfluidity is the presence of off-diagonal long-ranged order in the two-body density matrix (TBDM) in Eq. (2.77), the largest eigenvalue of which diverges as the number of particles N tends to infinity [168]. The superfluid condensate fraction c may be obtained by evaluating the quantity in Eq. (2.84) [107], which requires the evaluation of the translationally and rotationally averaged TBDM in Eq. (2.83) in the limit where $r \to \infty$. The one-body density matrix, by contrast, tends to zero in the $r \to \infty$ limit [168]. A full discussion of the methods used to evaluate the condensate fraction in QMC simulations can be found in Appendix B and the CASINO manual [107].

After fully training both the Slater FermiNet and the AGPs FermiNet for the $N = 38$ particle system, we used the resulting neural wave functions to compute the quantity $\frac{\Omega^2}{N_\uparrow}\rho^{(2)TR}_{\uparrow\downarrow}(r)$. The results are shown in Fig. 5.4, which provides further evidence that the Slater FermiNet fails to converge to the superfluid ground state; the quantity $\frac{\Omega^2}{N_\uparrow}\rho^{(2)TR}_{\uparrow\downarrow}(r)$ appears to be approaching zero

in the large pair-separation limit, implying that the condensate fraction is also zero. The same quantity for the AGPs FermiNet approaches a finite value which we estimated to be roughly $c = 0.44(1)$ using the eight data points with separations $r/r_s \geq 2.0$. This value is consistent with previous estimates from experiments and the most recent AFMC value from [72] (Table 5.2).

In addition, we also estimated the condensate fractions for the $N = 66$ UFG at a fixed $\mu = 12$ with two different densities: $r_s = 1$ ($k_F r_e = 0.32$) and $k_F = 1$ ($k_F r_e = 0.17$), respectively. We compute the quantity $\frac{\Omega^2}{N_\uparrow}\rho^{(2)TR}_{\uparrow\downarrow}(r)$ at five sequentially-spaced separations r near $r = L/2$, where the quantity has approached its asymptotic value. We then take the average of the five data points to get estimated values of the condensate fraction. [5]. Our estimate of the condensate fraction for $N = 66$ is $c = 0.42(1)$ at $k_F r_e = 0.32$, and $c = 0.52(1)$ at $k_F r_e = 0.17$, which are both within error bars of the experimental results. The results are summarised in Table 5.2.

Method	Value
Our estimate for $N = 38$ at $k_F r_e = 0.32$	0.44(1)
Our estimate for $N = 66$ at $k_F r_e = 0.32$	0.42(1)
Our estimate for $N = 66$ at $k_F r_e = 0.17$	0.52(1)
FN-DMC for $N = 38$ at $k_F r_e = 0.03$ [4]	0.61(2)
FN-DMC for $N = 66$ at $k_F r_e = 0.03$ [4]	0.57(2)
FN-DMC for $N = 128$ with VMC extrapolation at $k_F r_e = 0.32$ [105]	0.51
FN-DMC with $k_F r_e \to 0$ extrapolation for $N = 66$ [92]	0.56(1)
AFMC with $k_F r_e \to 0$ extrapolation for $N = 66$ [72]	0.43(2)
Experiment [173]	0.46(7)
Experiment [86]	0.47(7)

Table 5.2: Estimates of the superfluid condensate fraction at unitarity using various methods. The quantity $k_F r_e$ is a dimensionless number, indicating the deviation of the simulated system from a perfect UFG with zero-range interaction.

Although VMC methods are generally considered to be less accurate than FN-DMC methods, an important advantage of VMC methods is that almost any expectation value, including any reduced density matrix, may be estimated without bias. The same is not true of FN-DMC simulations, which sample the wave function instead of its square modulus and produce biased "one-sided" estimates of the expectation values of operators that do not commute with the Hamiltonian [49]. Thus, there are very few unbiased and accurate first-principles calculations of the condensate fraction. Our approach, having both the advantages of VMC and surpassing the accuracy of DMC, provides solutions to these problems and a more accurate way to estimate general expectation values.

[5] Due to limited resources, we only compute five data points at far separations instead of the whole range of separations from $r = 0$ to $r = L/2$ for $N = 66$.

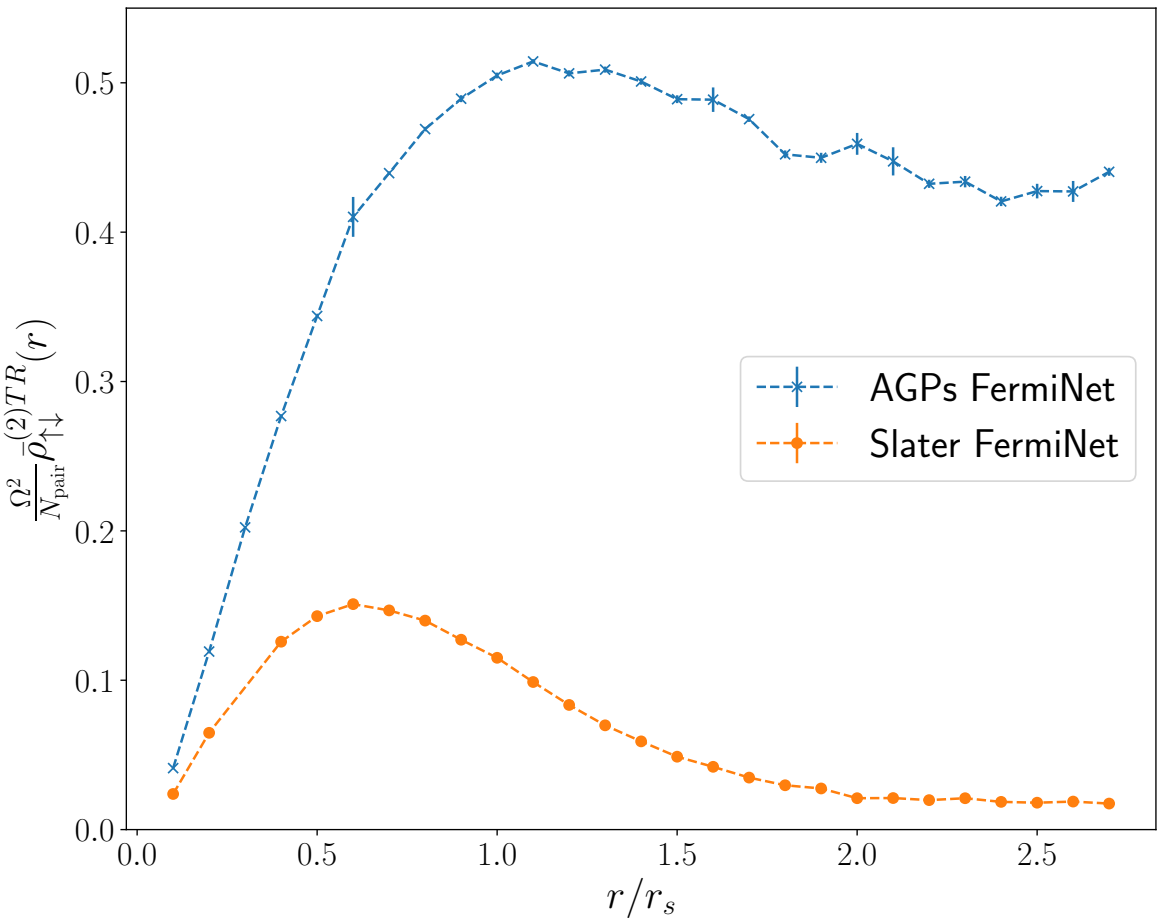


Figure 5.4: Comparison of the TBDM estimators calculated using the AGPs FermiNet and the Slater FermiNet with $N = 38$, $r_s = 1$ and $\mu = 12$. The error bars show the standard error of the TBDM estimator, accumulated over 2,000 inference steps. Most of the error bars are so small that they are obscured the symbols.

5.1.4 Scaling behaviours between the two Ansätze

Finally, we study how the number of block-diagonal determinants required to achieve a given accuracy scales with the number of particles in the system.

In the previous sections, we have demonstrated that the Slater FermiNet with 32 block-diagonal determinants is able to capture superfluidity in small systems but fails at larger systems. Therefore, we are interested in the scaling of the original block-diagonal determinant FermiNet wave function with respect to the system size, *i.e.* how many block-diagonal determinants are required for the Slater FermiNet to converge to the ground state at larger system sizes?

To answer this question, we choose six even-particle systems from $N = 4$ to $N = 14$ and compare the energies obtained using the Slater FermiNet with multiple block-diagonal determinants against energies obtained using a *single-determinant* (and thus single-geminal) AGPs FermiNet

wave function. First, we set the AGPs FermiNet with one determinant energies as baselines and plot the percentage difference of the Slater FermiNet from the baseline against the number of block-diagonal determinants used in the Slater FermiNet wave function, repeated at each system size from 4 to 14 even particles systems. The results are shown in Fig. 5.5.

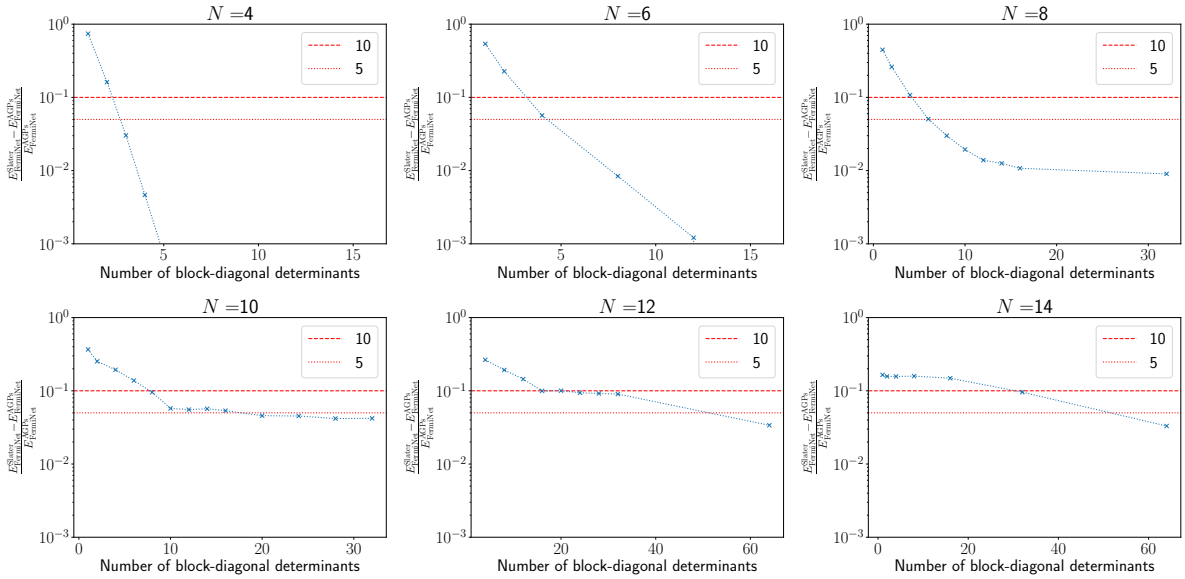


Figure 5.5: Percentage difference between the Slater FermiNet with various number of block-diagonal determinants and the AGPs FermiNet with one determinant at different system sizes, from 4 to 14 particles.

Note that as system size increases, it becomes more difficult for the Slater FermiNet to get close to the AGPs FermiNet baseline with a feasible number of determinants. Hence, due to the constraints on resources, it is not realistic to continue to increase the number of determinants until the Slater FermiNet achieves the same accuracy as the AGPs FermiNet. Instead, we decided to set two thresholds for the percentage difference between the two results. Here, we have used 5% and 10% as the thresholds, plotted as the two horizontal lines in Fig. 5.5. By plotting the x-intercept of the curves with the two threshold lines against system size, we can determine the relationship between the two, as shown in Fig. 5.6. As the y-axis in Fig. 5.6 is in logarithmic scale, a roughly linear relationship suggests an exponential scaling of the number of block-diagonal determinants as system size increases. This suggests that the number of block-diagonal determinants required to achieve a given percentage accuracy increases approximately exponentially with the number of particles.

These results suggest that the Slater FermiNet wave functions constructed using a neural network of fixed size is in fact capable of describing the ground state of the UFG accurately given the number of block-diagonal determinants is sufficiently large. However, the number of block-diagonal determinants required to learn the ground state increase *exponentially* as system size gets bigger and the number becomes inaccessible as the number of particles exceeds around

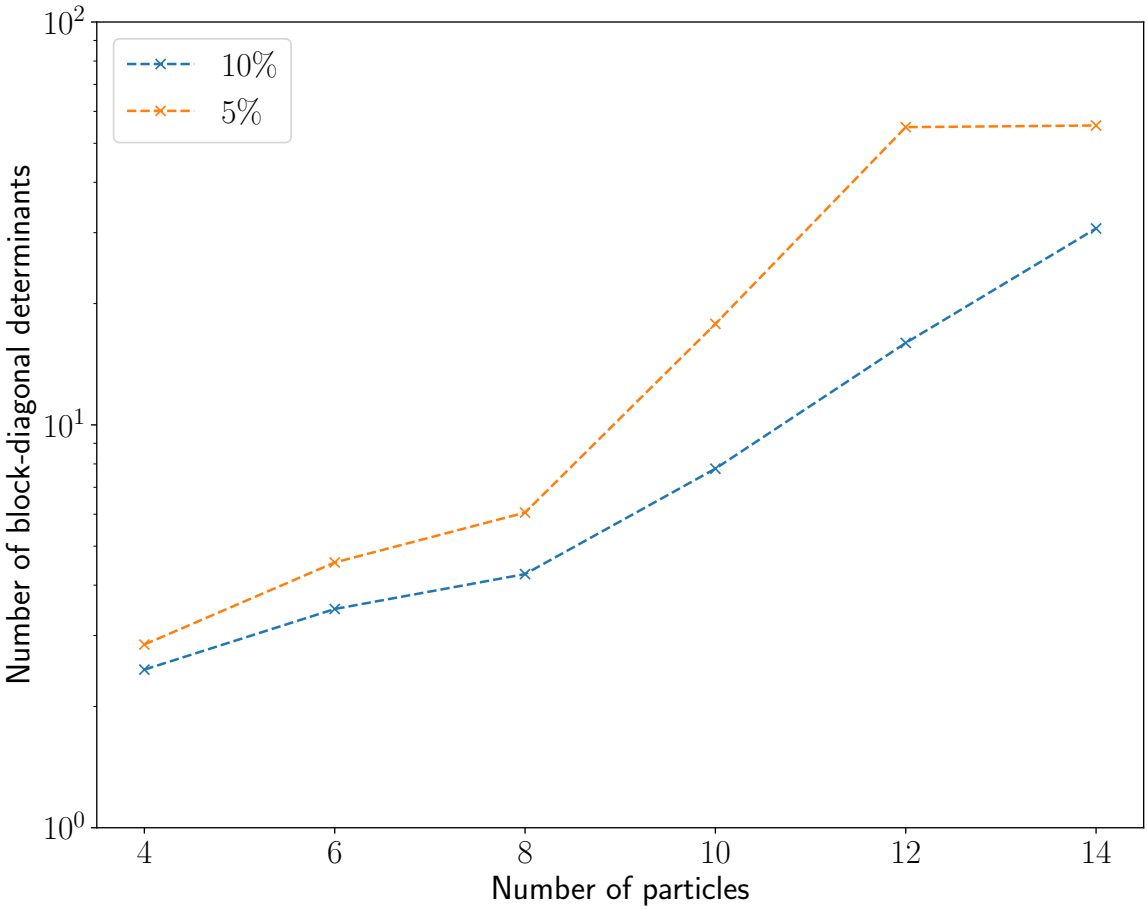


Figure 5.6: Number of block-diagonal determinants required for the multi-determinant Slater FermiNet energy to fall within 5% and 10% of the energy obtained using a single-determinant/single-geminal AGPs FermiNet.

14. Hence, in practice, using the AGPs FermiNet to studying paired systems is more feasible.

5.2 Discussion

In this chapter, we used the FermiNet NWF [121] to study the benchmark superfluid system known as the UFG. We showed that the Slater FermiNet NWF has difficulties in describing paired systems with strong, short-ranged attractive interactions between particles of opposite spin. Hence, we proposed a way to improve the variational Ansatz by using determinants of FermiNet geminals, similar to an AGPs or a BCS wave function. We showed mathematically that the Slater FermiNet is a limiting case of the AGPs FermiNet despite the use of fewer parameters in the latter. It follows that any FermiNet wave function can in principle be written as an AGPs FermiNet wave function.

We compared the total energies and energies per particle of the UFG as calculated using the Slater FermiNet and the AGPs FermiNet. The former fails to produce a paired state when the number of particles, N, is greater than around 10, while the AGPs FermiNet works very well.

As the UFG has a superfluid ground state, we computed the pairing gap and condensate fraction for the $N = 38$ system and compared estimates made with the Slater FermiNet and the AGPs FermiNet. There is a clear qualitative difference between the pairing gap obtained using the AGPs FermiNet and the Slater FermiNet, with the latter approaching zero as the number of particles N increases. Calculations of the superfluid condensate fraction show a similar behaviour: the AGPs FermiNet gives an accurate finite result, while the value obtained using the Slater FermiNet tends to zero in the limit of large system size. Although the AGPs pairing gap shows significant finite size errors, it lies close to the mean-field BCS result with Gorkov-Melik-Barkhudarov corrections [63]. Taken together, these results show that the Slater FermiNet is unable to represent large systems with superfluid ground states. The AGPs FermiNet is much more suitable for studying paired systems such as the UFG.

To demonstrate the success of the AGPs FermiNet, we also compared our calculated total energies with state-of-the-art fixed-node diffusion QMC energies obtained using a Jastrow-BCS Ansatz [47]. For all systems with more than a few particles, the AGPs FermiNet achieves lower (*i.e.*, better) variational energies than FN-DMC using the same model interaction and system parameters.

The inability of the Slater FermiNet Ansatz to accurately describe the UFG ground state comes as a surprise because the original FermiNet paper [121] argued that any many-body fermionic wave function can be represented as a single determinant of FermiNet orbitals. However, the mathematical argument relies on the construction of FermiNet orbitals with unphysical discontinuities. Whether or not any wave function can be represented as a single determinant of FermiNet orbitals of the type used in practice, which are differentiable everywhere except at electron-electron and electron-nuclear coalescence points, remains an open question. Another limitation is that the architecture of the FermiNet neural network, which is rather simple, may not be able to represent an arbitrary many-electron FermiNet orbital. Even if a single-determinant Slater FermiNet wave function is general in principle, there is no guarantee that it is equally easy to represent all wave functions. It may be that producing an accurate representation of a paired wave function requires the width and number of layers in the neural network to increase rapidly with system size. Furthermore, if a network of fixed size is used, it may be necessary to increase the number of Slater FermiNet determinants rapidly as the system size increases. The observation that the Slater FermiNet works well when $N \lessapprox 10$ but that the quality of the results degrades rapidly for larger systems, along with the scaling study presented in the final part of the Results section, suggests that this is, in fact, the case.

The AGPs FermiNet introduced in this paper shares many of the strengths of the Slater FermiNet. In particular, there is no need to construct and optimise a new basis set for every new system or particle type. If the AGPs FermiNet proves equally successful in other paired

systems, it may now be relatively easy to investigate the importance of pairing in molecules, electron-positron systems, electron-hole liquids, and other s-wave superfluids. Another strength of the AGPs FermiNet is the ease with which it is possible to optimise linear combinations of determinants of FermiNet pairing orbitals, such as the one in Eq. (4.11). This is much more difficult to accomplish with conventional wave functions based on explicit two-electron pairing orbitals or pairing orbitals represented as outer products of single-particle orbitals or basis functions. Just as the many-particle orbitals in a Slater FermiNet radically generalise single-particle orbitals by incorporating electron-electron terms in a permutation-equivariant fashion, so the pairing functions in an AGPs FermiNet generalise BCS-style pairing functions by incorporating the effects of the remaining electrons in a permutation-equivariant fashion.

The AGPs FermiNet introduced here has a straightforward Pfaffian extension and can thus be applied to non-s-wave and triplet pairing. Therefore, we expect it to become a powerful tool for understanding strongly correlated non-s-wave superfluid and superconducting systems such as Helium-3 or high-T_c and p-wave superconductors. Finally, our approach is not limited to the FermiNet neural network and can be readily adapted to use more recent architectures such as the Psiformer [162], GLOBE and MOON [54], and DeepErwin [138].

Chapter 6

Spin-Imbalanced Fermi Gas

As mentioned in the background chapter, the ground state of a two-component, spin-balanced and isotropic Fermi gas with zero-range attractive interactions at $T = 0$ is *always* an s-wave superfluid, the qualitative properties of which have been known for decades [128, 142]. Gradually increasing the interaction strength only gives a smooth crossover—the BCS-BEC crossover. In reality, there exist many more exotic phases in dilute Fermi gases, some of which were introduced in Chapter 2. To observe and model these more exotic phases, one needs to break certain symmetries. The most straightforward approach is to adjust the population of each spin species to get a spin-imbalanced Fermi gas, which mimics a superconductor in an external magnetic field. As such, great amount of effort has been put in this field to determine the different phases of matter that can exist in such a system.

Although most of these phases were predicted decades ago, they have proven notoriously difficult to observe in experiments. For example, the FFLO phase was predicted to exist in 3D ultracold spin-imbalanced Fermi gases. However, it only exists in a narrow range of interaction strength and polarisation [125] because the critical field in the Clogston-Chandrasekhar limit, $h_c^{\mathrm{CC}} = \frac{\Delta_0}{\sqrt{2}} \approx 0.707\Delta_0$, is very close to the critical field predicted by the FFLO theory, $h_c^{\mathrm{FF}} = 0.754\Delta_0$ [154], where Δ_0 is the BCS gap at zero field and $T = 0$. Hence, it is hard to observe the FFLO phase even in the highly tunable ultracold Fermi gases in 3D.

In contrast, 2D Fermi systems provide an excellent host for exotic phases such as the FFLO phase, as discussed in Chapter 2. This chapter describes our work using neural wave functions (NWF), specifically the FermiNet [121, 149], to study these phases in the spin-imbalanced Fermi gas.

6.1 Results

The Hamiltonian of a 2D SIFG is

$$\hat{H} = -\frac{1}{2}\sum_{i}^{N^{\uparrow}} \nabla^2_{\mathbf{r}_i^{\uparrow}} - \frac{1}{2}\sum_{i}^{N^{\downarrow}} \nabla^2_{\mathbf{r}_i^{\downarrow}} + \sum_{i}^{N^{\uparrow}}\sum_{j}^{N^{\downarrow}} U(\mathbf{r}_i^{\uparrow} - \mathbf{r}_j^{\downarrow}), \tag{6.1}$$

where

$$U(\mathbf{r}) = -\frac{2v_0\mu^2}{\cosh^2(\mu r)} \tag{6.2}$$

is the modified Pöschl-Teller potential that is also used in the study of the unitary Fermi gas. This potential is tuned by using the two parameters μ and v_0 to achieve the desired strength and range of the interaction. The strength of interaction at low energies is characterised by the s-wave scattering length a_s, which is always positive in 2D due to the fact that one can *always* find a bound-state in the two-body scattering problem, no matter how weak the interaction is. As such, we choose to characterise the interaction strength by the natural logarithm of the scattering length,

$$\eta = \ln(k_F a_s), \tag{6.3}$$

where k_F is the Fermi wavevector of a *spin-balanced* Fermi gas of the same total electron density, given by $k_F = \sqrt{2\pi n}$ in 2D. Under this scheme, the weakly interacting limit (BCS limit) is when $a_s \to 0^+$ and $\eta \to -\infty$, whereas the strongly interacting limit (BEC limit) is given by $a_s \to +\infty$ and $\eta \to +\infty$. Note that while analytical expression for a_s in terms of μ and v_0 exists in 3D [45], no such expression is available in 2D. As such, to reach the desired values of η, we solve the 2D scattering problem numerically and obtain the corresponding η for each choices of μ and v_0.

As in our previous work on the 3D unitary Fermi gas, we study the spin-imbalanced Fermi gas in a 2D periodic square box with constant density by setting $k_F = 1$. This corresponds to $r_s = \sqrt{2}$, where r_s, the radius of a circle that contains one particle on average, provides a convenient measure of the inter-particle distance. Throughout this work, we employ the dimensionless system based on Hartree atomic units: the unit of length is the Bohr radius, a_0, and the unit of energy is the Hartree. To ensure that the range of the interaction is small compared with the inter-particle separation, we set $\mu = 24$, while varying v_0 to drive the system to various scattering lengths. The chosen values of v_0 and the corresponding values of η are shown in Table (6.1).

We studied a system with $(N^{\uparrow}, N^{\downarrow}) = (25, 9)$. The Fermi surfaces (occupied momentum states) of this system without any interactions are shown in Fig. (6.1). With the definition in Eq. (2.47), this configuration gives a polarisation of $P_{(25,9)} = 0.47$. Using the formulae from Eq. (2.46), Eq. (2.53) and Eq. (2.68), we determine the approximate locations of the phase transitions in terms of the scattering length a_s. For the $(25, 9)$ Fermi gas, the Clogston-Chandrasekhar limit and the FF limit correspond to $\eta^{\mathrm{CC}}_{(25,9)} = 1.22$ and $\eta^{\mathrm{FF}}_{(25,9)} = 1.56$ respectively. Note that these values are by no means accurate predictions for the locations of the phase transitions,

v_0	η	$\lvert\epsilon_b\rvert/\epsilon_F$
0.10	4.367	4.061×10^{-4}
0.15	1.956	5.047×10^{-2}
0.20	0.747	5.661×10^{-1}
0.25	0.018	2.431
0.30	-0.470	6.457
0.35	-0.821	13.04
0.40	-1.087	22.18
0.45	-1.298	33.82
0.50	-1.465	47.19

Table 6.1: Values of v_0 and their corresponding values of $\eta = \ln(k_F a_s)$, as well as the binding energy per particle $|\epsilon_b|/\epsilon_F = 8e^{-2(\gamma+\eta)}$ at a fixed $\mu = 24$.

because the formulae are derived within mean-field theory and are only valid in the weakly interacting BCS limit. In addition, the derivations assume the thermodynamic limit, whereas we are studying the spin-imbalanced Fermi gas with a finite number of particles in a periodic simulation cell. Nonetheless, they provide reasonable rough estimates of the locations of the relevant phase transitions.

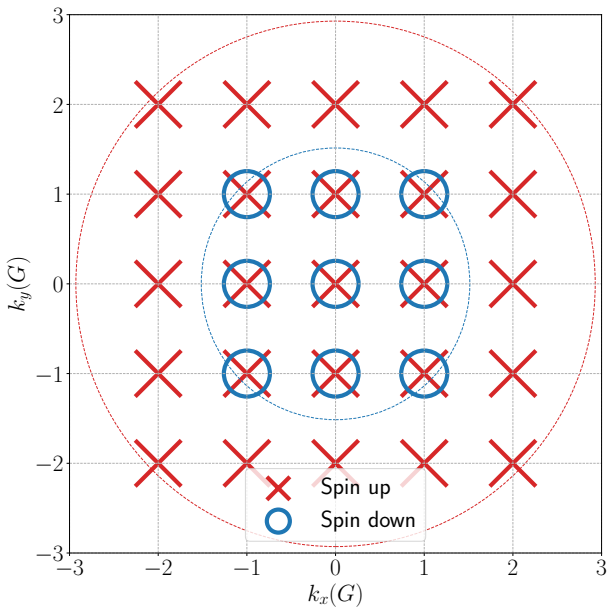
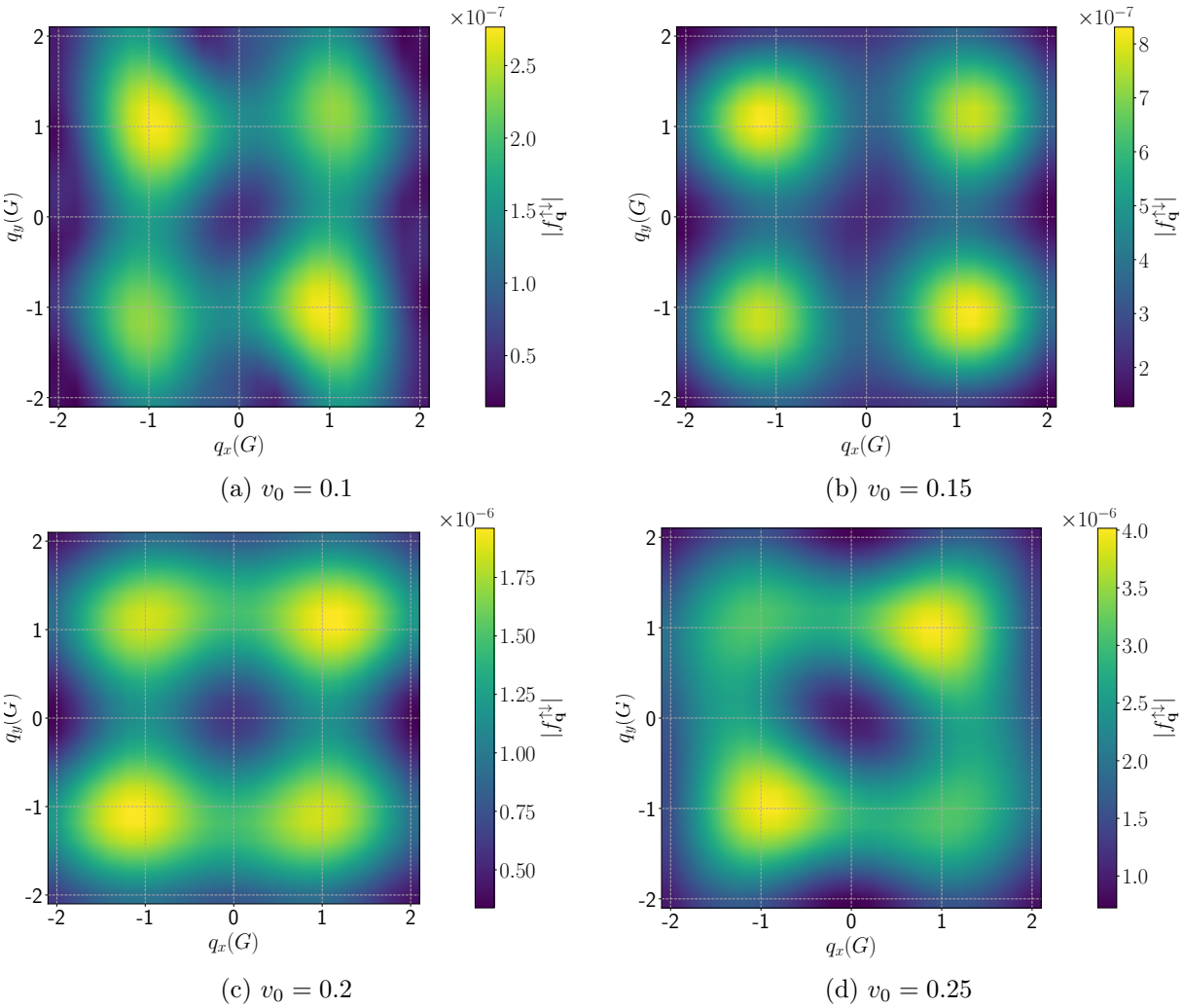


Figure 6.1: The occupied momentum states in the non-interacting (25, 9) Fermi gas. The red crosses and the blue circles denote the occupied states by the spin-up and spin-down particles, respectively. The wave vectors k_x and k_y are measured in units of $G = 2\pi/L$, where L is the length of the side of the square real-space simulation cell.

6.1.1 Pair-Momentum Density

As described in Chapter 2 (see Eq. (2.86)), the pair-momentum density is the Fourier transform of the translationally averaged two-body density matrix $\bar{\rho}^{(2)}_{\alpha\beta}(\mathbf{r})$. For an s-wave superfluid, the only nonzero pair-momentum density is $f_0^{\uparrow\downarrow}$ at $\mathbf{k} = 0$, *i.e.*, a delta peak. This indicates a condensation of zero-momentum Cooper pairs. In contrast, the phases with finite-momentum Cooper pairing, such as the FFLO phase, will have nonzero values of $f_{\mathbf{q}}^{\uparrow\downarrow}$ at the corresponding finite values of $\mathbf{q}$. Hence, the pair-momentum density can provide evidence for the existence of FFLO phases in the system.

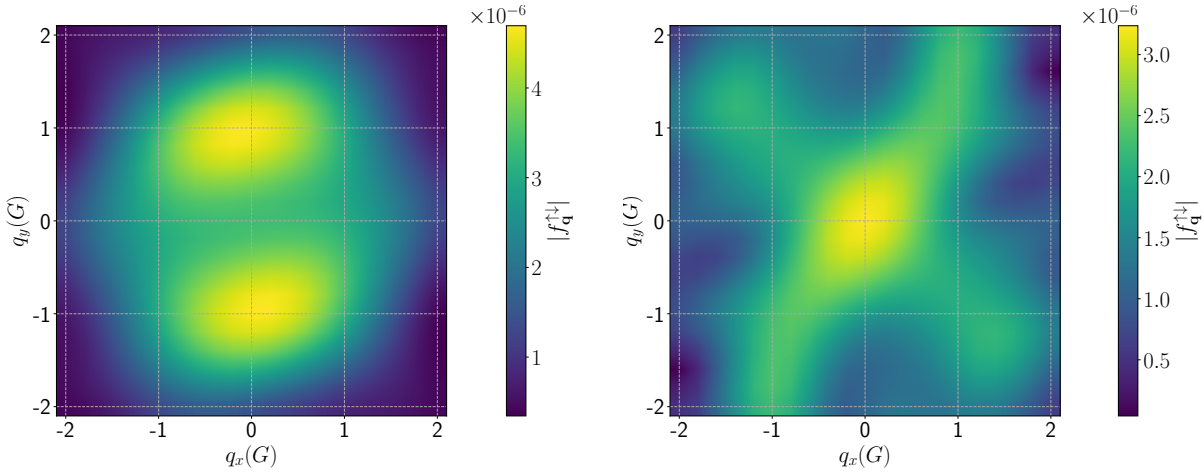


(a) $v_0 = 0.1$ (b) $v_0 = 0.15$

(c) $v_0 = 0.2$ (d) $v_0 = 0.25$

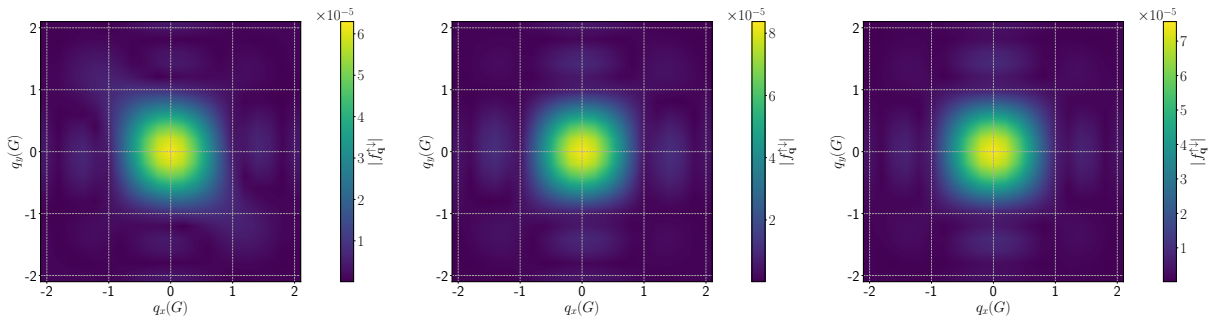
Figure 6.2: Pair-momentum density with signatures of the FFLO phase.

The pair-momentum density of the $(25, 9)$ Fermi gas in momentum space for different strengths of interaction can be seen in Fig. (6.2), Fig. (6.3) and Fig. (6.4). At weak interaction strengths, $v_0 = 0.1, 0.15, 0.2, 0.25$, the pair-momentum densities shown in Fig. (6.2) exhibit clear peaks at the four combinations of wavevectors $\mathbf{q} = (\pm G, \pm G)$, which indicates the existence of finite-



(a) $v_0 = 0.3$ (b) $v_0 = 0.35$

Figure 6.3: Pair-momentum density at the intermediate strength of interaction.



(a) $v_0 = 0.4$ (b) $v_0 = 0.45$ (c) $v_0 = 0.5$

Figure 6.4: Pair-momentum density with signatures of a standard BEC s-wave superfluid.

momentum Cooper pairs forming at the Fermi surface, connected by the four-fold rotational symmetry of the square simulation box. The magnitude of the signal increases with the strength of interaction, ranging from roughly 10^{-7} to 10^{-6} as v_0 increases from 0.1 to 0.25. All of the peaks are at least 6 sample standard deviations above zero, providing strong evidence for the presence of the FFLO phase. On the other hand, at strong interactions, such as $v_0 = 0.4, 0.45, 0.5$, Fig. (6.4) shows that the pair-momentum density is dominated by the $\mathbf{q} = 0$ peak, with negligible signal elsewhere. This indicates that the system has a condensate of zero-momentum Cooper pairs, making it a polarised s-wave superfluid.

In the meantime, in the intermediate regime with $v_0 = 0.3$ and $v_0 = 0.35$, the pair-momentum densities shown in Fig. (6.3) do not allow an unambiguous identification of the phase to be made. To investigate further, we need to take a look at other observables.

6.1.2 One-Particle Density

It is a well known that the ground state of a translationally invariant Hamiltonian is also translationally invariant. Hence, one would expect the one-particle number density $n(\mathbf{r})$ of the ground state of the spin-imbalanced Fermi gas to be uniform. This is, in fact, the case at weak interactions, *e.g.*, $v_0 = 0.1, 0.15, 0.2, 0.25$. However, if we compute the one-particle densities separately for different spin channels at the intermediate interaction strengths, $v_0 = 0.3, 0.35$, we observe a breaking of the translation symmetry. This can be seen in Fig. (6.5), where the minority spins (spin-down) are seen to be localised. The majority spin (spin-up) number density has a mostly uniform background with an increase in density at the locations of the localised minority spins. All of these observations suggest that the system is a crystal of pairs of fermions, submerged in a sea of uniform unpaired spin-up fermions. Slices of the 2D density $n^\sigma(\mathbf{r})$ with either a fixed $x = x_0$ or $y = y_0$ can be seen in Fig. (6.6), where the top and bottom rows show the majority and minority spin density, respectively. The first column shows a horizontal slice of the 2D density with a fixed $y = y_0 \approx -0.06L$, indicated by the *red dash-dotted horizontal* line on the figure in the third column. The second column shows a vertical slice of the 2D density with a fixed $x = x_0 \approx 0.17L$, indicated by the *red dashed vertical* line on the figure in the third column. In the 1D projection plots, the black dashed horizontal line represents the average number density, $\bar{n}^\sigma = \frac{N^\sigma}{\Omega}$, of fermions of spin σ in the simulation cell with volume Ω. The blue curves are the actual values of the density across the cell, with the orange curves being running average of the blue curves.

This is a very surprising set of results, as we are seeing crystal formation from a system of fermions with *purely attractive interactions*. To the best of our knowledge, this crystalline phase is previously unseen in the spin-imbalanced Fermi gas in 2D, perhaps because it only exists in a small range of values of η roughly between $\eta \sim 0$ and $\eta \sim -1$ in our simulations. The formation of a crystal could imply that the there exists an effective repulsion between quite strongly bound Cooper pairs of fermions that prevents their collapse. Indeed, if we increase the strength of the interaction further and push the system into the strongly interacting BEC limit, the dimers collapse and the system becomes phase separated in real space, as shown in Fig. (6.7). On the other hand, in the weakly interacting regime with $\eta > 0$, the attraction is not sufficiently strong to support the formation of bosonic dimers of fermions, and the translation symmetry is not broken. It is only in the intermediate region that the balance between the attraction and the effective repulsion leads to the formation of the crystalline phase.

6.2 Discussion

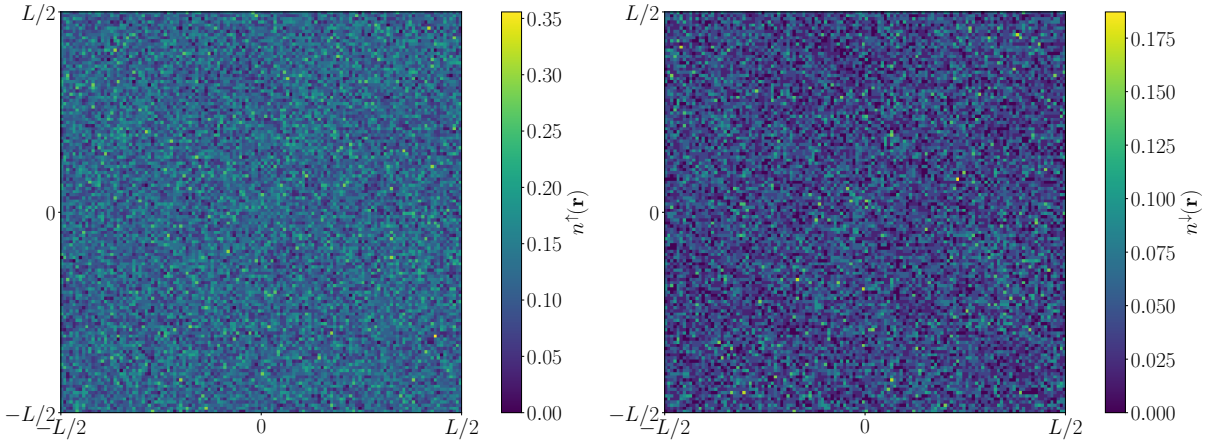
In this work, we have presented a study of the 2D two-component spin-imbalanced Fermi gas with short-range attractive interactions between opposite-spin fermions. We fixed the system size to be $N^\uparrow = 25$ and $N^\downarrow = 9$, varied the strength of interaction to drive the system to different

scattering lengths, and investigated the ground-state phase diagram. In the weakly interacting BCS limit, when $\eta = \ln(k_F a_s) \gtrsim 0.018$, we observed an FFLO phase with four-fold rotational symmetry. In the strongly interacting BEC limit, when $\eta = \ln(k_F a_s) \lesssim -4.7$, the system was a standard s-wave superfluid but became phase separated in real space, with the fully-paired minority-spin fermions clustered into islands.

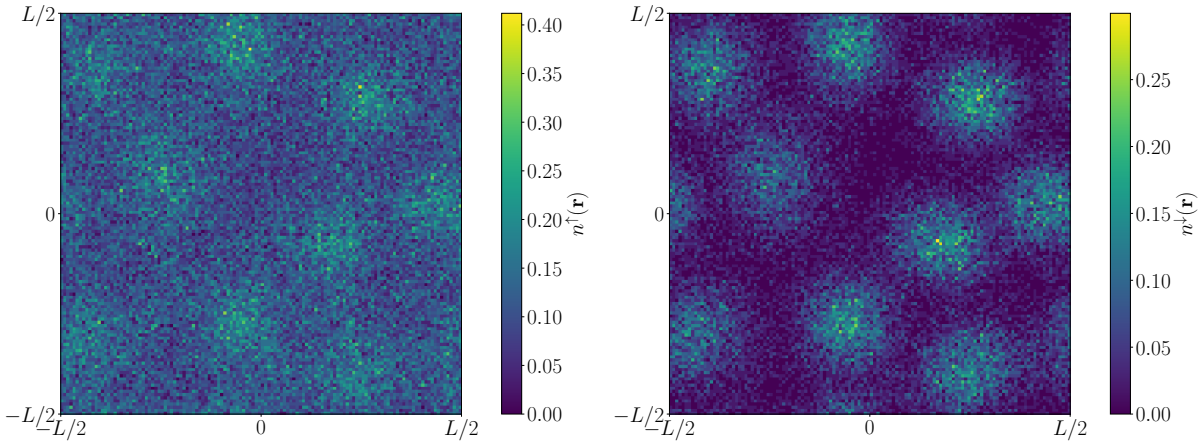
Between these two limits was surprising, the one-particle spin densities in real space were seen to exhibit translational symmetry breaking. Specifically, the strongly bounded Cooper pairs are localised and formed a crystal, while the unpaired majority spins form a uniform background. To the best of our knowledge, this $T = 0$ phase of the spin-imbalanced Fermi gas has not previously been observed or predicted. The formation of a crystal in a 2D Fermi gas is surprising as it implies the emergence of effective repulsions between the bosonic dimers in a *purely attractive* Fermi gas. While a complete theory is elusive, we have proposed a possible explanation of our results. Previous work based on diagrammatic perturbation theory has shown that, in a strongly interacting attractive Fermi gas, there exists an effective *repulsion* between the tightly bounded dimers of fermions, mediated by the unpaired fermions [11, 71, 120]. On either side of this intermediate regime, the attractive interaction is either too weak or too strong for a formation of crystals. Hence, we believe the emergence of the crystalline phase requires fine balance between the bare attraction and the effective repulsion.

As in the case of all variational methods, it is possible that the optimisation process may fail to converge sufficiently or that the neural network may be unable to represent the true ground state. The states we have found may be physically incorrect. However, we must also stress that the neural network variational Monte Carlo method in general and the AGPs FermiNet Ansatz in particular have shown themselves able to achieve accurate and unbiased results in superfluids [95], without requiring the use of any external data. Furthermore, in contrast to projector quantum Monte Carlo methods such as diffusion and auxiliary-field Monte Carlo, variational calculations such as ours are unaffected by the fermion sign problem. In addition, we reran the same simulations using the original FermiNet with the same network size and same number of dense determinants as the geminals [121, 149]. Both Ansätze gave similar results, especially on the crystal phase at $v_0 = 0.3$. Hence, we have good reasons to believe the reliability of our results.

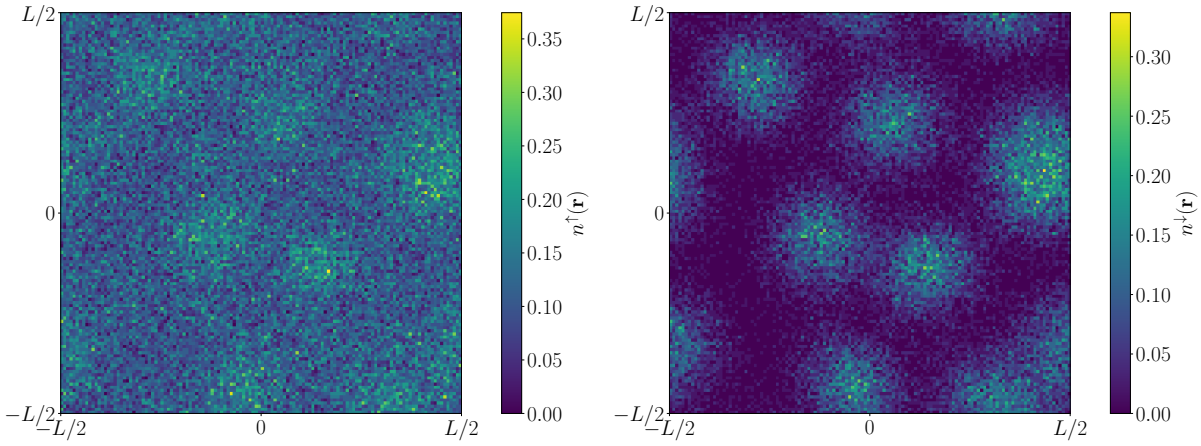
Future work could involve comparisons with results obtained using other neural wave functions and other numerical methods. The amount of work using numerical methods to study 2D spin-imbalanced attractive Fermi gas is very limited and we hope that the new directions suggested by our results will inspire further study using theoretical, computational and experimental methods.



(a) $v_0 = 0.25$



(b) $v_0 = 0.3$



(c) $v_0 = 0.35$

Figure 6.5: One-particle density of the $(25, 9)$ spin-imbalanced Fermi gas for a range of different interaction strengths near the weakly interacting BCS limit, plotted separately for different spin channels. The plots on the left-hand side are the densities of spin-up particles and the plots on the right-hand side are the densities of the spin-down particles.

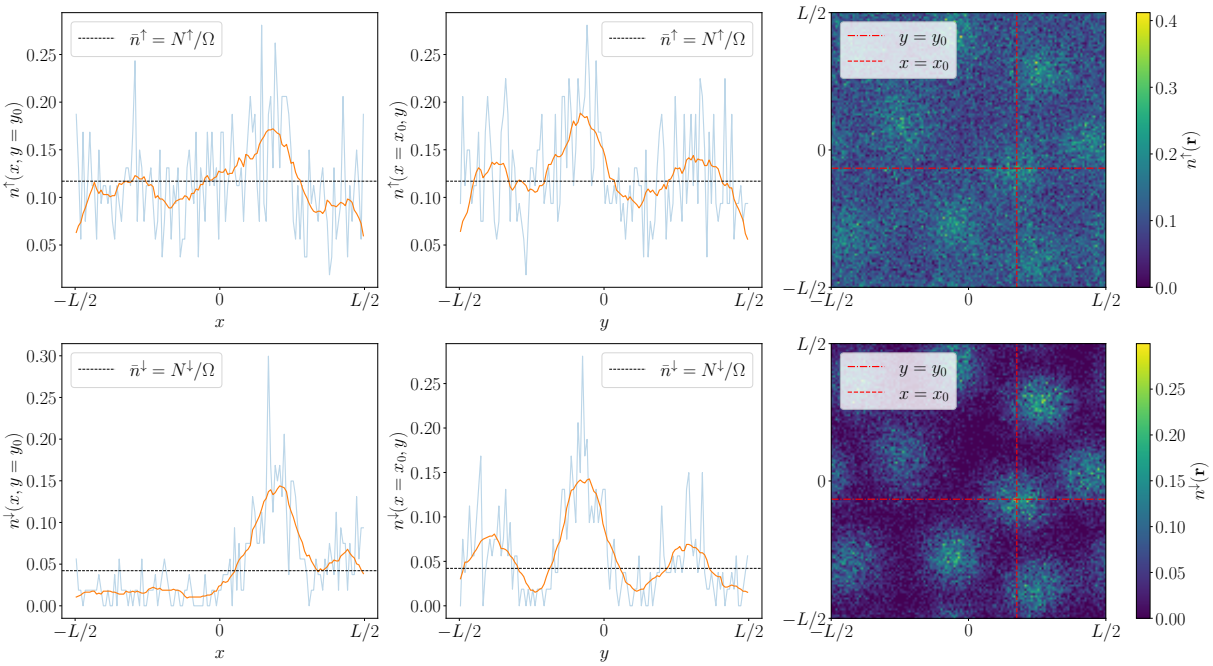
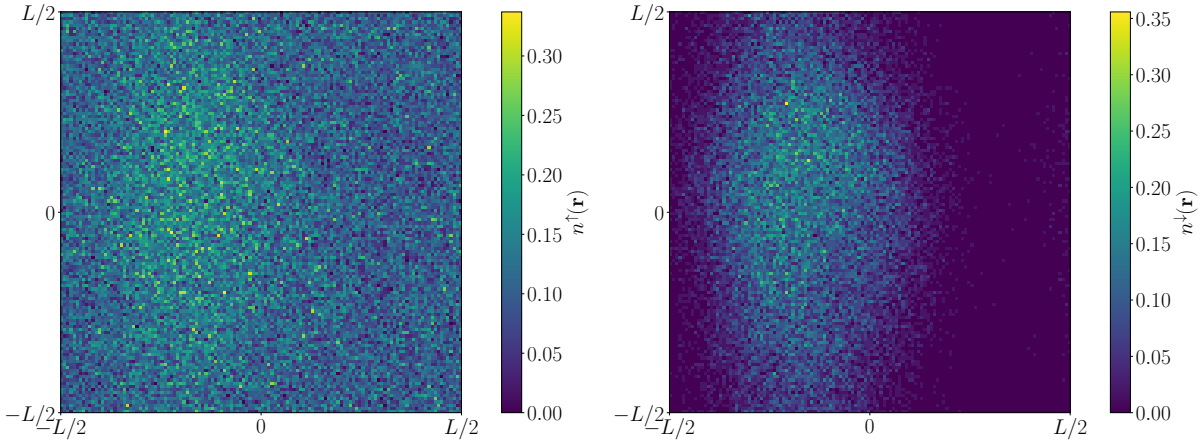
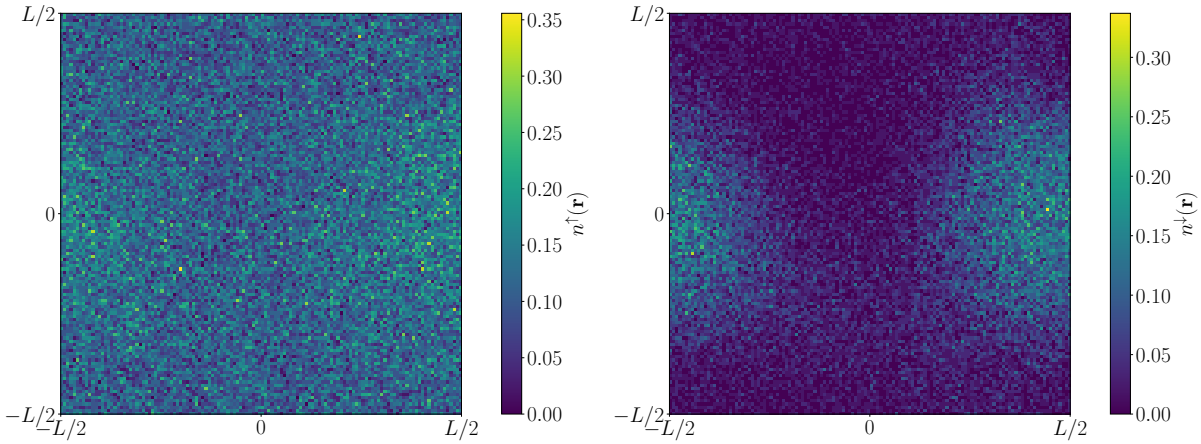


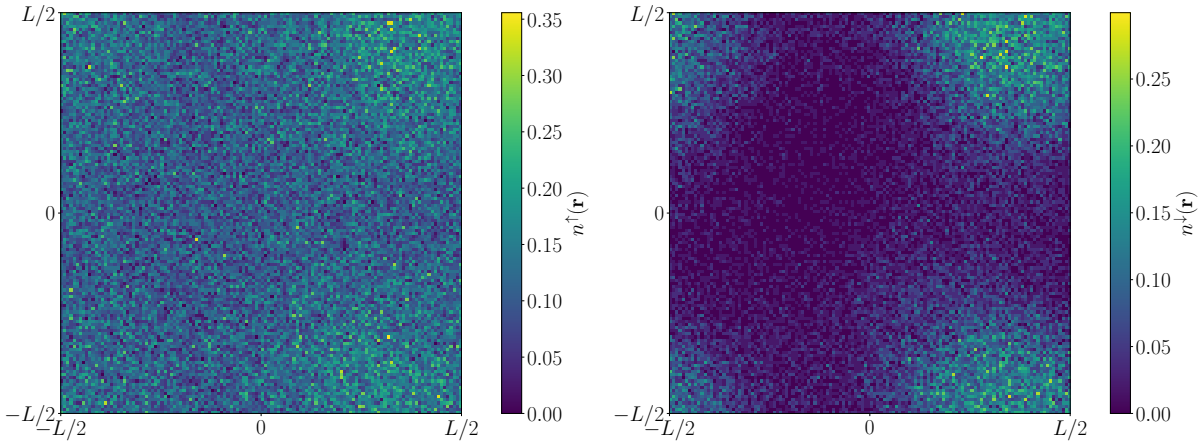
Figure 6.6: One-particle densities and their 1D projections for the $(25, 9)$ spin-imbalanced Fermi gas at $v_0 = 0.3$, plotted separately for different spin channels. The first row shows the majority spin (spin-up) density and the second row represent the minority spin (spin-down) density. The first column shows a horizontal slice of the 2D density with a fixed $y = y_0 \approx -0.06L$, indicated by the *red dash-dotted horizontal* line on the figure in the third column. The second column shows a vertical slice of the 2D density with a fixed $x = x_0 \approx 0.17L$, indicated by the *red dashed vertical* line on the figure in the third column. In the 1D projections, the black dashed horizontal line represents the average number density, $\bar{n}^{\sigma} = \frac{N^{\sigma}}{\Omega}$, of fermions of spin σ in the simulation cell with volume Ω. The blue curves are the actual values of the density across the cell, with the orange curves being running average of the blue curves.



(a) $v_0 = 0.4$



(b) $v_0 = 0.45$



(c) $v_0 = 0.5$

Figure 6.7: One-particle density of the $(25, 9)$ spin-imbalanced Fermi gas for a range of interaction strengths in the intermediate to strongly interacting regime, plotted separately for different spin channels. The plots on the left-hand side are the densities of spin-up particles and the plots on the right-hand side are the densities of the spin-down particles.

Chapter 7

Anisotropic Fermi Gas

In this chapter, we continue our study of 2D Fermi gases. This time, we study the 2D anisotropic Fermi gas with attractive interactions. The anisotropies arise from the spin- and spatial-direction-dependent masses of the particles, with the dispersion given in Eq. (2.74). While this particular system has been studied in the past using mean-field and numerical methods [41, 42, 66], it was only realised recently that this kind of d-wave anisotropy is one of the prime signatures of the newly discovered collinear magnetic phase – altermagnetism [144].

7.1 Results

The Hamiltonian of the anisotropic Fermi gas is

$$\hat{H} = \hat{H}_0 + \hat{U}, \tag{7.1}$$

$$\hat{H}_0 = -\frac{1}{2}\sum_i^{N^\uparrow}\left[\frac{1}{1+\delta}\frac{\partial^2}{\partial x_i^{\uparrow 2}} + \frac{1}{1-\delta}\frac{\partial^2}{\partial y_i^{\uparrow 2}}\right] - \frac{1}{2}\sum_i^{N^\downarrow}\left[\frac{1}{1-\delta}\frac{\partial^2}{\partial x_i^{\downarrow 2}} + \frac{1}{1+\delta}\frac{\partial^2}{\partial y_i^{\downarrow 2}}\right], \tag{7.2}$$

$$\hat{U} = \sum_i^{N^\uparrow}\sum_j^{N^\downarrow} U(\mathbf{r}_i^\uparrow - \mathbf{r}_j^\downarrow), \tag{7.3}$$

where

$$U(\mathbf{r}) = -\frac{2v_0\mu^2}{\cosh^2(\mu r)} \tag{7.4}$$

is the interaction between opposite spin particles. This is exactly the same interaction we used to study the unitary Fermi gas in Chapter (5) and the spin-imbalanced Fermi gas in Chapter (6). We study the anisotropic Fermi gas system with $(N^\uparrow, N^\downarrow) = (11, 11)$ and $\delta \in \{0.1, 0.3, 0.5, 0.7\}$, where δ, which is defined by Eq. (7.2), controls the anisotropy. The non-interacting Fermi surfaces of the two spin species are shown in Fig. (7.1). Note the values of δ considered in this chapter were chosen such that the set of occupied states of the non-interacting system is always as shown in Fig. (7.1). The one-electron energy levels change as δ is varied, but their $T = 0$

occupations do not.. As in the spin-imbalanced Fermi gas, we set the system density parameter r_s to $\sqrt{2}a_0$ and the interaction (inverse) length scale μ to $24a_0^{-1}$, where a_0 is the Bohr radius. The strength of the interaction is controlled by the parameter v_0. All values of v_0 in the set $\{0.05, 0.1, 0.2, 0.3, 0.4, 0.5\}$Hartrees$a_0^2$ were considered.. The corresponding values of η can be found in Tab. (6.1), except for the weakest interaction strength $v_0 = 0.05$, for which $\eta = 11.582$.

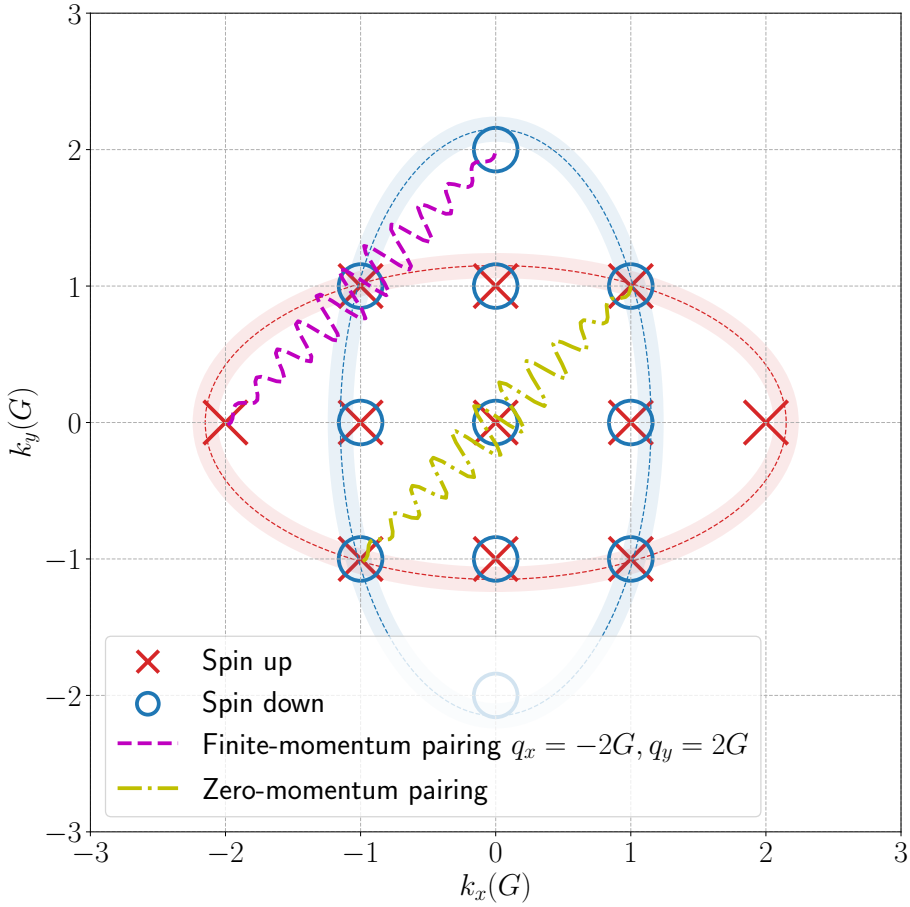


Figure 7.1: The occupied momentum states in the weakly-interacting (11, 11) anisotropic Fermi gas. The red crosses and the blue circles denote the occupied states by the spin-up and spin-down particles, respectively. The coloured regions at the two Fermi surfaces denote the region bounded by the Debye frequency. The dashed wavy line in magenta and the dash-dotted wavy line in olive indicates possible finite-momentum and zero-momentum pairing, respectively. This arrangement of states is the same for any of our chosen anisotropy $\delta \in \{0.1, 0.3, 0.5, 0.7\}$.

7.1.1 Condensate Fraction

The finite-momentum pairing phases we are interested in are characterised by several observables, the most important of which is the condensate fraction $c_{\uparrow\downarrow}$ as a function of v_0 and δ. The condensate fraction, defined in Eq. (2.84), represents the fraction of pairs that condense into the $\mathbf{k} = 0$ zero-momentum state in a superfluid. If the condensate fraction is greater than

zero, the system, or at least part of the system [1], is in the standard BCS superfluid state. In our unitary Fermi gas work in Chapter 5, we calculated this quantity with Eq. (2.84), but one can also calculate it using the pair-momentum density, $f_{\mathbf{q}}^{\uparrow\downarrow}$ at $\mathbf{q} = 0$, as shown in Eq. (2.88). Here, we use the latter method. Results are shown in Fig. (7.2), where the x-axis represents the strength of interaction v_0, and the y-axis represents the anisotropy δ. The number in each square represent the condensate fraction at a given v_0 and δ, which also matches with the colour scale of the pixel. While it is difficult to determine whether the system is in a superfluid state on the weakly interacting side of Fig. (7.2), as the numbers are very close to zero, we choose our criteria as follows: if the condensate fraction lies within one standard error of zero, we regard the system as not being in an s-wave superfluid state. Based on this criterion, the condensate fractions are essentially zero in the upper left-hand corner of the matrix of values in Fig. (7.2), where the anisotropy is strong and the interactions weak, implying that these systems are not s-wave superfluids. On the other hand, in the bottom right-hand corner, where the interactions are strong and the anisotropy is weak, the condensate fractions are of order unity, indicating the existence of a condensation of zero-momentum Cooper-pairs. The normal-to-superfluid phase boundary lies roughly on the diagonal from $v_0 = 0.1, \delta = 0.1$ to $v_0 = 0.4, \delta = 0.7$. To examine the states near the phase boundary more carefully, we need to look at other quantities, such as the two-body density matrix and the pair-momentum density at $\mathbf{k} \neq 0$.

7.1.2 Two-Body Density Matrix and Pair-Momentum Density

Next, we look at the real space, translationally averaged two-body density matrix, $\bar{\rho}_{\uparrow\downarrow}^{(2)}(\mathbf{r})$, and its Fourier components, also known as the pair-momentum density, $f_{\mathbf{q}}^{\uparrow\downarrow}$, defined in Eq. (2.81) and Eq. (2.86), respectively. As stated in Chapter 2, exotic paired phases, such as the FFLO phase [50, 87], are characterised by the existence of peaks in $f_{\mathbf{q}}^{\uparrow\downarrow}$ at non-zero values of $\mathbf{q}$. Notice that, to simplify comparison with the condensate fractions shown in Fig. (7.2), we scale the pair-momentum density by a factor of $\frac{\Omega^2}{N_{\text{pair}}}$.

When the interaction is weak, $v_0 = 0.05, 0.1, 0.2$, the zero-momentum condensate fraction is nearly zero for all anisotropies δ except $\delta = 0.1$ at $v_0 = 0.1$ and $\delta = 0.1, 0.3$ at $v_0 = 0.2$. At the weakest interaction considered, $v_0 = 0.05$, the condensate fraction is never significantly above zero, indicating that there is no zero-momentum Cooper-pair condensation in this system. In addition, the $\mathbf{q} \neq 0$ peaks in the pair-momentum density, which are shown in Fig. (C.3) of Appendix C.1, are all lower than the central peak, suggesting the system is not in any superfluid phase.

At $v_0 = 0.1$, the pair-momentum densities suggest a similar lack of superfluidity when $\delta = 0.3, 0.5, 0.7$. However, at the lowest anisotropy, $\delta = 0.1$, there is a small, the lowest anisotropy $\delta = 0.1$ has a small but non-zero condensate fraction, as well as four non-zero peaks at $(q_x, q_y) = (\pm 2G, \pm 2G)$, where G is the magnitude of the primitive reciprocal vector of the

[1] The system could be phase separated. We will discuss this in later sections.

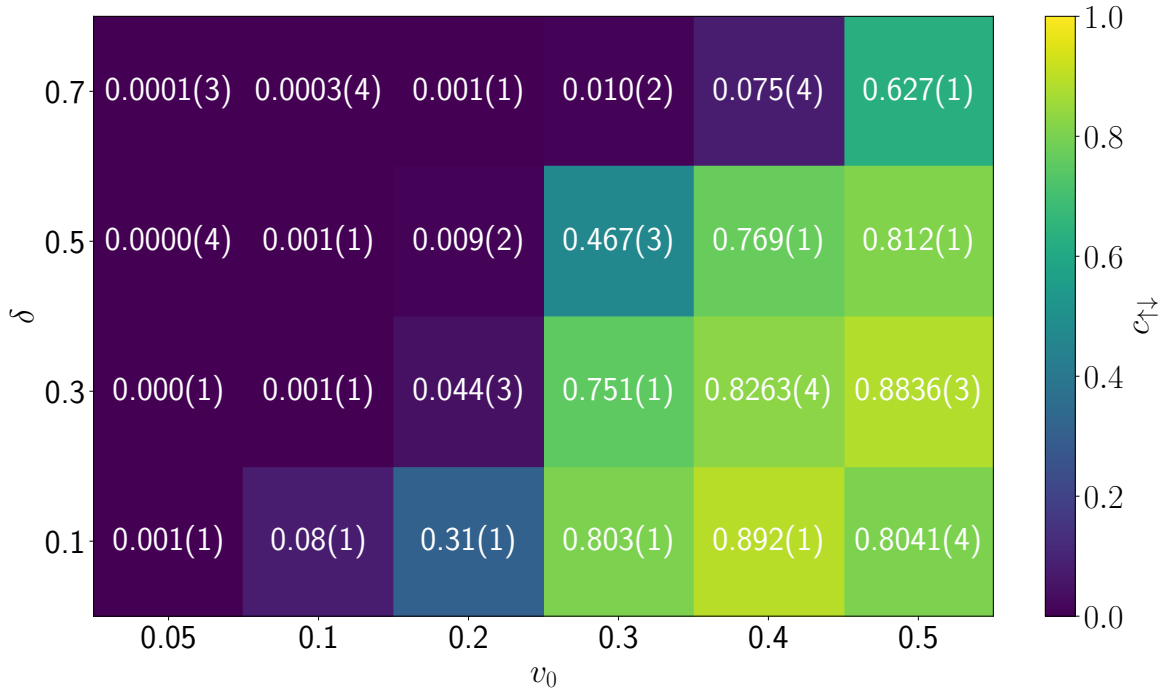


Figure 7.2: The superfluid condensate fraction $c_{\uparrow\downarrow}$ for a range of values of v_0 and δ. The white number in each square is the value of the condensate fraction at the corresponding v_0 and δ. The single digits in parentheses indicate the standard error in the final decimal place of the corresponding value of the condensate fraction.

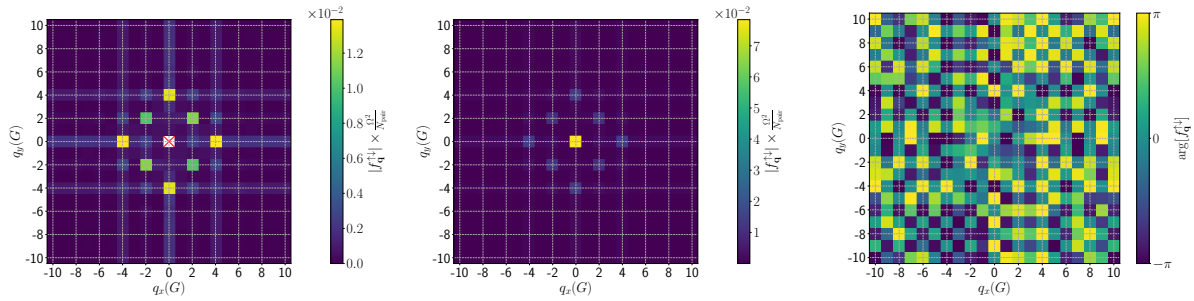
lattice of simulation cells. The pairing wave vectors $(q_x, q_y) = (\pm 2G, \pm 2G)$ are illustrated in Fig. (7.3a). These four peaks are consistent with the existence of pairing between the tips of the two Fermi surfaces in Fig. (7.1), indicated by the dashed wavy line in magenta, which is also consistent with the mean-field calculations in Ref. [143]. This is a strong indication of the existence of a pair-density wave (PDW) in this system [2]. Indeed, as predicted by FFLO theory, the real space translationally averaged two-body density matrix $\bar{\rho}^{(2)}_{\uparrow\downarrow}(\mathbf{r})$ in Fig (7.4a) shows a very clear spatial modulation across the simulation cell. Moreover, Fig. (7.3a) also shows four other peaks at $q_x = \pm 4G, q_y = 0$ and $q_x = 0, q_y = \pm 4G$ that are not seen in previous mean-field calculations.

At the next strength of interaction $v_0 = 0.2$ in Fig. (7.2), the $\mathbf{q} = 0$ condensate fraction for $\delta = 0.1, 0.3, 0.5$ is larger than when $v_0 = 0.1$. The condensate fractions for $\delta = 0.1, 0.3$ are now significantly above zero, while the condensate fraction for $\delta = 0.5$ is only slightly (roughly four standard errors) above zero. In Fig. (7.4b), the two-body density matrix displays a clear spatial modulation across the simulation cell, leading to multiple $\mathbf{q} \neq 0$ peaks in the pair-momentum density in Fig. (7.3b). However, all of these peaks are overshadowed by the $\mathbf{q} = 0$ peak, which

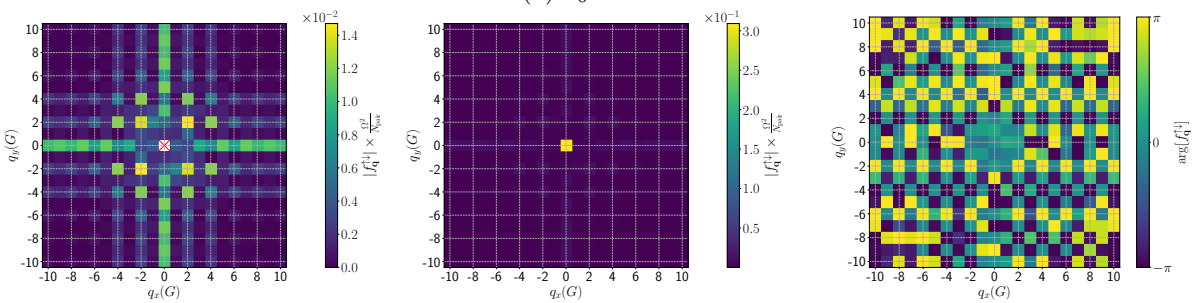
[2] As mentioned in Chapter 2, the term "pair-density wave" is more suitable here than the term "FFLO phase" because the former describes general superconducting states with spatially modulated order parameters, whereas the latter rely on spin imbalance.

is roughly one order of magnitude larger than the peaks with $\mathbf{q} \neq 0$.

These results suggest the systems at $v_0 = 0.1$ and $v_0 = 0.2$ with $\delta = 0.1$ are close to a phase transition between the BCS and PDW phases. Indeed, as we increase the interaction strength further to $v_0 = 0.3$, we get a BCS superfluid up to $\delta = 0.5$, as Fig. (7.2) suggests. At $v_0 = 0.3, \delta = 0.1, 0.3$, $v_0 = 0.4, \delta = 0.1, 0.3, 0.5$ and $v_0 = 0.5, \delta = 0.1, 0.3, 0.5, 0.7$, the condensate fraction $c_{\uparrow\downarrow}$ of the BCS superfluid is around 0.8. The pair-momentum density for $v_0 = 0.3, 0.4, 0.5$ at $\delta = 0.1$ can be found in Fig. (7.5a), Fig. (7.5b) and Fig. (7.5c). Results for all δ can be found in Fig. (C.6), Fig. (C.8) and Fig. (C.9) in Appendix C.1. Note that the left-hand pair-momentum density plots in Fig. (7.3b), Fig. (7.5a), Fig. (7.5b) and Fig. (7.5c), all of which have had the dominant $\mathbf{q} = 0$ component removed, contain large "crosses" centred on the origin. These are artefacts caused by Fourier transforming an isotropic function inside a square simulation cell. A much fainter version of this artefact can also be seen in Fig. (7.3a).



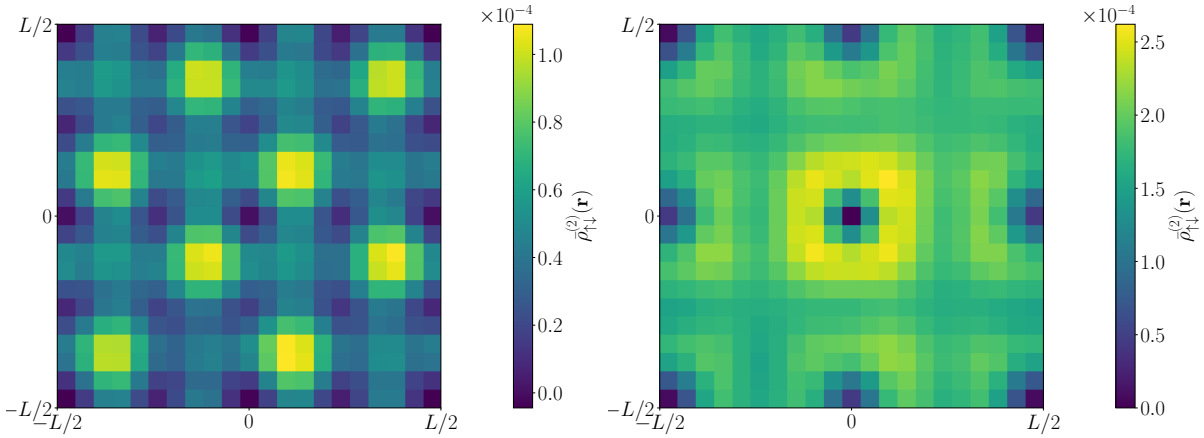
(a) $v_0 = 0.1$



(b) $v_0 = 0.2$

Figure 7.3: The pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.1, 0.2$ and $\delta = 0.1$. The left-hand figures are the absolute values of the pair-momentum density with the dominant peak at $\mathbf{k} = 0$ masked out, such that other weaker signals can be seen. The middle figures are the original unmasked absolute values of the pair-momentum density. The right-hand figures are the phases of the pair-momentum density.

Lastly, we briefly discuss the possibility of the existence of a Cooper-pair Bose metal (CPBM) phase in our anisotropic Fermi gas model. The CPBM phase has both Cooper pairs and gapless excitations, without any condensation in pair-momentum space. It was first proposed as a possible explanation of the long-standing pseudo-gap problem in high-T_c superconductors [123], and has been predicted to exist in 2D systems with anisotropic spin-dependent Fermi surfaces



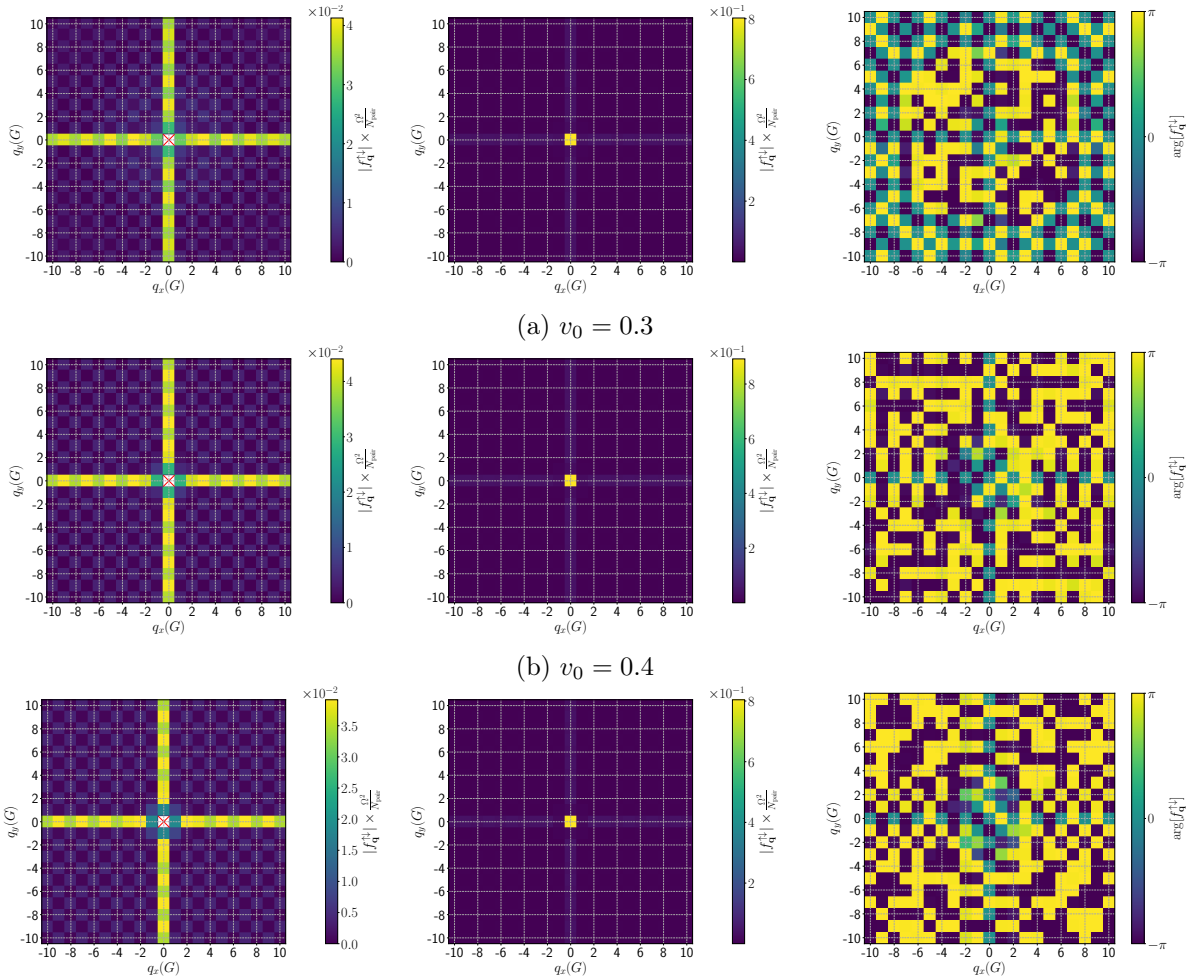
(a) $v_0 = 0.1$, $\delta = 0.1$ (b) $v_0 = 0.2$, $\delta = 0.1$

Figure 7.4: The real space, translationally averaged two-body density matrix, $\bar{\rho}^{(2)}_{\uparrow\downarrow}(\mathbf{r})$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.1, 0.2$ and $\delta = 0.1$.

[17, 18, 41, 42]. Recent simulations using the constrained-path quantum Monte Carlo method and the functional renormalisation group method [17, 18] have observed CPBM phases in lattice models with anisotropic hopping terms. In particular, the CPBM state was favoured in the strong anisotropy regime, where s-wave pairing is suppressed. These results prompted us to look at systems with large δ and zero or very small condensate fractions.

In Fig. (7.6) and Fig. (7.7), we compare the absolute values of the pair-momentum density, multiplied by the factor $\frac{\Omega^2}{N_{\text{pair}}}$, for $v_0 = 0.2, 0.3, 0.35, 0.4$ and $\delta = 0.5, 0.7$. Note that these figures are the "zoomed-in" version of the previous pair-momentum density plots, since the pair-momentum densities are zero when $|\mathbf{q}|$ is large. Fig. (7.6) shows results for $\delta = 0.5$ and Fig. (7.7) shows results for $\delta = 0.7$. At $v_0 = 0.2$ when $\delta = 0.5$, and at all values of v_0 when $\delta = 0.7$, we observe a very weak $\mathbf{k} = 0$ peak and a diffuse region that spreads diagonally to the four points $(q_x, q_y) = (\pm G, \pm G)$ at $v_0 = 0.2$. These $\mathbf{q}$ points differ from the ones that gain intensity from FFLO-type pairing, as can be seen by comparing Fig. (7.6) and Fig. (7.7) with Fig. (7.3a). Furthermore, the peaks at $(q_x, q_y) = (\pm G, \pm G)$ are even weaker than the nearly-zero $\mathbf{q} = 0$ peak. Hence, we rule out the possibility of the existence of a PDW in this regime. As the strength of interaction is increased further, the system falls back to a pure s-wave BCS superfluid, with the Fourier transform artefacts surrounding the $\mathbf{k} = 0$ bright peak.

Although these non-trivial results are inconsistent with both the normal and BCS superfluid phases, they are not sufficient to establish the presence of a CPBM. As emphasised in Ref. [18], the existence of a CPBM would require measurements of additional observables, such as the bosonic density structure factor and density–density correlations. Due to time constraints, we were unable to compute these quantities, and therefore cannot confirm or rule out the existence of the CPBM phase. We leave this question for future investigation.



(a) $v_0 = 0.3$

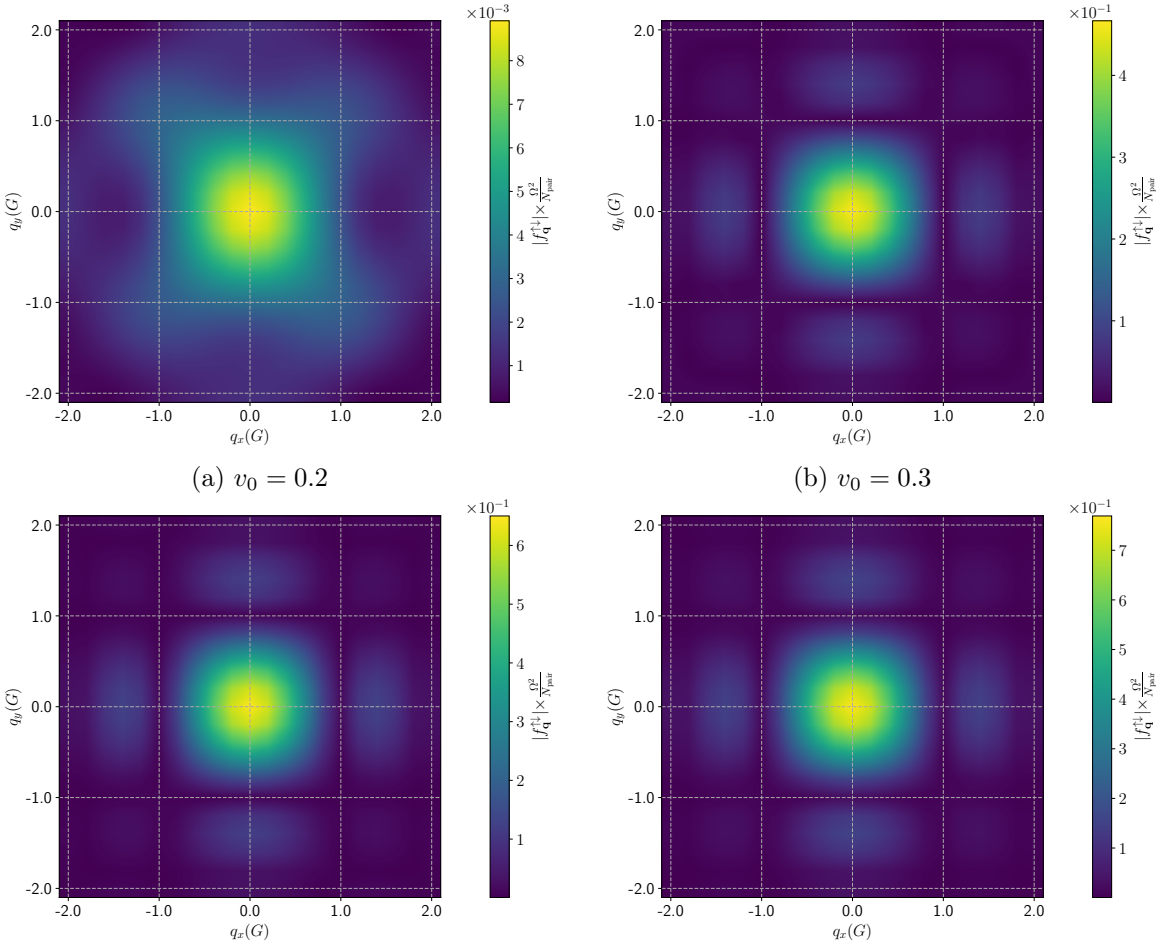
(b) $v_0 = 0.4$

(c) $v_0 = 0.5$

Figure 7.5: The pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.4, 0.5$ and $\delta = 0.1$. The left-hand figures are the absolute values of the pair-momentum density with the dominant peak at $\mathbf{k} = 0$ masked out, such that other weaker signals can be seen. The middle figures are the original unmasked absolute values of the pair-momentum density. The right-hand figures are the phases of the pair-momentum density.

7.2 Discussion

In this work, we used the neural network variational Monte Carlo (NNVMC) method to study the interaction-anisotropy phase diagram of the 2D anisotropic Fermi gas with a d-wave-like spin-dependent splitting of its dispersion relation in momentum space. We focused our attention on the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ spin-balanced Fermi gas with short-range attractive interactions, tuning the anisotropy and strength of interaction across the BCS-BEC crossover to obtain a rough 2D phase diagram of this system. Fig. (7.2) shows that the BCS superfluid condensate fraction decreases as the anisotropy increases at fixed interaction strength or the interaction strength decreases at fixed anisotropy. In particular, at small anisotropies and weak interactions, we see



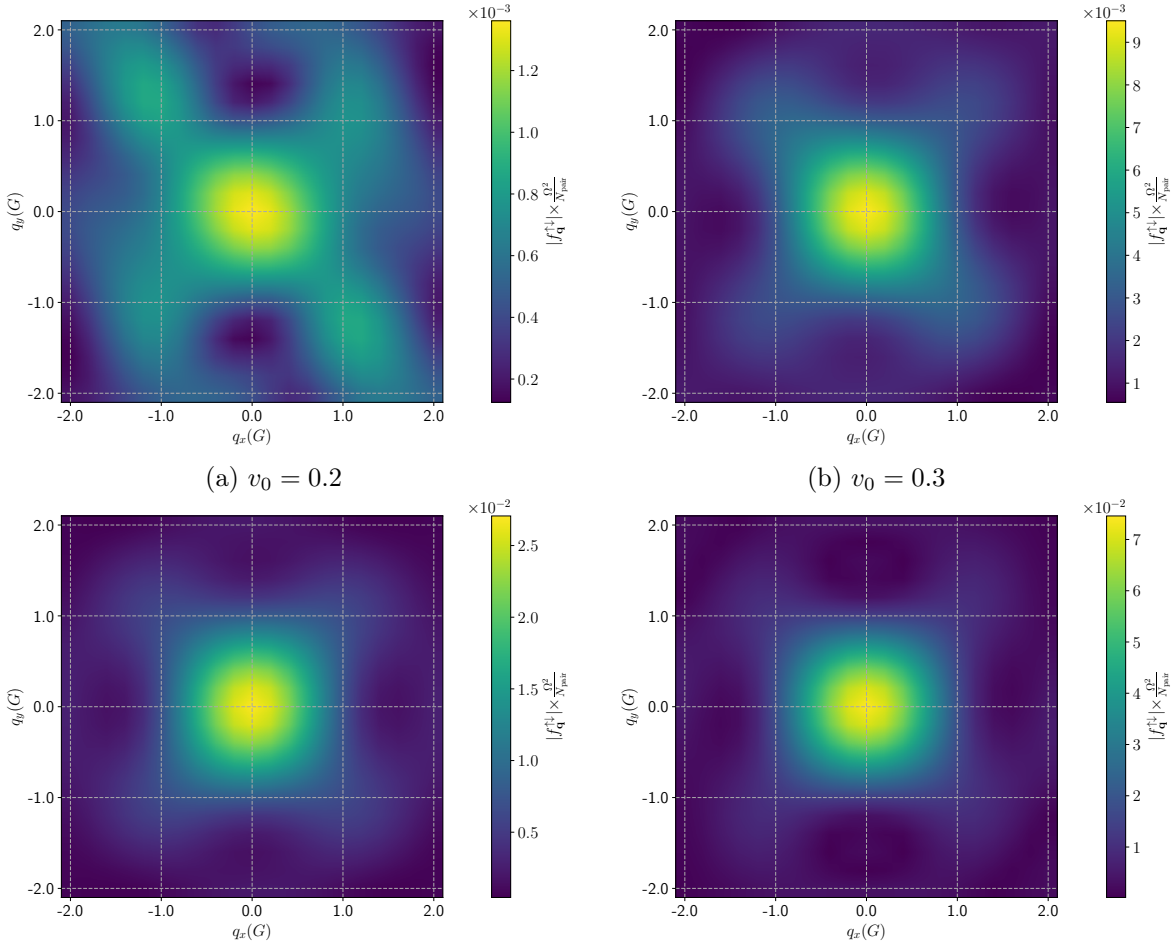
(a) $v_0 = 0.2$ (b) $v_0 = 0.3$

(c) $v_0 = 0.35$ (d) $v_0 = 0.4$

Figure 7.6: The absolute values of the pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.2, 0.3, 0.35, 0.4$ and fixed $\delta = 0.5$. Note that we have cropped out the regions outside $\mathbf{q} = (\pm 2G, \pm 2G)$, as the signals in those regions are essentially zero. The whole uncropped version of the pair-momentum density can be found in the figures in Appendix C.1.

the co-existence of BCS superfluidity with zero-momentum pairing and a PDW phase with finite-momentum pairing. The presence of BCS superfluidity is demonstrated by the non-zero value of the $\mathbf{q} = 0$ condensate fraction, whilst the non-zero peaks in the pair-momentum density and the spatial modulation of the translationally averaged two-body density matrix demonstrate the presence of a PDW. The co-existence disappears as the interactions become stronger, until the system becomes a pure s-wave BCS superfluid, which survives up to intermediate ($\delta \sim 0.3 - 0.5$) and even large anisotropies ($\delta > 0.5$).

On the other hand, in the large anisotropy and intermediate to strong interaction regime, various mean-field calculations and quantum Monte Carlo simulations of anisotropic Hubbard models [17, 18, 41, 42] have predicted the existence of a CPBM phase. While we have seen



(a) $v_0 = 0.2$ (b) $v_0 = 0.3$

(c) $v_0 = 0.35$ (d) $v_0 = 0.4$

Figure 7.7: The absolute values of the pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.2, 0.3, 0.35, 0.4$ and fixed $\delta = 0.7$. Note that we have cropped out the regions outside $\mathbf{q} = (\pm 2G, \pm 2G)$, as the signals in those regions are essentially zero. The whole uncropped version of the pair-momentum density can be found in the figures in Appendix C.1.

non-trivial peaks in the pair-momentum density in this regime, our results are not sufficient to support or rule out the existence of the CPBM phase. As such, we leave this question to future work.

To the best of our knowledge, this chapter describes the first application of the NNVMC method to the anisotropic Fermi gas with d-wave-like spin-dependent splitting in momentum space. The study of d-wave anisotropic Fermi gases has recently gained considerable attention, largely motivated by the discovery of a novel type of collinear magnetism, termed altermagnetism [144, 145]. This new class of materials highlights the rich interplay between spin splitting and metallic properties, an interplay that creates fertile ground for exotic phases of matter at low temperatures, including pair-density wave (PDW) states [27, 143]. We hope that this work will

provide a useful starting point for exploring the microscopic mechanisms underpinning these phenomena, as well as a demonstration of the potential of neural-network-based quantum many-body simulations in addressing open problems in correlated fermionic systems. Looking ahead, a natural extension of this work is to investigate 2D Fermi gases with other underlying symmetries, which may host an even broader landscape of unconventional quantum phases.

Chapter 8

Conclusion

In this thesis, we have utilised the novel neural network variational Monte Carlo (NNVMC) and neural wave function (NWF) methods to study exotic phases of matter. In Chapter 1 and Chapter 2, we provided a brief introduction to the Bardeen-Cooper-Schrieffer (BCS) mean-field theory [8] of dilute Fermi gases with attractive interactions. Specifically, we mentioned how one can make use of the "diluteness" of the system to replace the model-dependent interaction with a universal scale – the s-wave scattering length – by regularising the interaction. This regularisation allows one to tune the system through the Bardeen-Cooper-Schrieffer to Bose-Einstein condensate (BCS-BEC) crossover regime by adjusting only a single quantity. We then extended our discussion to Fermi gases with certain broken symmetries, namely the spin-imbalanced Fermi gas and the anisotropic Fermi gas. In relation to these two systems, we explained how to extend the BCS theory from zero-momentum pairing to finite-momentum pairing. The latter is thought to be the main ingredient in several exotic phases of matter, such as the Fulde-Ferrell-Larkin-Ovchinnikov (FFLO) phase and the pair-density wave (PDW) state. Lastly, to bridge the theory discussed in Chapter 2 with the NNVMC method, we provided a discussion of the important observables that one can calculate with Monte Carlo methods.

In Chapter 3, we provided a brief introduction to the NNVMC method. Starting from the many-body time-independent Schrödinger equation (TISE), we introduced the variational Monte Carlo (VMC) method and explained how the problem of solving the TISE can be turned into an optimisation problem. After that, we briefly discussed how to construct an optimisable trial wave function that satisfies fermionic exchange symmetry using Slater determinants, thereby obtaining the well-known Slater-Jastrow wave function. We then introduced the idea of a backflow transformation, which transforms the real particle coordinates into a set of quasiparticle coordinates, further improving the accuracy of the wave function. The backflow transformation imparts an important property to the orbital functions – *permutation invariance* – which makes the wave function *permutation equivariant*. The former term means that the outputs of the function are invariant under any permutation of its inputs, while the latter one means that if the inputs of the function are permuted, its outputs are also permuted in the same way. This is one of the

important concepts that led to the invention of neural network wave functions (NNWF), also known as neural wave functions (NWF), for continuous fermionic system. The most successful early examples of NNWF were the PauliNet [73] and the FermiNet [121, 149]. Following a short introduction to artificial neural networks (ANNs), we briefly discussed earlier work that used ANNs to represent the wave functions of lattice and spin models, an approach pioneered by Troyer and Carleo in 2017. Neural wave functions for lattice and spin systems are often called neural network quantum states (NNQS) or neural quantum states (NQS) [19].

In Chapter 4, we discussed a different way to construct a NWF, generalising the idea of the well-known antisymmetric geminal power (AGP) wave functions [6, 7, 25, 26, 57, 100]. Originating from the superconducting BCS wave function, the AGP wave function replaces the single-particle orbital functions used in SJB-type Ansätze by pairing orbital functions, or geminals, which are functions of two particle coordinates instead of one. This class of wave function is known to be better at capturing correlations between pairs of particles, especially in symmetry-broken phases such as superconductors and superfluids. Following a brief introduction to the AGP wave functions, we discussed how one can construct such wave functions using neural networks to represent generalised pairing orbitals that depend not only on the positions of two particles but also, in a permutation invariant manner, on the positions of the remaining $N-2$ electrons. An AGP neural wave function based on a network closely analogous to a FermiNet was presented as an example. In addition, we presented the relationships between the original FermiNet (also referred to as the Slater FermiNet in the main text) and the different AGP FermiNet wave functions.

In Chapter 5, we studied the unitary Fermi gas (UFG), a well-known example of a strongly correlated system with a superfluid ground state [128]. This system is a dilute Fermi gas with an infinite s-wave scattering length, and displays a form of universality, meaning that its properties do not depend on the details of the interaction potential. We utilised the FermiNet wave function to study the ground-state properties of the UFG and found that the original FermiNet, or the Slater FermiNet, is unable to accurately represent the ground states of large systems. By modifying the Slater FermiNet neural network to make it represent a generalised AGP singlet (AGPs) wave function, we were able to cure this problem completely. We found that the AGPs FermiNet reaches lower energies than the Slater FermiNet for all system sizes studied, and is able to capture the superfluid properties of the UFG, such as the non-zero pairing gap and the condensate fraction, producing results consistent with other simulations, theory, and experiments. The AGPs FermiNet results are more accurate than state-of-the-art fixed-node diffusion Monte Carlo (FN-DMC) results from the literature [47]; this is encouraging as FN-DMC is viewed as one of the most accurate methods available for studying interacting many-fermion systems in continuous space. Lastly, we studied the severity of the problems we had encountered using the Slater FermiNet, taking the AGPs FermiNet as a baseline. We found that the number of determinants required for the Slater FermiNet to reach a given accuracy relative to the AGPs FermiNet scales exponentially with the system size, indicating that Slater-determinant-based

antisymmetrisation method limits the performance of the Slater FermiNet in superfluid systems.

In Chapter 6, we extended the use of the AGPs FermiNet to study the spin-imbalanced Fermi gas (SIFG). The SIFG is a dilute Fermi gas system with unequal populations of the two spin species, and is difficult to describe with both theoretical and numerical methods due to its strongly interacting nature as well as the notorious sign problem. In particular, we focused our attention on the 2D SIFG, as it provides better Fermi surface nesting, creating more favourable conditions for the existence of exotic phases of matter, such as the FFLO phase. Having fixed the density and particle number, we varied the strength of interaction, sweeping through the BCS-BEC crossover. By analysing the relevant observables, we found evidence for the existence of the FFLO phase at weak interactions and the existence of a polarised superfluid phase at strong interactions, consistent with previous studies in the literature. In addition, we found a previously unknown phase of matter that breaks the translational symmetry of the system and forms a crystalline structure in the density profile, with several non-zero peaks in the pair-momentum density. In this regime, all minority spins paired up with the same number of majority spins, forming a localised crystal, while the remaining of the majority spins acts as a uniform background on top of the crystals. While we are unable to determine the exact nature of this phase, we hypothesise that it is a type of charge density wave (CDW) phase, or a Wigner-crystal-like phase arising because the Pauli principle induces an effective repulsion between the minority-spin fermions. To the best of our knowledge, this is the first time that such a phase has been reported in the 2D SIFG. While we do not rule out the possibility that it may be an artefact caused by finite-size or convergence issues, we stress that multiple independent calculations all show the same behaviour. We hope that this finding will motivate further studies in the future to confirm or refute our findings.

In Chapter 7, we studied another type of symmetry-broken Fermi gas with attractive interactions in 2D, which we refer to as the anisotropic Fermi gas. The anisotropic Fermi gas is a dilute Fermi gas with equal spin populations, but with an anisotropic effective mass tensor, such that the non-interacting Fermi surfaces are elliptical. The Fermi surfaces of the two spin species are related by a 90° rotation in momentum space. The study of this system was motivated by the recent discovery of a novel type of collinear magnetism, known as the altermagnetism, which has a similar Fermi surface structure. We studied the ground state properties of this system for a range of anisotropies and interaction strengths, using the AGPs FermiNet. At small anisotropies, we found that the system transited from the normal phase to a pair-density wave (PDW) state at weak interactions, and becomes an s-wave superfluid phase at strong interactions. Like the FFLO phase, the PDW state is characterised by finite-momentum Cooper pairing with spatial modulations in the pair-momentum density. The location of the various phase transitions gradually shifts towards stronger interactions as the anisotropy is increased. Lastly, we discussed the possibility of the existence of the Cooper-pair-Bose-metal (CPBM) phase at intermediate to strong interactions and large anisotropies, as suggested by previous theoretical and numerical studies. While we have seen possible signatures of the CPBM phase in our results, we are unable

to confirm its existence due to the limitations of our current calculations. As such, we leave this as an open question for future studies.

In conclusion, we have demonstrated the power of the NNVMC method in studying strongly correlated fermionic superfluids with exotic phases. Not only have we proven that the NNVMC method can provide accurate descriptions of the ground state properties of such systems and phases, but we have also shown that these methods are capable of discovering novel phases of matter in strongly correlated superfluid systems. We hope our work will motivate further studies using these novel methods, extending their applications to solve many other problems in condensed matter physics, quantum chemistry, and beyond.

Bibliography

[1] S. K. Adhikari, W. G. Gibson, and T. K. Lim. "Effective-range Theory in Two Dimensions". In: *The Journal of Chemical Physics* 85.10 (Nov. 1986), pp. 5580–5583. DOI: 10.1063/1.451572.

[2] Sadhan K. Adhikari. "Quantum Scattering in Two Dimensions". In: *American Journal of Physics* 54.4 (Apr. 1986), pp. 362–367. DOI: 10.1119/1.14623.

[3] Mark G. Alford et al. "Color Superconductivity in Dense Quark Matter". In: *Reviews of Modern Physics* 80.4 (Nov. 2008), pp. 1455–1515. DOI: 10.1103/RevModPhys.80.1455.

[4] G. E. Astrakharchik et al. "Momentum Distribution and Condensate Fraction of a Fermion Gas in the BCS-BEC Crossover". In: *Physical Review Letters* 95.23 (Dec. 2005), p. 230405. DOI: 10.1103/PhysRevLett.95.230405.

[5] J. E. Baarsma and P. Törmä. "Larkin-Ovchinnikov Phases in Two-Dimensional Square Lattices". In: *Journal of Modern Optics* 63.18 (Oct. 2016), pp. 1795–1804. DOI: 10.1080/09500340.2015.1128009. arXiv: 1511.01701 [cond-mat].

[6] M. Bajdich et al. "Pfaffian Pairing Wave Functions in Electronic-Structure Quantum Monte Carlo Simulations". In: *Physical Review Letters* 96.13 (Apr. 2006), p. 130201. DOI: 10.1103/PhysRevLett.96.130201.

[7] Michal Bajdich. *Generalized Pairing Wave Functions and Nodal Properties for Electronic Structure Quantum Monte Carlo*. Dec. 2007. arXiv: 0712.3066 [cond-mat].

[8] J. Bardeen, L. N. Cooper, and J. R. Schrieffer. "Microscopic Theory of Superconductivity". In: *Phys. Rev.* 106.1 (Apr. 1957), pp. 162–164. DOI: 10.1103/PhysRev.106.162.

[9] Federico Becca and Sandro Sorella. *Quantum Monte Carlo Approaches for Correlated Systems*. Cambridge: Cambridge University Press, 2017. DOI: 10.1017/9781316417041.

[10] Paulo F. Bedaque, Heron Caldas, and Gautam Rupak. "Phase Separation in Asymmetrical Fermion Superfluids". In: *Physical Review Letters* 91.24 (Dec. 2003). DOI: 10.1103/physrevlett.91.247002.

[11] G. Bertaina and S. Giorgini. "BCS-BEC Crossover in a Two-Dimensional Fermi Gas". In: *Physical Review Letters* 106.11 (Mar. 2011), p. 110403. DOI: 10.1103/PhysRevLett.106.110403.

[12] *Beryllium Dimer—Caught in the Act of Bonding*. DOI: 10.1126/science.1174326.

[13] I. Boettcher et al. "Sarma Phase in Relativistic and Non-Relativistic Systems". In: *Physics Letters B* 742 (Mar. 2015), pp. 86–93. DOI: 10.1016/j.physletb.2015.01.014. arXiv: 1409.5232 [cond-mat, physics:hep-ph, physics:nucl-th].

[14] J.P. Bouchaud, A. Georges, and C. Lhuillier. "Pair Wave Functions for Strongly Correlated Fermions and Their Determinantal Representation". In: *Journal de Physique* 49.4 (1988), pp. 553–559. DOI: 10.1051/jphys:01988004904055300.

[15] Bjørnulf Brekke, Arne Brataas, and Asle Sudbø. "Two-Dimensional Altermagnets: Superconductivity in a Minimal Microscopic Model". In: *Physical Review B* 108.22 (Dec. 2023), p. 224421. DOI: 10.1103/PhysRevB.108.224421.

[16] Heron Caldas. "Cold Asymmetrical Fermion Superfluids". In: *Physical Review A* 69.6 (June 2004). DOI: 10.1103/physreva.69.063602.

[17] Zhangkai Cao et al. "${d}_{{x}^{2}\ensuremath{-}{y}^{2}}$-Wave Bose Metal Induced by next-Nearest-Neighbor Hopping Anisotropy". In: *Physical Review B* 111.13 (Apr. 2025), p. 134512. DOI: 10.1103/PhysRevB.111.134512.

[18] Zhangkai Cao et al. *Exotic D-Wave Bose Metal in Two Dimensions*. May 2024. arXiv: 2405.13405 [cond-mat].

[19] Giuseppe Carleo and Matthias Troyer. "Solving the Quantum Many-Body Problem with Artificial Neural Networks". In: *Science* 355.6325 (2017), pp. 602–606. DOI: 10.1126/science.aag2302.

[20] J Carlson et al. "Auxiliary-Field Quantum Monte Carlo Method for Strongly Paired Fermions". In: *Physical Review A* 84.6 (2011), 061602(R).

[21] J. Carlson, S. Gandolfi, and Alexandros Gezerlis. "Superfluid Pairing in Neutrons and Cold Atoms". In: *Fifty Years of Nuclear BCS*. WORLD SCIENTIFIC, Mar. 2013, pp. 348–359. DOI: 10.1142/9789814412490_0026.

[22] J. Carlson et al. "Superfluid Fermi Gases with Large Scattering Length". In: *Phys. Rev. Lett.* 91.5 (July 2003), p. 050401. DOI: 10.1103/PhysRevLett.91.050401.

[23] G. Cassella et al. *Neural Network Variational Monte Carlo for Positronic Chemistry*. Feb. 2024. DOI: 10.48550/arXiv.2310.05607. arXiv: 2310.05607 [cond-mat, physics:physics].

[24] Gino Cassella et al. "Discovering Quantum Phase Transitions with Fermionic Neural Networks". In: *Physical Review Letters* 130.3 (Jan. 2023), p. 036401. DOI: 10.1103/PhysRevLett.130.036401.

[25] Michele Casula, Claudio Attaccalite, and Sandro Sorella. "Correlated Geminal Wave Function for Molecules: An Efficient Resonating Valence Bond Approach". In: *J. Chem. Phys* 121 (2004), p. 7110. DOI: 10.1063/1.1794632.

[26] Michele Casula and Sandro Sorella. "Geminal Wavefunctions with Jastrow Correlation: A First Application to Atoms". In: *The Journal of Chemical Physics* 119.13 (Oct. 2003), pp. 6500–6511. DOI: 10.1063/1.1604379. arXiv: cond-mat/0305169.

[27] Debmalya Chakraborty. "Zero-Field Finite-Momentum and Field-Induced Superconductivity in Altermagnets". In: *Physical Review B* 110.6 (2024). DOI: 10.1103/PhysRevB.110.L060508.

[28] Subhas J. Chakravorty et al. "Ground-State Correlation Energies for Atomic Ions with 3 to 18 Electrons". In: *Physical Review A* 47.5 (May 1993), pp. 3649–3670. DOI: 10.1103/PhysRevA.47.3649.

[29] B. S. Chandrasekhar. "A NOTE ON THE MAXIMUM CRITICAL FIELD OF HIGH-FIELD SUPERCONDUCTORS". In: *Applied Physics Letters* 1.1 (Sept. 1962), pp. 7–8. DOI: 10.1063/1.1777362.

[30] S. Y. Chang and V. R. Pandharipande. "Ground-State Properties of Fermi Gases in the Strongly Interacting Regime". In: *Phys. Rev. Lett.* 95.8 (Aug. 2005), p. 080402. DOI: 10.1103/PhysRevLett.95.080402.

[31] S. Y. Chang et al. "Quantum Monte Carlo Studies of Superfluid Fermi Gases". In: *Physical Review A* 70.4 (Oct. 2004), p. 043602. DOI: 10.1103/PhysRevA.70.043602.

[32] Qijin Chen et al. "BCS–BEC Crossover: From High Temperature Superconductors to Ultracold Superfluids". In: *Physics Reports* 412.1 (June 2005), pp. 1–88. DOI: 10.1016/j.physrep.2005.02.005.

[33] F. Chevy and C. Salomon. "Thermodynamics of Fermi Gases". In: *The BCS-BEC Crossover and the Unitary Fermi Gas*. Ed. by Wilhelm Zwerger. Berlin, Heidelberg: Springer, 2012, pp. 407–446. DOI: 10.1007/978-3-642-21978-8_11.

[34] Cheng Chin et al. "Feshbach Resonances in Ultracold Gases". In: *Reviews of Modern Physics* 82.2 (Apr. 2010), pp. 1225–1286. DOI: 10.1103/RevModPhys.82.1225.

[35] Kenny Choo, Antonio Mezzacapo, and Giuseppe Carleo. "Fermionic Neural-Network States for Ab-Initio Electronic Structure". In: *Nature Communications* 11.1 (Dec. 2020), p. 2368. DOI: 10.1038/s41467-020-15724-9.

[36] Kenny Choo, Titus Neupert, and Giuseppe Carleo. "Two-Dimensional Frustrated J1-J2 Model Studied with Neural Network Quantum States". In: *Physical Review B* 100.12 (Sept. 2019), p. 125124. DOI: 10.1103/PhysRevB.100.125124.

[37] A. M. Clogston. "Upper Limit for the Critical Field in Hard Superconductors". In: *Physical Review Letters* 9.6 (Sept. 1962), pp. 266–267. DOI: 10.1103/PhysRevLett.9.266.

[38] Purnendu Das et al. *Realizing Altermagnetism in Fermi-Hubbard Models with Ultracold Atoms*. Dec. 2023. DOI: 10.48550/arXiv.2312.10151. arXiv: 2312.10151 [cond-mat, physics:quant-ph].

[39] V. J. Emery. "Reaction Matrix Singularities and the Energy Gap in an Infinite System of Fermions". In: *Nuclear Physics* 19 (Sept. 1960), pp. 154–163. DOI: 10.1016/0029-5582(60)90228-5.

[40] Andres Perez Fadon et al. *Interaction-Induced Symmetry Breaking in Circular Quantum Dots*. Oct. 2024. DOI: 10.48550/arXiv.2410.01652. arXiv: 2410.01652 [cond-mat].

[41] Adrian E. Feiguin and Matthew P. A. Fisher. "Exotic Paired Phases in Ladders with Spin-Dependent Hopping". In: *Physical Review B* 83.11 (Mar. 2011), p. 115104. DOI: 10.1103/PhysRevB.83.115104.

[42] Adrian E. Feiguin and Matthew P. A. Fisher. "Exotic Paired States with Anisotropic Spin-Dependent Fermi Surfaces". In: *Physical Review Letters* 103.2 (July 2009), p. 025303. DOI: 10.1103/PhysRevLett.103.025303.

[43] Rafael M. Fernandes. "Topological Transition from Nodal to Nodeless Zeeman Splitting in Altermagnets". In: *Physical Review B* 109.2 (2024). DOI: 10.1103/PhysRevB.109.024404.

[44] R. P. Feynman and Michael Cohen. "Energy Spectrum of the Excitations in Liquid Helium". In: *Physical Review* 102.5 (June 1956), pp. 1189–1204. DOI: 10.1103/PhysRev.102.1189.

[45] Siegfried Flügge. *Practical Quantum Mechanics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1971. DOI: 10.1007/978-3-642-61995-3.

[46] Michael McNeil Forbes, Stefano Gandolfi, and Alexandros Gezerlis. "Effective-Range Dependence of Resonantly Interacting Fermions". In: *Physical Review A* 86.5 (Nov. 2012), p. 053603. DOI: 10.1103/PhysRevA.86.053603.

[47] Michael McNeil Forbes, Stefano Gandolfi, and Alexandros Gezerlis. "Resonantly Interacting Fermions in a Box". In: *Phys. Rev. Lett.* 106.23 (June 2011), p. 235303. DOI: 10.1103/PhysRevLett.106.235303.

[48] Michael McNeil Forbes et al. "Stability Criteria for Breached-Pair Superfluidity". In: *Physical Review Letters* 94.1 (Jan. 2005), p. 017001. DOI: 10.1103/PhysRevLett.94.017001.

[49] W. M. C. Foulkes et al. "Quantum Monte Carlo Simulations of Solids". In: *Reviews of Modern Physics* 73.1 (Jan. 2001), pp. 33–83. DOI: 10.1103/RevModPhys.73.33.

[50] Peter Fulde and Richard A. Ferrell. "Superconductivity in a Strong Spin-Exchange Field". In: *Physical Review* 135.3A (Aug. 1964), A550–A563. DOI: 10.1103/PhysRev.135.A550.

[51] Alexander Galea et al. "Diffusion Monte Carlo Study of Strongly Interacting Two-Dimensional Fermi Gases". In: *Physical Review A* 93.2 (Feb. 2016), p. 023602. DOI: 10.1103/PhysRevA.93.023602.

[52] Stefano Gandolfi, Alexandros Gezerlis, and J. Carlson. "Neutron Matter from Low to High Density". In: *Annual Review of Nuclear and Particle Science* 65.1 (2015), pp. 303–328. DOI: 10.1146/annurev-nucl-102014-021957. eprint: https://doi.org/10.1146/annurev-nucl-102014-021957.

[53] Nicholas Gao and Stephan Günnemann. *Ab-Initio Potential Energy Surfaces by Pairing GNNs with Neural Wave Functions*. Mar. 2022. DOI: 10.48550/arXiv.2110.05064. arXiv: 2110.05064 [cs].

[54] Nicholas Gao and Stephan Günnemann. "Generalizing Neural Wave Functions". In: *Proceedings of the 40th International Conference on Machine Learning*. Ed. by Andreas Krause et al. Vol. 202. Proceedings of Machine Learning Research. PMLR, July 2023, pp. 10708–10726.

[55] Nicholas Gao and Stephan Günnemann. *Sampling-Free Inference for Ab-Initio Potential Energy Surface Networks*. Mar. 2023. DOI: 10.48550/arXiv.2205.14962. arXiv: 2205.14962 [physics].

[56] I. M. Gel′fand and G. E. Shilov. *Generalized Functions, Volume 1*. American Mathematical Soc., Apr. 2016.

[57] Claudio Genovese et al. "General Correlated Geminal Ansatz for Electronic Structure Calculations: Exploiting Pfaffians in Place of Determinants". In: *Journal of Chemical Theory and Computation* 16.10 (Oct. 2020), pp. 6114–6131. DOI: 10.1021/acs.jctc.0c00165.

[58] Leon Gerard et al. "Gold-Standard Solutions to the Schrödinger Equation Using Deep Learning: How Much Physics Do We Need?" In: *Advances in Neural Information Processing Systems*. Ed. by S. Koyejo et al. Vol. 35. Curran Associates, Inc., 2022, pp. 10282–10294.

[59] Alexandros Gezerlis and J. Carlson. "Low-Density Neutron Matter". In: *Physical Review C* 81.2 (Feb. 2010), p. 025803. DOI: 10.1103/PhysRevC.81.025803.

[60] Stefano Giorgini, Lev P. Pitaevskii, and Sandro Stringari. "Theory of Ultracold Atomic Fermi Gases". In: *Rev. Mod. Phys.* 80.4 (Oct. 2008), pp. 1215–1274. DOI: 10.1103/RevModPhys.80.1215.

[61] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. Adaptive Computation and Machine Learning. Cambridge, Massachusetts: The MIT Press, 2016.

[62] *Google-Deepmind/Ferminet*. Google DeepMind. Apr. 2024.

[63] L P Gor'Kov and T K Melik-Barkhudarov. "CONTRIBUTION TO THE THEORY OF SUPERFLUIDITY IN AN IMPERFECT FERMI GAS". In: *Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki* 40 (May 1961), pp. 1452–1458.

[64] Olga Goulko and Matthew Wingate. "Numerical Study of the Unitary Fermi Gas across the Superfluid Transition". In: *Physical Review A: Atomic, Molecular, and Optical Physics* 93.5 (May 2016), p. 053604. DOI: 10.1103/PhysRevA.93.053604.

[65] Izrail Solomonovich Gradshteyn and Daniel Zwillinger. *Table of Integrals, Series, and Products*. 8th edition. Amsterdam Boston: Elsevier, Academic Press is an imprint of Elsevier, 2015.

[66] Jan Gukelberger et al. "p-Wave Superfluidity by Spin-Nematic Fermi Surface Deformation". In: *Physical Review Letters* 113.19 (Nov. 2014), p. 195301. DOI: 10.1103/PhysRevLett.113.195301.

[67] Kai Guther et al. "Binding Curve of the Beryllium Dimer Using Similarity-Transformed FCIQMC: Spectroscopic Accuracy with Triple-Zeta Basis Sets". In: *The Journal of Chemical Physics* 155.1 (July 2021), p. 011102. DOI: 10.1063/5.0055575.

[68] Jiequn Han, Linfeng Zhang, and Weinan E. "Solving Many-Electron Schrödinger Equation Using Deep Neural Networks". In: *Journal of Computational Physics* 399 (Dec. 2019), p. 108929. DOI: 10.1016/j.jcp.2019.108929.

[69] Satoru Hayami, Yuki Yanagi, and Hiroaki Kusunose. "Momentum-Dependent Spin Splitting by Collinear Antiferromagnetic Ordering". In: *Journal of the Physical Society of Japan* 88.12 (Dec. 2019), p. 123702. DOI: 10.7566/JPSJ.88.123702.

[70] Lianyi He and Pengfei Zhuang. "Stable Sarma State in Two-band Fermi Systems". In: *Physical Review B* 79.2 (Jan. 2009), p. 024511. DOI: 10.1103/PhysRevB.79.024511. arXiv: 0805.4699 [cond-mat, physics:hep-ph].

[71] Lianyi He et al. "Quantum Fluctuations in the BCS-BEC Crossover of Two-Dimensional Fermi Gases". In: *Physical Review A* 92.2 (Aug. 2015), p. 023620. DOI: 10.1103/PhysRevA.92.023620. arXiv: 1506.07156 [cond-mat].

[72] Rongzheng He et al. "Superfluid Condensate Fraction and Pairing Wave Function of the Unitary Fermi Gas". In: *Physical Review A* 101.6 (June 2020), p. 063615. DOI: 10.1103/PhysRevA.101.063615.

[73] Jan Hermann, Zeno Schätzle, and Frank Noé. "Deep-Neural-Network Solution of the Electronic Schrödinger Equation". In: *Nature Chemistry* 12.10 (Oct. 2020), pp. 891–897.

[74] Jan Hermann et al. "Ab-Initio Quantum Chemistry with Neural-Network Wavefunctions". In: *Nature Reviews Chemistry* 7.10 (Aug. 2023), pp. 692–709. DOI: 10.1038/s41570-023-00516-8. arXiv: 2208.12590 [physics].

[75] Mohamed Hibat-Allah et al. "Recurrent Neural Network Wave Functions". In: *Physical Review Research* 2.2 (June 2020), p. 023358. DOI: 10.1103/PhysRevResearch.2.023358.

[76] P. Hohenberg and W. Kohn. "Inhomogeneous Electron Gas". In: *Physical Review* 136.3B (Nov. 1964), B864–B871. DOI: 10.1103/PhysRev.136.B864.

[77] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. "Multilayer Feedforward Networks Are Universal Approximators". In: *Neural Networks* 2.5 (1989), pp. 359–366. DOI: 10.1016/0893-6080(89)90020-8.

[78] Hui Hu and Xia-Ji Liu. "Mean-Field Phase Diagrams of Imbalanced Fermi Gases near a Feshbach Resonance". In: *Physical Review A* 73.5 (May 2006), p. 051603. DOI: 10.1103/PhysRevA.73.051603.

[79] Masao Ishikawa and Masato Wakayama. "Minor Summation Formula of Pfaffians". In: *Linear and Multilinear Algebra* 39.3 (Aug. 1995), pp. 285–305. DOI: 10.1080/03081089508818403.

[80] Masao Ishikawa and Masato Wakayama. "Minor Summation Formula of Pfaffians and Schur Function Identities". In: *Proceedings of the Japan Academy, Series A, Mathematical Sciences* 71.3 (Jan. 1995). DOI: 10.3792/pjaa.71.54.

[81] S. Jensen, C. N. Gilbreth, and Y. Alhassid. "The Contact in the Unitary Fermi Gas across the Superfluid Phase Transition". In: *Physical Review Letters* 125.4 (July 2020), p. 043402. DOI: 10.1103/PhysRevLett.125.043402. arXiv: 1906.10117 [cond-mat, physics:hep-lat, physics:nucl-th].

[82] Tosio Kato. "On the Eigenfunctions of Many-particle Systems in Quantum Mechanics". In: *Communications on Pure and Applied Mathematics* 10.2 (Jan. 1957), pp. 151–177. DOI: 10.1002/cpa.3160100201.

[83] N. N. Khuri et al. "Low-Energy Potential Scattering in Two and Three Dimensions". In: *Journal of Mathematical Physics* 50.7 (July 2009), p. 072105. DOI: 10.1063/1.3167803.

[84] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. Jan. 2017. DOI: 10.48550/arXiv.1412.6980. arXiv: 1412.6980 [cs].

[85] W. Kohn and L. J. Sham. "Self-Consistent Equations Including Exchange and Correlation Effects". In: *Physical Review* 140.4A (Nov. 1965), A1133–A1138. DOI: 10.1103/PhysRev.140.A1133.

[86] W. J. Kwon et al. "Strongly Correlated Superfluid Order Parameters from Dc Josephson Supercurrents". In: *Science* 369.6499 (2020), pp. 84–88. DOI: 10.1126/science.aaz2463. eprint: https://www.science.org/doi/pdf/10.1126/science.aaz2463.

[87] A. I. Larkin and Y. N. Ovchinnikov. "Nonuniform State of Superconductors". In: *Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki* 47 (1964), pp. 1136–1146.

[88] A. J. Leggett. "Diatomic Molecules and Cooper Pairs". In: *Modern Trends in the Theory of Condensed Matter*. Ed. by Andrzej Pękalski and Jerzy A. Przystawa. Berlin, Heidelberg: Springer, 1980, pp. 13–27. DOI: 10.1007/BFb0120125.

[89] Anthony James Leggett. *Quantum Liquids: Bose Condensation and Cooper Pairing in Condensed-Matter Systems*. 1st ed. Oxford University PressOxford, Sept. 2006. DOI: 10.1093/acprof:oso/9780198526438.001.0001.

[90] Xiang Li, Zhe Li, and Ji Chen. "Ab Initio Calculation of Real Solids via Neural Network Ansatz". In: *Nature Communications* 13.1 (Dec. 2022), p. 7895. DOI: 10.1038/s41467-022-35627-1.

[91] Xiang Li et al. "Fermionic Neural Network with Effective Core Potential". In: *Physical Review Research* 4.1 (Jan. 2022), p. 013021. DOI: 10.1103/PhysRevResearch.4.013021.

[92] Xin Li, Jindř řich Kolorenč č, and Lubos Mitas. "Atomic Fermi Gas in the Unitary Limit by Quantum Monte Carlo Methods: Effects of the Interaction Range". In: *Physical Review A: Atomic, Molecular, and Optical Physics* 84.2 (Aug. 2011), p. 023615. DOI: 10.1103/PhysRevA.84.023615.

[93] Jeffmin Lin, Gil Goldshlager, and Lin Lin. "Explicitly Antisymmetrized Neural Network Layers for Variational Monte Carlo Simulation". In: *Journal of Computational Physics* 474 (Feb. 2023), p. 111765. DOI: 10.1016/j.jcp.2022.111765. arXiv: 2112.03491 [physics].

[94] W. Vincent Liu. "Breached Pair Superfluidity: A Brief Review". In: Apr. 2022, pp. 163–175. DOI: 10.1142/9789811251948_0013. arXiv: 2109.02208 [cond-mat].

[95] Wan Tong Lou et al. "Neural Wave Functions for Superfluids". In: *Physical Review X* 14.2 (May 2024), p. 021030. DOI: 10.1103/PhysRevX.14.021030.

[96] Di Luo and Bryan K. Clark. "Backflow Transformations via Neural Networks for Quantum Many-Body Wave-Functions". In: *Physical Review Letters* 122.22 (June 2019), p. 226401. DOI: 10.1103/PhysRevLett.122.226401. arXiv: 1807.10770 [cond-mat, physics:physics].

[97] Mathias Macêdo-Lima and Lucas Madeira. "Scattering Length and Effective Range of Microscopic Two-Body Potentials". In: *Revista Brasileira de Ensino de Física* 45 (2023), e20230079. DOI: 10.1590/1806-9126-rbef-2023-0079.

[98] Kazushige Machida and Hiizu Nakanishi. "Superconductivity under a Ferromagnetic Molecular Field". In: *Physical Review B* 30.1 (July 1984), pp. 122–133. DOI: 10.1103/PhysRevB.30.122.

[99] Massimo Mannarelli, Giuseppe Nardulli, and Marco Ruggieri. "Evaluating the Phase Diagram of Superconductors with Asymmetric Spin Populations". In: *Physical Review A* 74.3 (Sept. 2006), p. 033606. DOI: 10.1103/PhysRevA.74.033606.

[100] Mariapia Marchi et al. "Resonating Valence Bond Wave Function with Molecular Orbitals: Application to First-Row Molecules". In: *The Journal of Chemical Physics* 131.15 (Oct. 2009), p. 154116. DOI: 10.1063/1.3249966.

[101] M. Marini, F. Pistolesi, and G. C. Strinati. "Evolution from BCS Superconductivity to Bose Condensation: Analytic Results for the Crossover in Three Dimensions". In: *The European Physical Journal B* 1.2 (Feb. 1998), pp. 151–159. DOI: 10.1007/s100510050165. arXiv: cond-mat/9703160.

[102] James Martens and Roger Grosse. "Optimizing Neural Networks with Kronecker-factored Approximate Curvature". In: *Proceedings of the 32nd International Conference on Machine Learning*. Ed. by Francis Bach and David Blei. Vol. 37. Proceedings of Machine Learning Research. Lille, France: PMLR, July 2015, pp. 2408–2417.

[103] Igor I. Mazin et al. "Prediction of Unconventional Magnetism in Doped $FeSb_2$". In: *Proceedings of the National Academy of Sciences* 118.42 (Oct. 2021), e2108924118. DOI: 10.1073/pnas.2108924118.

[104] W. Meissner and R. Ochsenfeld. "Ein neuer Effekt bei Eintritt der Supraleitfähigkeit". In: *Die Naturwissenschaften* 21.44 (Nov. 1933), pp. 787–788. DOI: 10.1007/BF01504252.

[105] Andrew J. Morris, P. López Ríos, and R. J. Needs. "Ultracold Atoms at Unitarity within Quantum Monte Carlo Methods". In: *Physical Review A* 81.3 (Mar. 2010), p. 033619. DOI: 10.1103/PhysRevA.81.033619.

[106] Olexei I. Motrunich and Matthew P. A. Fisher. "D-Wave Correlated Critical Bose Liquids in Two Dimensions". In: *Physical Review B* 75.23 (June 2007). DOI: 10.1103/physrevb.75.235116.

[107] R. J. Needs et al. "Variational and Diffusion Quantum Monte Carlo Calculations with the CASINO Code". In: *The Journal of Chemical Physics* 152.15 (Apr. 2020), p. 154106. DOI: 10.1063/1.5144288. arXiv: 2003.06506 [cond-mat, physics:physics].

[108] Predrag Nikolić and Subir Sachdev. "Renormalization-Group Fixed Points, Universal Phase Diagram, and 1/ N Expansion for Quantum Liquids with Interactions near the Unitarity Limit". In: *Physical Review A: Atomic, Molecular, and Optical Physics* 75.3 (2007), p. 033608.

[109] Yusuke Nishida and Dam Thanh Son. "ϵ Expansion for a Fermi Gas at Infinite Scattering Length". In: *Physical Review Letters* 97.5 (2006), p. 050403.

[110] P. Nozières and S. Schmitt-Rink. "Bose Condensation in an Attractive Fermion Gas: From Weak to Strong Coupling Superconductivity". In: *Journal of Low Temperature Physics* 59.3 (May 1985), pp. 195–211. DOI: 10.1007/BF00683774.

[111] Soichi Okada. "On the Generating Functions for Certain Classes of Plane Partitions". In: *Journal of Combinatorial Theory, Series A* 51.1 (May 1989), pp. 1–23. DOI: 10.1016/0097-3165(89)90073-3.

[112] Georgios Palkanoglou, Fotis K. Diakonos, and Alexandros Gezerlis. "From Odd-Even Staggering to the Pairing Gap in Neutron Matter". In: *Physical Review C* 102.6 (Dec. 2020), p. 064324. DOI: 10.1103/PhysRevC.102.064324.

[113] C.-H. Pao, Shin-Tza Wu, and S.-K. Yip. "Superfluid Stability in BEC-BCS Crossover". In: *Physical Review B* 73.13 (Apr. 2006), p. 132506. DOI: 10.1103/PhysRevB.73.132506. arXiv: cond-mat/0506437.

[114] T. Papenbrock and G. F. Bertsch. "Pairing in Low-Density Fermi Gases". In: *Phys. Rev. C* 59.4 (Apr. 1999), pp. 2052–2055. DOI: 10.1103/PhysRevC.59.2052.

[115] Meera M. Parish. "The BCS-BEC Crossover". In: vol. 3. Nov. 2014, pp. 179–197. DOI: 10.1142/9781783264766_0009. arXiv: 1402.5171 [cond-mat].

[116] Robert G. Parr and Yang Weitao. *Density-Functional Theory of Atoms and Molecules*. Oxford University Press, Jan. 1995. DOI: 10.1093/oso/9780195092769.001.0001.

[117] Gabriel Pescia et al. "Neural-Network Quantum States for Periodic Systems in Continuous Space". In: *Physical Review Research* 4.2 (May 2022), p. 023138. DOI: 10.1103/PhysRevResearch.4.023138.

[118] C J Pethick and D G Ravenhall. "Matter at Large Neutron Excess and the Physics of Neutron-Star Crusts". In: *Annual Review of Nuclear and Particle Science* 45.1 (1995), pp. 429–484. DOI: 10.1146/annurev.ns.45.120195.002241.

[119] Christopher Pethick and Henrik Smith. *Bose-Einstein Condensation in Dilute Gases*. Reprint. with corr. Cambridge: Cambridge University Press, 2006.

[120] D. S. Petrov, M. A. Baranov, and G. V. Shlyapnikov. "Superfluid Transition in Quasi-Two-Dimensional Fermi Gases". In: *Physical Review A* 67.3 (Mar. 2003), p. 031601. DOI: 10.1103/PhysRevA.67.031601.

[121] David Pfau et al. "Ab-Initio Solution of the Many-Electron Schr\"odinger Equation with Deep Neural Networks". In: *Physical Review Research* 2.3 (Sept. 2020), p. 033429. DOI: 10.1103/PhysRevResearch.2.033429. arXiv: 1909.02487 [physics].

[122] David Pfau et al. *Natural Quantum Monte Carlo Computation of Excited States*. Feb. 2024. DOI: 10.48550/arXiv.2308.16848. arXiv: 2308.16848 [physics, physics:quant-ph].

[123] Philip Phillips and Denis Dalidovich. "The Elusive Bose Metal". In: *Science* 302.5643 (Oct. 2003), pp. 243–247. DOI: 10.1126/science.1088253.

[124] P Pieri and G C Strinati. "Popov Approximation for Composite Bosons in the BCS-BEC Crossover". In: *PHYSICAL REVIEW B* ().

[125] M. Pini, P. Pieri, and G. Calvanese Strinati. "Evolution of an Attractive Polarized Fermi Gas: From a Fermi Liquid of Polarons to a Non-Fermi Liquid at the Fulde-Ferrell-Larkin-Ovchinnikov Quantum Critical Point". In: *Physical Review B* 107.5 (Feb. 2023), p. 054505. DOI: 10.1103/PhysRevB.107.054505.

[126] L. Pisani, P. Pieri, and G. Calvanese Strinati. "Spatial Emergence of Off-Diagonal Long-Range Order throughout the BCS-BEC Crossover". In: *Physical Review B* 105.5 (Feb. 2022), p. 054505. DOI: 10.1103/PhysRevB.105.054505. arXiv: 2112.00614.

[127] Leo Radzihovsky. "Fluctuations and Phase Transitions in Larkin-Ovchinnikov Liquid Crystal States of Population-Imbalanced Resonant Fermi Gas". In: *Physical Review A* 84.2 (Aug. 2011), p. 023611. DOI: 10.1103/PhysRevA.84.023611. arXiv: 1102.4903 [cond-mat].

[128] M. Randeria, W. Zwerger, and M. Zwierlein. "The BCS–BEC Crossover and the Unitary Fermi Gas". In: *The BCS-BEC Crossover and the Unitary Fermi Gas*. Ed. by Wilhelm Zwerger. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 1–32. DOI: 10.1007/978-3-642-21978-8_1.

[129] Mohit Randeria. "Pre-Pairing for Condensation". In: *Nature Physics* 6.8 (Aug. 2010), pp. 561–562. DOI: 10.1038/nphys1748.

[130] Mohit Randeria, Ji-Min Duan, and Lih-Yir Shieh. "Bound States, Cooper Pairing, and Bose Condensation in Two Dimensions". In: *Physical Review Letters* 62.9 (Feb. 1989), pp. 981–984. DOI: 10.1103/PhysRevLett.62.981.

[131] Mohit Randeria, Ji-Min Duan, and Lih-Yir Shieh. "Superconductivity in a Two-Dimensional Fermi Gas: Evolution from Cooper Pairing to Bose Condensation". In: *Physical Review B* 41.1 (Jan. 1990), pp. 327–343. DOI: 10.1103/PhysRevB.41.327.

[132] Weiluo Ren et al. "Towards the Ground State of Molecules via Diffusion Monte Carlo on Neural Networks". In: *Nature Communications* 14.1 (Apr. 2023), p. 1860. DOI: 10.1038/s41467-023-37609-3. arXiv: 2204.13903 [physics].

[133] K. F. Riley, M. P. Hobson, and S. J. Bence. *Mathematical Methods for Physics and Engineering*. 3rd ed. Cambridge: Cambridge University Press, 2006. DOI: 10.1017/cbo9780511810763.

[134] Leonard Sidney Rodberg, Raphael Morten Thaler, and Roy M. Thaler. *Introduction to the Quantum Theory of Scattering*. 2. printing. Pure and Applied Physics 26. New York: Academic Press, 1970.

[135] C. A. R. Sá de Melo, Mohit Randeria, and Jan R. Engelbrecht. "Crossover from BCS to Bose Superconductivity: Transition Temperature and Time-Dependent Ginzburg-Landau Theory". In: *Physical Review Letters* 71.19 (Nov. 1993), pp. 3202–3205. DOI: 10.1103/PhysRevLett.71.3202.

[136] Jun J. Sakurai. *Modern Quantum Mechanics*. 3rd ed. Cambridge: Cambridge University Press, 2020.

[137] G. Sarma. "On the Influence of a Uniform Exchange Field Acting on the Spins of the Conduction Electrons in a Superconductor". In: *Journal of Physics and Chemistry of Solids* 24.8 (Aug. 1963), pp. 1029–1032. DOI: 10.1016/0022-3697(63)90007-6.

[138] Michael Scherbela, Leon Gerard, and Philipp Grohs. *Towards a Foundation Model for Neural Network Wavefunctions*. Mar. 2023. DOI: 10.48550/arXiv.2303.09949. arXiv: 2303.09949 [physics].

[139] Matsuo Shigemasa et al. "Phase Diagram of the Fulde-Ferrell-Larkin-Ovchinnikov State in a Three-Dimensional Superconductor". In: *Journal of the Physical Society of Japan* (Nov. 2013). DOI: 10.1143/JPSJ.67.280.

[140] Hiroshi Shimahara. "Fulde-Ferrell State in Quasi-Two-Dimensional Superconductors". In: *Physical Review B* 50.17 (Nov. 1994), pp. 12760–12765. DOI: 10.1103/PhysRevB.50.12760.

[141] Hiroshi Shimahara. "Structure of the Fulde-Ferrell-Larkin-Ovchinnikov State in Two-Dimensional Superconductors". In: *Journal of the Physical Society of Japan* 67.3 (Mar. 1998), pp. 736–739. DOI: 10.1143/JPSJ.67.736. arXiv: cond-mat/9711017.

[142] Yong-il Shin et al. "Phase Diagram of a Two-Component Fermi Gas with Resonant Interactions". In: *Nature* 451.7179 (Feb. 2008), pp. 689–693. DOI: 10.1038/nature06473.

[143] GiBaik Sim and Johannes Knolle. *Pair Density Waves and Supercurrent Diode Effect in Altermagnets*. July 2024. DOI: 10.48550/arXiv.2407.01513. arXiv: 2407.01513 [cond-mat].

[144] Libor Šmejkal, Jairo Sinova, and Tomas Jungwirth. "Emerging Research Landscape of Altermagnetism". In: *Physical Review X* 12.4 (Dec. 2022), p. 040501. DOI: 10.1103/PhysRevX.12.040501.

[145] Libor Šmejkal et al. "Crystal Time-Reversal Symmetry Breaking and Spontaneous Hall Effect in Collinear Antiferromagnets". In: *Science Advances* 6.23 (June 2020), eaaz8809. DOI: 10.1126/sciadv.aaz8809.

[146] D. T. Son and M. A. Stephanov. "Phase Diagram of a Cold Polarized Fermi Gas". In: *Physical Review A* 74.1 (July 2006), p. 013614. DOI: 10.1103/PhysRevA.74.013614.

[147] Young-Ho Song et al. "Quantum Many-Body Calculations Using Body-Centered Cubic Lattices". In: *Physical Review C* 104.4 (Oct. 2021), p. 044304. DOI: 10.1103/PhysRevC.104.044304.

[148] Sandro Sorella. "Green Function Monte Carlo with Stochastic Reconfiguration". In: *Physical Review Letters* 80.20 (May 1998), pp. 4558–4561. DOI: 10.1103/PhysRevLett.80.4558.

[149] James S. Spencer et al. *Better, Faster Fermionic Neural Networks*. Nov. 2020. arXiv: 2011.07125 [physics].

[150] Elias M. Stein and Guido Weiss. *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton University Press, Nov. 1971.

[151] G. C. Strinati. "Pairing Fluctuations Approach to the BCS–BEC Crossover". In: *The BCS-BEC Crossover and the Unitary Fermi Gas*. Ed. by Wilhelm Zwerger. Berlin, Heidelberg: Springer, 2012, pp. 99–126. DOI: 10.1007/978-3-642-21978-8_4.

[152] Giancarlo Calvanese Strinati et al. "The BCS–BEC Crossover: From Ultra-Cold Fermi Gases to Nuclear Systems". In: *Physics Reports-review Section of Physics Letters* 738 (2018), pp. 1–76. DOI: 10.1016/j.physrep.2018.02.004.

[153] Attila Szabo and Neil S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Mineola, N.Y: Dover Publications, 1996.

[154] Satoshi Takada and Takeo Izuyama. "Superconductivity in a Molecular Field. I". In: *Progress of Theoretical Physics* 41.3 (Mar. 1969), pp. 635–663. DOI: 10.1143/PTP.41.635.

[155] E. Taylor, A. Griffin, and Y. Ohashi. "Spin-Polarized Fermi Superfluids as Bose-Fermi Mixtures". In: *Physical Review A* 76.2 (Aug. 2007), p. 023614. DOI: 10.1103/PhysRevA.76.023614. arXiv: cond-mat/0703433.

[156] Michael Tinkham. *Introduction to Superconductivity: Second Edition*. Courier Corporation, June 2004.

[157] Julien Toulouse and C. J. Umrigar. "Optimization of Quantum Monte Carlo Wave Functions by Energy Minimization". In: *The Journal of Chemical Physics* 126.8 (Feb. 2007), p. 084102. DOI: 10.1063/1.2437215.

[158] Léon Van Hove. "Abnormal Behaviour of the Dilute Fermi Gas for Negative Two Body Scattering Length". In: *Physica* 25.7 (Jan. 1959), pp. 849–858. DOI: 10.1016/0031-8914(59)90008-4.

[159] James Richard Varley. "Supersolidity and Superfluidity in Two Dimensions". PhD thesis.

[160] B. J. Verhaar et al. "Scattering Length and Effective Range in Two Dimensions: Application to Adsorbed Hydrogen Atoms". In: *Journal of Physics A: Mathematical and General* 17.3 (Feb. 1984), p. 595. DOI: 10.1088/0305-4470/17/3/020.

[161] Luciano Loris Viteritti, Riccardo Rende, and Federico Becca. "Transformer Variational Wave Functions for Frustrated Quantum Spin Systems". In: *Physical Review Letters* 130.23 (June 2023), p. 236401. DOI: 10.1103/PhysRevLett.130.236401.

[162] Ingrid von Glehn, James S. Spencer, and David Pfau. *A Self-Attention Ansatz for Ab-initio Quantum Chemistry*. Apr. 2023. DOI: 10.48550/arXiv.2211.13672. arXiv: 2211.13672 [physics].

[163] G. N. 1886- Watson. *A Treatise on the Theory of Bessel Functions - Scholar's Choice Edition*. Creative Media Partners, LLC, Feb. 2015.

[164] Darryl Foo Chuan Wei. "Communal Pairing in Fermi Gases with Attractive Interactions". PhD thesis.

[165] Max Wilson et al. "Neural Network Ansatz for Periodic Wave Functions and the Homogeneous Electron Gas". In: *Physical Review B* 107.23 (June 2023), p. 235139. DOI: 10.1103/PhysRevB.107.235139.

[166] Max Wilson et al. *Simulations of State-of-the-Art Fermionic Neural Network Wave Functions with Diffusion Monte Carlo*. Mar. 2021. arXiv: 2103.12570 [physics, physics:quant-ph].

[167] Samuel S. M. Wong. *Introductory Nuclear Physics*. 1st ed. Wiley, Nov. 1998. DOI: 10.1002/9783527617906.

[168] C. N. Yang. "Concept of Off-Diagonal Long-Range Order and the Quantum Phases of Liquid He and of Superconductors". In: *Rev. Mod. Phys.* 34.4 (Oct. 1962), pp. 694–704. DOI: 10.1103/RevModPhys.34.694.

[169] D. C. Youla. "A Normal Form for a Matrix under the Unitary Congruence Group". In: *Canadian Journal of Mathematics* 13 (1961), pp. 694–704. DOI: 10.4153/CJM-1961-059-8.

[170] Lin-Ding Yuan et al. "Giant Momentum-Dependent Spin Splitting in Centrosymmetric Low-Z Antiferromagnets". In: *Physical Review B* 102.1 (July 2020), p. 014422. DOI: 10.1103/PhysRevB.102.014422.

[171] Song-Bo Zhang, Lun-Hui Hu, and Titus Neupert. "Finite-Momentum Cooper Pairing in Proximitized Altermagnets". In: *Nature Communications* 15.1 (Feb. 2024), p. 1801. DOI: 10.1038/s41467-024-45951-3. arXiv: 2302.13185 [cond-mat].

[172] Bruno Zumino. "Normal Forms of Complex Matrices". In: *Journal of Mathematical Physics* 3.5 (Sept. 1962), pp. 1055–1057. DOI: 10.1063/1.1724294.

[173] M. W. Zwierlein et al. "Formation Dynamics of a Fermion Pair Condensate". In: *Physical Review Letters* 94.18 (May 2005), p. 180401. DOI: 10.1103/PhysRevLett.94.180401.

[174] Martin W Zwierlein. "High-Temperature Superfluidity in an Ultracold Fermi Gas". PhD thesis.

Appendix A

Scattering Theory

In this section, we provide a brief introduction to the scattering theory in non-relativistic quantum mechanics. First, we take a general approach to scattering that is based on the Lippmann-Schwinger equation, and describe the asymptotic behaviour of the wave function in both 2D and 3D. We also derive the Lippmann-Schwinger equations for the scattering amplitude, which will be used in later sections. Next, we introduce the method of partial waves separately in 2D and 3D, and define the notion of an s-wave scattering length in both cases. Finally, we combine the two sections together and obtain the relations between the scattering length and the strength of the contact interaction that we used in the background chapter.

A.1 Lippmann-Schwinger Equation

In a broad sense, the aim of scattering theory is to study the properties of some unknown object by shooting particles at it, and obtain the properties based on how the particles are deflected. To formulate the scattering problem in the language of quantum mechanics, the unknown object can be treated as a potential $V(\mathbf{r})$ [1], and the properties of the particle can be found by solving the following time-independent Schrödinger equation (TISE)

$$\hat{H}\left|\psi\right\rangle = \left[\frac{\hat{\mathbf{p}}^2}{2m} + \hat{V}\right]\left|\psi\right\rangle = E\left|\psi\right\rangle . \tag{A.1}$$

In the absence of the potential $\hat{V}$, the TISE reduces to $\hat{H}_0\left|\phi\right\rangle = \frac{p^2}{2m}\left|\phi\right\rangle$ where $\hat{H}_0 = \frac{\hat{\mathbf{p}}^2}{2m}$. Note that the eigenstates $\left|\phi\right\rangle$ of this TISE are the same as the momentum eigenstates $\left|\mathbf{p}\right\rangle$, but for generality, we write them as $\left|\phi\right\rangle$ for now. If we assume the potential goes to zero as $|\mathbf{r}| \to \infty$, we would expect the particle to be in a momentum eigenstate at large distances. If we make a further assumption that the scattering is *elastic*, the energy must be conserved throughout the whole scattering process, which means that the free-particle state $\left|\phi\right\rangle$ has the same energy E in

[1] By defining the potential to be $V(\mathbf{r})$, we have implicitly assumed it to be *local*, *i.e.* $\left\langle\mathbf{r}\right|\hat{V}\left|\mathbf{r}'\right\rangle = V(\mathbf{r})\delta^{(d)}(\mathbf{r}-\mathbf{r}')$. This is sufficient for our work because we only work with local potentials.

Eq. (A.1). This allows us to write down an alternative form for Eq. (A.1)

$$|\psi\rangle = |\phi\rangle + \frac{1}{E - \hat{H}_0 \pm i0}\hat{V}\,|\psi\rangle\,, \tag{A.2}$$

where the denominator of the second term is made slightly complex to avoid singularities [2]. This equation is known as the *Lippmann-Schwinger equation* in its most general form, as it is independent of any representation. Next, we define the operator $\hat{T}$ such that it satisfies

$$\hat{V}\,|\psi\rangle = \hat{T}\,|\phi\rangle\,. \tag{A.3}$$

Combining this equation with the Lippmann-Schwinger equation and projecting it to momentum space, we get

$$\langle \mathbf{k}'|\,\hat{T}\,|\phi\rangle = \langle \mathbf{k}'|\,\hat{V}\,|\phi\rangle + \langle \mathbf{k}'|\,\hat{V}\frac{1}{E - \hat{H}_0 + i0}\hat{T}\,|\phi\rangle\,. \tag{A.4}$$

Using the fact that $|\phi\rangle$ are momentum eigenstates, we set $|\phi\rangle = |\mathbf{k}\rangle$ to obtain the following Lippmann-Schwinger equation for the matrix $T_{\mathbf{k}'\mathbf{k}}$,

$$T_{\mathbf{k}'\mathbf{k}} = V_{\mathbf{k}'\mathbf{k}} + \frac{2m}{\hbar^2}\sum_{\mathbf{k}''} V_{\mathbf{k}'\mathbf{k}''}\frac{1}{k^2 - k''^2 + i0}T_{\mathbf{k}''\mathbf{k}}. \tag{A.5}$$

But how is this related to studying a system of interacting particles, such as the dilute Fermi gas? To see that, we first consider the simplest case – two particles of equal mass m' interacting in space. It is possible to separate their motion in the centre-of-mass frame from their relative motion. If we ignore their bulk properties and only focus on their relative motions, the two-body problem can be reduced to an effective one-body problem of a particle with *half the mass* of the original two particles $m = m'/2$, interacting under the influence of a potential. Although we have used the symbol m throughout this section, readers should keep in mind that when applied to two-particle problems, one should divide the mass by two $m \to m/2$. For now, we will stick with this convention unless stated otherwise.

A.1.1 Scattering in dilute Fermi gases

To apply the Lippmann-Schwinger equation to dilute Fermi gases, we first define the basis and the boundary conditions of the system. As we are interested in simulating a dilute Fermi gas in a box, we set our basis to be plane waves, given by $\langle \mathbf{r}|\mathbf{k}\rangle = \frac{1}{\sqrt{\Omega}}e^{i\mathbf{k}\cdot\mathbf{r}}$ with the normalisations chosen to be

$$\langle \mathbf{k}|\mathbf{k}'\rangle = \delta_{\mathbf{k}\mathbf{k}'}, \qquad \langle \mathbf{r}|\mathbf{r}'\rangle = \delta^{(d)}(\mathbf{r} - \mathbf{r}') \tag{A.6}$$

[2] The term "$i0$" might be written as $i\epsilon$ or $i\delta$ in some literature, but they all serve the same purpose, which is to represent an infinitesimal imaginary number to avoid divergences. The $\pm$ sign indicates the direction (incoming or outgoing) of the spherical wave, where we have let it to be positive from here, as it does not affect our disucssion.

where $\delta_{\mathbf{k}\mathbf{k}'}$ is the Kronecker delta and $\delta^{(d)}(\mathbf{r}-\mathbf{r}')$ is the Dirac delta function. Now, we make use of the Lippmann-Schwinger equation again and project it onto real space

$$\psi_{\mathbf{k}}(\mathbf{r}) = \frac{e^{i\mathbf{k}\cdot\mathbf{r}}}{\sqrt{\Omega}} + \frac{2m}{\hbar^2\Omega}\int d\mathbf{r}' V(\mathbf{r}')\psi_{\mathbf{k}}(\mathbf{r}')\sum_{\mathbf{k}'}\frac{e^{i\mathbf{k}'\cdot(\mathbf{r}-\mathbf{r}')}}{k^2-k'^2+i0} \tag{A.7}$$

where the subscript $\mathbf{k}$ indicates the momentum of the incident wave, and we have assumed the potential to be local $V(\mathbf{r},\mathbf{r}') = V(\mathbf{r})\delta^{(d)}(\mathbf{r}-\mathbf{r}')$. The final step is to turn the summation over momentum into an integral. Up to this point, our the derivation is valid for any spatial dimension. This transformation, however, is dependent on the spatial dimension. As we worked in both 2D and 3D, we will discuss both cases in this section.

Using the approximation $\sum_{\mathbf{k}} = \int \frac{\Omega}{(2\pi)^d} d\mathbf{k}$ to calculate the summation over momentum

$$\sum_{\mathbf{k}'}\frac{e^{i\mathbf{k}'\cdot(\mathbf{r}-\mathbf{r}')}}{k^2-k'^2+i0} \approx \int d\mathbf{k}'\frac{\Omega}{(2\pi)^d}\frac{e^{i\mathbf{k}'\cdot(\mathbf{r}-\mathbf{r}')}}{k^2-k'^2+i0}, \tag{A.8}$$

and is evaluated in 2D and 3D respectively [2, 65, 134, 136]

$$\text{2D:} \qquad \int d\mathbf{k}'\frac{\Omega}{(2\pi)^2}\frac{e^{i\mathbf{k}'\cdot(\mathbf{r}-\mathbf{r}')}}{k^2-k'^2+i0} = -\frac{i\Omega}{4}H_0^{(1)}(k|\mathbf{r}-\mathbf{r}'|) \tag{A.9}$$

$$\text{3D:} \qquad \int d\mathbf{k}'\frac{\Omega}{(2\pi)^3}\frac{e^{i\mathbf{k}'\cdot(\mathbf{r}-\mathbf{r}')}}{k^2-k'^2+i0} = -\frac{\Omega}{4\pi}\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|} \tag{A.10}$$

where $H_0^{(1)}(x)$ is the Hankel function of the first kind, and we have used contour integration to arrive at the final expression [3]. The wave functions in 2D and 3D thus become

$$\text{2D:} \qquad \psi_{\mathbf{k}}^{\text{2D}}(\mathbf{r}) = \frac{e^{i\mathbf{k}\cdot\mathbf{r}}}{\sqrt{\Omega}} - i\frac{2m}{4\hbar^2}\int d\mathbf{r}' V(\mathbf{r}')\psi_{\mathbf{k}}^{\text{2D}}(\mathbf{r}')H_0^{(1)}(k|\mathbf{r}-\mathbf{r}'|) \tag{A.11}$$

$$\text{3D:} \qquad \psi_{\mathbf{k}}^{\text{3D}}(\mathbf{r}) = \frac{e^{i\mathbf{k}\cdot\mathbf{r}}}{\sqrt{\Omega}} - \frac{2m}{4\pi\hbar^2}\int d\mathbf{r}' V(\mathbf{r}')\psi_{\mathbf{k}}^{\text{3D}}(\mathbf{r}')\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}. \tag{A.12}$$

As we are only interested in the asymptotic limit where $r \to \infty$ for dilute Fermi gases, where the microscopic detail of the interaction becomes irrelevant, we can approximate $|\mathbf{r}-\mathbf{r}'| \approx r - \frac{\mathbf{r}\cdot\mathbf{r}'}{r}$ and define the scattered wave vector to be $\mathbf{k}' = k'\hat{\mathbf{r}}$ with $k' = |\mathbf{k}'| = k$ from energy conservation. The asymptotic behaviour of the wave functions under this limit become

$$\text{2D:} \qquad \psi_{\mathbf{k}}^{\text{2D}}(\mathbf{r}) \xrightarrow{r\to\infty} \frac{1}{\sqrt{\Omega}}\left[e^{i\mathbf{k}\cdot\mathbf{r}} + f_{\mathbf{k}'\mathbf{k}}^{\text{2D}}\frac{e^{ikr}}{\sqrt{r}}\right] \tag{A.13}$$

$$\text{3D:} \qquad \psi_{\mathbf{k}}^{\text{3D}}(\mathbf{r}) \xrightarrow{r\to\infty} \frac{1}{\sqrt{\Omega}}\left[e^{i\mathbf{k}\cdot\mathbf{r}} + f_{\mathbf{k}'\mathbf{k}}^{\text{3D}}\frac{e^{ikr}}{r}\right], \tag{A.14}$$

[3] The detailed derivation is widely available in various literature on scattering theory and quantum mechanics, such as Ref. [65, 134, 136].

where

$$f^{2D}_{\mathbf{k}'\mathbf{k}} = -\frac{m}{\hbar^2}\sqrt{\frac{i\Omega}{2\pi k}}\int d\mathbf{r}' V(\mathbf{r}')\psi^{2D}_{\mathbf{k}}(\mathbf{r}')e^{-i\mathbf{k}'\cdot\mathbf{r}'} = -\frac{m\Omega}{\hbar^2}\sqrt{\frac{i}{2\pi k}}T^{2D}_{\mathbf{k}'\mathbf{k}} \tag{A.15}$$

$$f^{3D}_{\mathbf{k}'\mathbf{k}} = -\frac{m\sqrt{\Omega}}{2\pi\hbar^2}\int d\mathbf{r}' V(\mathbf{r}')\psi^{3D}_{\mathbf{k}}(\mathbf{r}')e^{-i\mathbf{k}'\cdot\mathbf{r}'} = -\frac{m\Omega}{2\pi\hbar^2}T^{3D}_{\mathbf{k}'\mathbf{k}} \tag{A.16}$$

are known as the *scattering amplitude*, and the second equalities follow from the relation

$$T_{\mathbf{k}'\mathbf{k}} = \langle \mathbf{k}' | \hat{V} | \psi \rangle = \frac{1}{\sqrt{\Omega}}\int d\mathbf{r} V(\mathbf{r})\psi_{\mathbf{k}}(\mathbf{r})e^{-i\mathbf{k}'\cdot\mathbf{r}}. \tag{A.17}$$

Hence, we obtain the Lippmann-Schwinger equations for the scattering amplitudes

$$f^{2D}_{\mathbf{k}'\mathbf{k}} = -\frac{m\Omega}{\hbar^2}\sqrt{\frac{i}{2\pi k}}V_{\mathbf{k}'\mathbf{k}} + \frac{2m}{\hbar^2}\sum_{\mathbf{k}''}V_{\mathbf{k}'\mathbf{k}''}\frac{1}{k^2 - k''^2 + i0}f^{2D}_{\mathbf{k}''\mathbf{k}} \tag{A.18}$$

$$f^{3D}_{\mathbf{k}'\mathbf{k}} = -\frac{m\Omega}{2\pi\hbar^2}V_{\mathbf{k}'\mathbf{k}} + \frac{2m}{\hbar^2}\sum_{\mathbf{k}''}V_{\mathbf{k}'\mathbf{k}''}\frac{1}{k^2 - k''^2 + i0}f^{3D}_{\mathbf{k}''\mathbf{k}}. \tag{A.19}$$

We will use these equations in later sections after introducing the method of partial waves.

A.2 Method of Partial Waves

In this section, we will take an alternative approach to scattering by looking at the time-independent Schrödinger equation in detail, and discuss how it connects with the previous section. As we have already seen in the last section, despite having taken a general approach to scattering theory, the asymptotic behaviours of the scattered waves in 2D and 3D can still be vastly different. As such, we will discuss the partial wave expansions in 2D and 3D separately. First, we will talk about the better known case of the two: the partial wave expansion in 3D.

A.2.1 Partial wave expansions in 3D

Consider a particle under the influence of a central potential $V(\mathbf{r})$. The Schrödinger equation in the real-space representation can be written as

$$\hat{H}\psi(\mathbf{r}) = -\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r}) + V(\mathbf{r})\psi(\mathbf{r}) = E\psi(\mathbf{r}). \tag{A.20}$$

Making a further assumption that the potential is isotropic such that $V(\mathbf{r}) \equiv V(r)$, the Hamiltonian commutes with the angular momentum operator, $\left[\hat{H}, \hat{\mathbf{L}}^2\right] = 0$, as the former does not depend on the polar angle θ or the azimuthal angle ϕ. This implies that the wave function is an eigenstate of $\hat{\mathbf{L}}^2$ such that $\hat{\mathbf{L}}^2\psi = l(l+1)\psi$. If we write the Laplacian in spherical coordinates,

Eq. (A.20) becomes

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\psi(\mathbf{r})\right)+\left[k^2-\frac{l(l+1)}{r^2}+\frac{2m}{\hbar^2}V(r)\right]\psi(\mathbf{r})=0 \tag{A.21}$$

where we have used $E=\frac{\hbar^2k^2}{2m}$. The wave function can be written as a sum of products of radial functions and angular functions

$$\psi(\mathbf{r})=\psi(r,\theta,\phi)=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}a_lR_l(r)Y_l^m(\theta,\phi) \tag{A.22}$$

where $Y_l^m(\theta,\phi)$ are the spherical harmonics. Due to the azimuthal symmetry of the potential, it is also possible to choose an orientation such that only the $m=0$ term is retained in the summation, giving

$$\psi(\mathbf{r})=\psi(r,\theta)=\sum_{l=0}^{\infty}\sqrt{\frac{(2l+1)}{4\pi}}a_lR_l(r)P_l(\cos\theta) \tag{A.23}$$

where we have used $Y_l^0(\theta,\phi)=\sqrt{\frac{(2l+1)}{4\pi}}P_l(\cos\theta)$. Under this setting, the direction of the incident wave $\mathbf{k}$ aligns with the z axis, *i.e.* $\mathbf{k}=k\hat{\mathbf{z}}$. As we are interested in the region where the details of the potential are irrelevant, we can solve the Schrödinger equation in that region by setting $V(r)=0$, provided that $V(r)$ is short range. This is done by substituting Eq. (A.23) into Eq. (A.20) and solving the separated differential equations for $R_l(r)$, yielding [133, 167]

$$R_l(r)=\alpha_lj_l(kr)+\beta_ln_l(kr)\xrightarrow{r\to\infty}\frac{\alpha_l'e^{ikr}-(-1)^l\beta_l'e^{-ikr}}{i^{l+1}kr} \tag{A.24}$$

where α_l, β_l, α_l' and β_l' are constants, independent of r, $j_l(kr)$ and $n_l(kr)$ are spherical Bessel functions of the first and second kinds, respectively [4]. The last expression is the asymptotic form of $R_l(r)$ in the asymptotic limit where $r\to\infty$. The asymptotic form of the wave function is then given by

$$\psi(r,\theta)=\sum_{l=0}^{\infty}\frac{(2l+1)}{2ikr}\frac{a_l\beta_l'}{i^l\sqrt{\pi(2l+1)}}\left[e^{2i\delta_l(k)}e^{ikr}-(-1)^le^{-ikr}\right]P_l(\cos\theta) \tag{A.25}$$

where we have defined $e^{2i\delta_l(k)}=\frac{\alpha_l'}{\beta_l'}$. Note that we write the factor $\frac{(2l+1)}{2ikr}$ out explicitly because it allows us to compare directly with the partial wave expansion of Eq. (A.14) that we derived in the previous section, as we can see below.

We now relate Eq. (A.14) with the partial wave expansion in Eq. (A.25). Using Rayleigh's

[4] Note that the constants, α_l, β_l, α_l' and β_l', are independent of r *only*, which means that they can depend on the other quantities like k and l. This is subtle but important, as we will define the phase shift, which also has k and l dependence, based on the latter two constants.

formula

$$e^{ikz} = \sum_{l=0}^{\infty}(2l+1)i^l j_l(kr)P_l(\cos\theta) \xrightarrow{r\to\infty} \sum_{l=0}^{\infty}\frac{(2l+1)}{2ikr}\left[e^{ikr} - (-1)^l e^{-ikr}\right]P_l(\cos\theta), \tag{A.26}$$

Eq. (A.14) becomes

$$\psi_{\mathbf{k}}^{3\mathrm{D}}(\mathbf{r}) \xrightarrow{r\to\infty} \frac{1}{\sqrt{\Omega}}\left[\sum_{l=0}^{\infty}\frac{(2l+1)}{2ikr}\left[e^{ikr} - (-1)^l e^{-ikr}\right]P_l(\cos\theta) + f_{\mathbf{k}'\mathbf{k}}^{3\mathrm{D}}\frac{e^{ikr}}{r}\right]. \tag{A.27}$$

Comparing with Eq. (A.25) and equating coefficients for each value of l, we finally obtain the partial wave expansion of the scattering amplitude $f_{\mathbf{k}'\mathbf{k}}^{3\mathrm{D}}$ in 3D

$$f_{\mathbf{k}'\mathbf{k}}^{3\mathrm{D}} = \sum_{l=0}^{\infty}\frac{(2l+1)}{2ik}(e^{2i\delta_l(k)} - 1)P_l(\cos\theta) = \sum_{l=0}^{\infty}(2l+1)f_l(k)P_l(\cos\theta) \tag{A.28}$$

with

$$f_l(k) = \frac{(e^{2i\delta_l(k)} - 1)}{2ik} = \frac{e^{i\delta_l(k)}\sin\left[\delta_l(k)\right]}{k} = \frac{1}{k\cot\left[\delta_l(k)\right] - ik}. \tag{A.29}$$

The matching of coefficients shows that the product $a_l\beta_l'$ is required to be $a_l\beta_l' = i^l\sqrt{\frac{\pi(2l+1)}{\Omega}}$.

Low-energy limit and the scattering length

As we are concerned about the low-energy limit ($k \to 0$) of the scattering, the only relevant term in the partial wave expansion is the $l = 0$ term, also known as the "s-wave" term. This is because higher angular momentum terms $l > 0$ are suppressed due to the centrifugal barrier $\frac{l(l+1)}{r^2}$ in the Hamiltonian. In the distant region where $V(r) \approx 0$, keeping only the outgoing wave with $l = 0$, the wave function and scattering amplitude become

$$\psi(r) = \frac{e^{i\delta_0(k)}}{kr\sqrt{\Omega}}\sin\left[kr + \delta_0(k)\right] \qquad f_{\mathbf{k}'\mathbf{k}}^{3\mathrm{D}} = f_0(k) = \frac{e^{i\delta_0(k)}\sin\left[\delta_0(k)\right]}{k}. \tag{A.30}$$

Note that the two quantities no longer depend on θ, eliminating any dependence on the directions of $\mathbf{k}$ and $\mathbf{k}'$ and ensuring that the wave function and scattering amplitude depend on $k = |\mathbf{k}| = |\mathbf{k}'|$ only.

On the other hand, if we set $k = 0$, $l = 0$ and $V(r) = 0$ in Eq. (A.21), and assume a trial solution of the form $\psi(r) = u_0(r)/r$, we get a very simple differential equation for $u_0(r)$

$$\frac{d^2u_0(r)}{dr^2} = 0 \tag{A.31}$$

which has a general solution of $u_0(r) = c_1(r - c_2)$, where c_1, c_2 are constants. Expanding

Eq. (A.30) and matching its $k \to 0$ asymptotic form with this solution, we get

$$c_2 = -\frac{1}{k \cot\left[\delta_0(k)\right]}\bigg|_{k\to 0}. \tag{A.32}$$

The quantity c_2 here is also known as the (s-wave) *scattering length*, denoted as $a_s^{3\mathrm{D}}$. Hence, we arrive at the definition of scattering length in 3D

$$\lim_{k\to 0} k \cot\left[\delta_0(k)\right] = -\frac{1}{a_s^{3\mathrm{D}}}. \tag{A.33}$$

Under this limit, the scattering amplitude is related to the scattering length by

$$\lim_{k\to 0} f_0(k) = -a_s^{3\mathrm{D}}. \tag{A.34}$$

For finite range potentials with effective range r_e, one could also include the next order term [51, 136, 167]

$$k \cot\left[\delta_0(k)\right]\bigg|_{k\to 0} = -\frac{1}{a_s^{3\mathrm{D}}} + \frac{r_e k^2}{2}. \tag{A.35}$$

Let us take a break here and try to understand the physical significance of the scattering length. Mathematically, it is simply the r-intercept of the asymptotic ($r \to \infty$) wave function. With further calculations [136], it is possible to show that for *repulsive* potentials, $a_s^{3\mathrm{D}} > 0$. For *attractive* potentials, $a_s^{3\mathrm{D}} < 0$ if there are no bound states, and $a_s^{3\mathrm{D}} > 0$ if the potential is sufficiently strong to host bound states. Physically, the scattering length captures the microscopic details of the potential and integrates them into a single quantity. This is valid in dilute Fermi gases, since the scattering occurs at low energy and the particles are sufficiently far apart that they are rarely affected by the microscopic details of the interaction.

A.2.2 Partial wave expansions in 2D

Having discussed the partial wave expansion in 3D, we now turn our attention to 2D. The time-independent Schrödinger equation for 2D in polar coordinates with an isotropic potential $V(r)$ can be written as

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\psi(\mathbf{r})\right) + \frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}\psi(\mathbf{r}) + \left[k^2 - \frac{2m}{\hbar^2}V(r)\right]\psi(\mathbf{r}) = 0 \tag{A.36}$$

which is separable in r and θ. This implies that the solution can be written as [2, 51]

$$\psi(\mathbf{r}) = \psi(r,\theta) = \sum_{l=0}^{\infty} a_l R_l(r) T_l(\theta) \tag{A.37}$$

where the functions $R_l(r)$ and $T_l(\theta)$ satisfy the following differential equations respectively

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{d}{dr}R_l(r)\right)+\left[k^2-\frac{l^2}{r^2}-\frac{2m}{\hbar^2}V(r)\right]R_l(r)=0, \qquad \frac{d^2}{d\theta^2}T_l(\theta)+l^2T_l(\theta)=0. \tag{A.38}$$

In addition, we choose the incident wave vector $\mathbf{k}$ to align with the x-axis, *i.e.* $\mathbf{k}=k\hat{\mathbf{x}}$. Under this convention, the function $T_l(\theta)$ must be even due to symmetry. Together with the normalisation condition $\int_0^{2\pi}|T_l(\theta)|^2d\theta=1$, we get

$$T_l(\theta)=\frac{1}{\sqrt{\pi}}\cos(l\theta). \tag{A.39}$$

On the other hand, in the region where $V(r)=0$, the general solution for $R_l(r)$ is [83, 133, 160]

$$R_l(r)=\alpha_lJ_l(kr)+\beta_lN_l(kr)\xrightarrow{r\to\infty}\frac{\alpha_l'(-1)^li^l}{\sqrt{2\pi kr}}\left[e^{ikr-i\pi/4}e^{2i\delta_l(k)}+(-1)^le^{-ikr+i\pi/4}\right] \tag{A.40}$$

where α_l, β_l, α_l' and δ_l are independent of r, and $J_l(x)$ and $N_l(x)$ are Bessel functions of the first and second kinds, respectively. From the expansion of the incident wave, we have

$$e^{ikx}=\sum_{l=0}^{\infty}c_li^l\cos(l\theta)J_l(kr)\xrightarrow{r\to\infty}\sum_{l=0}^{\infty}\frac{c_l}{\sqrt{2\pi kr}}\cos(l\theta)\left[e^{ikr-i\pi/4}+(-1)^le^{-ikr+i\pi/4}\right] \tag{A.41}$$

with $c_0=1$ and $c_l=2$ for $l>0$. Carrying out a similar procedure as for the 3D case in the previous section, we obtain the following expression for the scattering amplitude

$$f^{2\mathrm{D}}_{\mathbf{k}'\mathbf{k}}=\sum_{l=0}^{\infty}c_l\sqrt{\frac{2i}{\pi k}}e^{i\delta_l(k)}\sin[\delta_l(k)]\cos(l\theta)=\sum_{l=0}^{\infty}c_lf_l(k)\cos(l\theta) \tag{A.42}$$

with

$$f_l(k)=\sqrt{\frac{2i}{\pi k}}e^{i\delta_l(k)}\sin[\delta_l(k)]=\sqrt{\frac{2i}{\pi k}}\frac{1}{\cot[\delta_l(k)]-i} \tag{A.43}$$

and $a_l\alpha_l'=i^lc_l\sqrt{\frac{\pi}{\Omega}}$.

Low-energy limit and the scattering length

Again, we now discuss the behaviour of the wave function at low energy. With $l=0$ and in the region where $V(r)=0$, the radial wave function and the scattering amplitude in the limit of $k\to 0$ become

$$\begin{aligned}R_0(r)&=-\frac{2}{\pi}e^{i\delta_0(k)}\sin[\delta_0(k)]\left[\ln(kr)+\gamma-\ln 2-\frac{\pi}{2}\cot[\delta_0(k)]\right]\\ f^{2\mathrm{D}}_{\mathbf{k}'\mathbf{k}}&=f_0(k)=\sqrt{\frac{2i}{\pi k}}e^{i\delta_0(k)}\sin[\delta_0(k)],\end{aligned} \tag{A.44}$$

where $\gamma \approx 0.577$ is the Euler-Mascheroni constant.

On the other hand, setting $l = 0$, $k = 0$ and $V(r) = 0$ in Eq. (A.38) gives the simple differential equation

$$\frac{d}{dr}\left(r\frac{dR_0(r)}{dr}\right) = 0 \tag{A.45}$$

which has a general solution of $R_0(r) = c_1 \ln(r/c_2)$, where c_1, c_2 are constants of r. Defining the 2D scattering length $a_s^{2\mathrm{D}}$ to be the r-intercept of the wave function, we have $c_2 = a_s^{2\mathrm{D}}$. Comparing the two expressions of $R_0(r)$, we get

$$\cot[\delta_0(k)]\Big|_{k\to 0} = \frac{2}{\pi}\left[\gamma + \ln\left(\frac{ka_s^{2\mathrm{D}}}{2}\right)\right]_{k\to 0} = \frac{2}{\pi}\ln\left(\frac{ka_s^{2\mathrm{D}}e^{\gamma}}{2}\right)\Big|_{k\to 0}, \tag{A.46}$$

and which can also be regarded as the definition of the scattering length in 2D. Note that in some literature, the constant $e^{\gamma}/2$ is absorbed into the 2D scattering length, such that $\frac{e^{\gamma}}{2}a_s^{2\mathrm{D}} \approx 0.89a_s^{2\mathrm{D}} \to a_s^{2\mathrm{D}}$. The scattering amplitude in the limit as $k \to 0$ is given by

$$f_{\mathbf{k}'\mathbf{k}}^{2\mathrm{D}}\Big|_{k\to 0} = f_0(k)\Big|_{k\to 0} = \sqrt{\frac{i\pi}{2k}}\frac{1}{\ln\left(\frac{ka_s^{2\mathrm{D}}}{2}\right)\big|_{k\to 0} + \gamma - \frac{i\pi}{2}}. \tag{A.47}$$

For a potential with a finite range r_e, we have [83, 160, 164]

$$\cot[\delta_0(k)]\Big|_{k\to 0} \approx \frac{2}{\pi}\left[\gamma + \ln\left(\frac{ka_s^{2\mathrm{D}}}{2}\right)\Big|_{k\to 0}\right] + \frac{r_e^2k^2}{2\pi}. \tag{A.48}$$

Compared to the 3D case that has a finite limit as $k \to 0$, the 2D scattering amplitude is divergent in this limit. In addition, it is well known that in the attractive 2D scattering problem, there always exists a bound state regardless of how weak the attractive potential is [51]. The direct implication is that the 2D scattering length $a_s^{2\mathrm{D}}$ for an attractive potential is *always non-negative*. Hence, in the main text, we characterise the strength of the interaction in 2D with $\ln\big(k_F a_s^{2\mathrm{D}}\big)$, which has a range of $(-\infty, \infty)$.

A.3 Regularising the Interaction

In this section, we combine the results from the previous two sections and derive the relation between the strength of the contact interaction and the scattering length in both 2D and 3D in the low energy limit. First, we focus on the 3D case.

As we have discussed in the previous section, in the limit as $k \to 0$, higher angular momenta scatterings are suppressed due to the centrifugal barrier. Hence, only the s-wave $l = 0$ term survives in the partial wave expansion of the scattering amplitude, implying that $f_{\mathbf{k}'\mathbf{k}}^{3\mathrm{D}} = f_0(k)$. This also implies that the scattering becomes isotropic as the scattering amplitude is now independent of the direction of the scattered wave $\mathbf{k}'$. As such, together with the use of the contact

(delta) interaction $V_{\mathbf{k}'\mathbf{k}} = -\frac{U_0}{\Omega}$, we can replace $f^{3\mathrm{D}}_{\mathbf{k}'\mathbf{k}}$ by $f_0(k)$ in Eq. (A.19) and divide the whole equation by $U_0 f_0(k)$ to get

$$\frac{1}{U_0} = \frac{m}{4\pi\hbar^2}\frac{1}{f_0(k)} - \frac{m}{\hbar^2\Omega}\sum_{\mathbf{k}''}\frac{1}{k^2 - k''^2 + i0}. \tag{A.49}$$

Note that we have replaced the mass m by the reduced mass $m/2$ to ensure that Eq. (A.49) describes the relative motion of two interacting particles of mass m. In the $k \to 0$ limit, $f_0(k) \to -a_s^{3\mathrm{D}}$ and we arrive at the following relation

$$\frac{m}{4\pi\hbar^2 a_s^{3\mathrm{D}}} = -\frac{1}{U_0} + \frac{1}{\Omega}\sum_{\mathbf{k}''}\frac{1}{2\epsilon_{\mathbf{k}''}} \tag{A.50}$$

which is used in the main text to regularise the BCS gap equation.

Now we carry out the same procedure for 2D, setting $f^{2\mathrm{D}}_{\mathbf{k}'\mathbf{k}} = f_0(k)$, $V_{\mathbf{k}'\mathbf{k}} = -\frac{U_0}{\Omega}$ and $m \to m/2$ in Eq. (A.18) to get

$$\frac{1}{U_0} = \frac{m}{2\hbar^2}\sqrt{\frac{i}{2\pi k}}\frac{1}{f_0(k)} - \frac{m}{\hbar^2\Omega}\sum_{\mathbf{k}''}\frac{1}{k^2 - k''^2 + i0}. \tag{A.51}$$

Note that the prefactor of the $\frac{1}{f_0(k)}$ term differs from the prefactor in the 3D case. In the $k \to 0$ limit, $f_0(k) \to \sqrt{\frac{i\pi}{2k}}\left[\ln\left(\frac{ka_s^{2\mathrm{D}}}{2}\right)\big|_{k\to 0} + \gamma - \frac{i\pi}{2}\right]^{-1}$ and we arrive at the following relation for 2D

$$\frac{m}{2\pi\hbar^2}\left[\ln\left(\frac{ka_s^{2\mathrm{D}}}{2}\right)\bigg|_{k\to 0} + \gamma - \frac{i\pi}{2}\right] = \frac{1}{U_0} - \frac{1}{\Omega}\sum_{\mathbf{k}''}\frac{1}{2\epsilon_{\mathbf{k}''}}. \tag{A.52}$$

At first glance, this equation is problematic in many sense. First, compared to the 3D equation, it still has a term that contains the momentum k, even though we have taken the limit $k \to 0$. This is because the equation is divergent in this limit as $\ln(k)\big|_{k\to 0} \to -\infty$. Therefore, we have to keep that term even though one should keep in mind that the momentum k should tends to zero. Secondly, there is another source of divergence on the right-hand side, which is the summation over $\mathbf{k}''$. This sum contains both infra-red and ultraviolet divergences in 2D, although one could argue that the infra-red divergence should be cancelled by the $\ln k$ on the right. The last problem came from the imaginary term $\frac{i\pi}{2}$ on the left-hand side, when there is no obvious imaginary term on the right-hand side. Nonetheless, if one combines it with the BCS gap equation in Eq. (2.18), all of the three problems go away. A detailed derivation of the regularised BCS gap equation in 2D can be found in the background chapter. We will stop our discussion of scattering theory at this point, as we have derived everything we need. The following references provide fuller and more complete analyses of scattering theory in 2D and 3D [1, 2, 51, 83, 97, 136, 160, 164, 167].

Appendix B

Expectation Values in QMC

In this Appendix we provide details on the QMC estimators of the relevant observables.

B.1 Density Matrices

The one- and two-body density matrices, $\rho_\alpha^{(1)}(\mathbf{r}_1', \mathbf{r}_1)$ and $\rho_{\alpha\beta}^{(2)}(\mathbf{r}_1', \mathbf{r}_2'; \mathbf{r}_1, \mathbf{r}_2)$ are given by [107]

$$\rho_\alpha^{(1)}(\mathbf{r}_1', \mathbf{r}_1) = N^\alpha \frac{\int |\Psi(\mathbf{R})|^2 \left[\frac{\Psi(\mathbf{r}_1', \dots)}{\Psi(\mathbf{r}_1, \dots)}\right] d\mathbf{r}_2 \cdots d\mathbf{r}_N}{\int |\Psi(\mathbf{R})|^2 d\mathbf{R}} \tag{B.1}$$

$$\rho_{\alpha\beta}^{(2)}(\mathbf{r}_1', \mathbf{r}_2'; \mathbf{r}_1, \mathbf{r}_2)) = N^\alpha N^\beta \frac{\int |\Psi(\mathbf{R})|^2 \left[\frac{\Psi(\mathbf{r}_1', \mathbf{r}_2', \dots)}{\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots)}\right] d\mathbf{r}_3 \cdots d\mathbf{r}_N}{\int |\Psi(\mathbf{R})|^2 d\mathbf{R}} \tag{B.2}$$

where $\mathbf{R} \equiv \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N\}$, $\alpha, \beta \in \{\uparrow, \downarrow\}$, and the normalisations are given in Eq. (2.79). The translationally averaged of the two density matrices are given in the main text in Eq. (2.80) and Eq. (2.81). They can be calculated using Monte Carlo integration

$$\frac{\int |\Psi(\mathbf{R})|^2 f(\mathbf{R}) d\mathbf{R}}{\int |\Psi(\mathbf{R})|^2 d\mathbf{R}} \approx \frac{1}{M} \sum_{i=1}^{M} f(\mathbf{R}_i) \tag{B.3}$$

where $f(\mathbf{R})$ is a general function of the particle coordinates $\mathbf{R}$, and $\{\mathbf{R}_i | i = 1, \dots, M\}$ is a set of Markov chain Monte Carlo (MCMC) samples, usually obtained from Metropolis-Hasting sampling of the probability distribution $|\Psi(\mathbf{R})|^2$. Most of the density matrix results in Chapter 6 and Chapter 7 are obtained using this method. For example, if one wish to calculate the translationally averaged one-body density matrix, $\bar{\rho}_\alpha^{(1)}(\mathbf{r})$, one would need to accumulate the quantity $\frac{\Psi(\mathbf{r}_1+\mathbf{r}, \dots)}{\Psi(\mathbf{r}_1, \dots)}$ using Eq. (B.3). In addition, to minimise any potential bias introduced when choosing a specific particle (particle-1 in this case) to calculate $\frac{\Psi(\mathbf{r}_1+\mathbf{r}, \dots)}{\Psi(\mathbf{r}_1, \dots)}$, we also average this quantity

over a batch of randomly chosen particles. In this case, the quantity $\frac{\Psi(\mathbf{r}_1+\mathbf{r},\dots)}{\Psi(\mathbf{r}_1,\dots)}$ becomes

$$\frac{\Psi(\mathbf{r}_1+\mathbf{r},\dots)}{\Psi(\mathbf{r}_1,\dots)} \to \frac{1}{M_s}\sum_{i_k\in\mathcal{S}_1}\frac{\Psi(\dots,\mathbf{r}_{i_k}+\mathbf{r},\dots)}{\Psi(\dots,\mathbf{r}_{i_k},\dots)} \tag{B.4}$$

where $\mathcal{S}_1 = \{i_k|k=1,\dots,M_s\}$ is a set of randomly chosen particle indices. The number M_s is the number of random samples in the set $\mathcal{S}_1$, which we choose to be a fraction of the number of particles of species-α, N^α. This can also be generalised to calculate the translationally averaged two-body density matrix in Eq. (2.81) by generating a set of random pairs of two indices, *i.e.*, $\mathcal{S}_2 = \{(i_k,j_k)|k=1,\dots,M_s\}$, and

$$\frac{\Psi(\mathbf{r}_1+\mathbf{r},\mathbf{r}_2+\mathbf{r},\dots)}{\Psi(\mathbf{r}_1,\mathbf{r}_2,\dots)} \to \frac{1}{M_s}\sum_{(i_k,j_k)\in\mathcal{S}_2}\frac{\Psi(\dots,\mathbf{r}_{i_k}+\mathbf{r},\dots,\mathbf{r}_{j_k}+\mathbf{r},\dots)}{\Psi(\dots,\mathbf{r}_{i_k},\dots,\mathbf{r}_{j_k},\dots)}. \tag{B.5}$$

Lastly, the pair-momentum density results in the main text are calculated by Fourier transforming the QMC accumulated translationally averaged two-body density matrix, as given in Eq. (2.86).

The translationally and rotationally averaged density matrices are given by Eq. (2.82) and Eq. (2.83), which have one extra integral $\int\delta(|\mathbf{r}'|-r)[\cdots]d\mathbf{r}'$ compared to Eq. (2.80) and Eq. (2.81). Hence, to calculate these quantities, one would need to average them over a sphere (or circle in 2D) of radius r, on top of the MCMC average and the particle index average. In this case, the average is given by

$$\int\delta(|\mathbf{r}'|-r)f(\mathbf{r}')d\mathbf{r}' \approx \frac{1}{M_b}\sum_{i=1}^{M_b}f(\mathbf{r}'_i) \tag{B.6}$$

where $\{\mathbf{r}'_i|i=1,\dots,M_b;|\mathbf{r}'_i|=r\}$ is a set of randomly generated points on a sphere (or circle in 2D) of radius r. This is also how the quantity $\frac{\Omega^2}{N_\uparrow}\rho^{(2)TR}_{\uparrow\downarrow}(r)$ is calculated in Fig. (5.4) in Chapter 5.

B.1.1 Improved Estimators for the Condensate Fraction

In a superfluid, while the definition in Eq. (2.84) is valid in the thermodynamic limit, where the one-body contributions tend to zero at far separation $r\to\infty$, it might not be the best estimator in real simulations since the one-body effects might not vanish at the edges of the simulation cell [4, 107]. As such, one often needs to subtract the one-body contributions from the TBDM, namely

$$\bar{\rho}^{TR(2)}_{\uparrow\downarrow}(r) \to \left[\rho^{(2)}_{\uparrow\downarrow}(\mathbf{r}_1+\mathbf{r}',\mathbf{r}_2+\mathbf{r}';\mathbf{r}_1,\mathbf{r}_2) - \rho^{(1)}_{\uparrow}(\mathbf{r}_1+\mathbf{r}';\mathbf{r}_1)\rho^{(1)}_{\downarrow}(\mathbf{r}_2+\mathbf{r}';\mathbf{r}_2)\right]^{TR} \tag{B.7}$$

where $[\cdots]^{TR}$ denotes the translational and rotational average of the quantity in the squared brackets. If we approximate

$$\left[\rho^{(1)}_{\uparrow}(\mathbf{r}_1+\mathbf{r}';\mathbf{r}_1)\rho^{(1)}_{\downarrow}(\mathbf{r}_2+\mathbf{r}';\mathbf{r}_2)\right]^{TR} \approx \bar{\rho}^{TR(1)}_{\uparrow}(r)\bar{\rho}^{TR(1)}_{\downarrow}(r), \tag{B.8}$$

we obtain the following improved estimator for the condensate fraction

$$c_{\uparrow\downarrow} = \frac{\Omega^2}{N_{\text{pair}}} \lim_{r\to\infty} \left[\bar{\rho}_{\uparrow\downarrow}^{TR(2)}(r) - \bar{\rho}_{\uparrow}^{TR(1)}(r)\bar{\rho}_{\downarrow}^{TR(1)}(r) \right]. \tag{B.9}$$

Another possible estimator, proposed in the CASINO manual [107], is to redefine the TBDM such that it includes the one-body terms

$$\begin{aligned} &\tilde{\rho}_{\uparrow\downarrow}^{(2)}(\mathbf{r}_1+\mathbf{r}', \mathbf{r}_2+\mathbf{r}'; \mathbf{r}_1, \mathbf{r}_2) \\ &\quad = N^{\uparrow}N^{\downarrow} \frac{\int |\Psi(\mathbf{R})|^2 \left[\frac{\Psi(\mathbf{r}_1+\mathbf{r}', \mathbf{r}_2+\mathbf{r}', \dots)}{\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots)} - \frac{\Psi(\mathbf{r}_1+\mathbf{r}', \mathbf{r}_2, \dots)}{\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots)} \frac{\Psi(\mathbf{r}_1, \mathbf{r}_2+\mathbf{r}', \dots)}{\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots)} \right] d\mathbf{r}_3 \cdots d\mathbf{r}_N}{\int |\Psi(\mathbf{R})|^2 d\mathbf{R}} \end{aligned} \tag{B.10}$$

and the condensate fraction is estimated as

$$c_{\uparrow\downarrow} = \frac{\Omega^2}{N_{\text{pair}}} \lim_{r\to\infty} \bar{\tilde{\rho}}_{\uparrow\downarrow}^{TR(2)}(r). \tag{B.11}$$

with $\bar{\tilde{\rho}}_{\uparrow\downarrow}^{TR(2)}(r)$ being the translationally and rotationally averaged density matrix in Eq. (B.10). This improved estimator is used throughout this work, unless stated otherwise.

Appendix C

Supplementary Data and Figures

Here we present data and figures that are not shown in the main text.

C.1 Anisotropic Fermi Gas

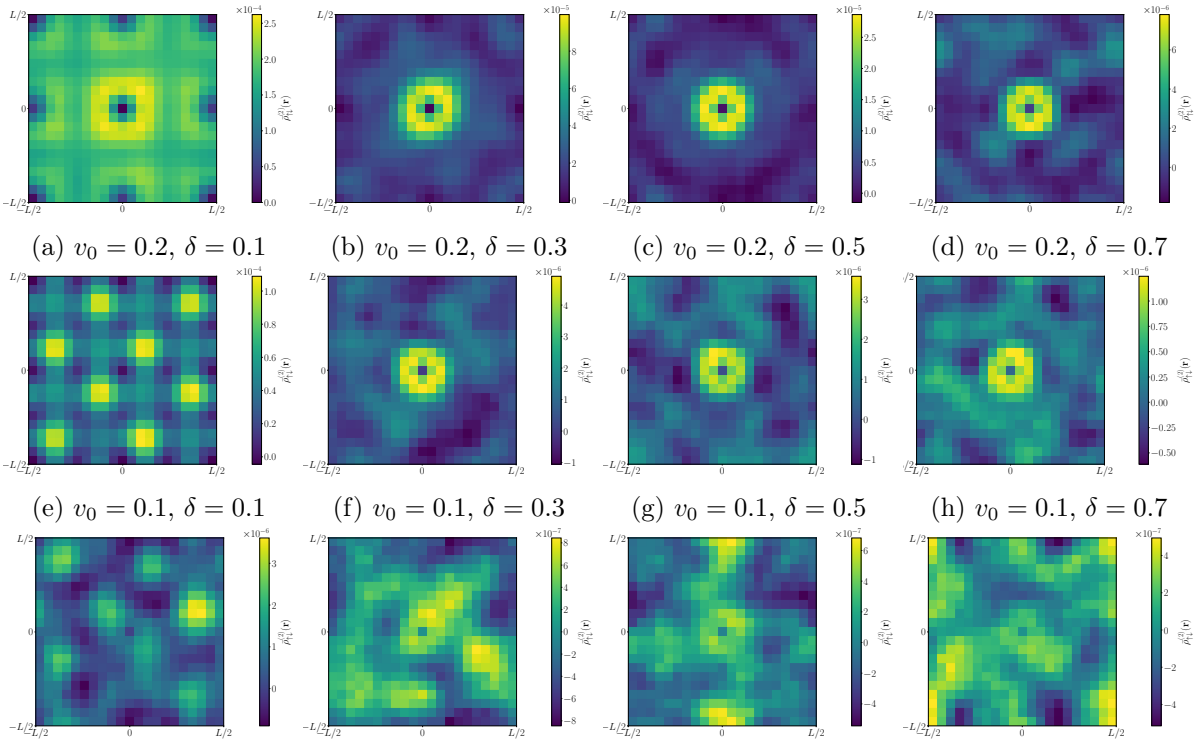
In our work on the anisotropic Fermi gas, we have done a grid scan of the phase diagram spanned by the interaction strength v_0 and the anisotropy δ, and we have presented some of the most important results in the main text. In this section, we present the full set of results from our work.

C.1.1 Two-Body Density Matrix

In this section, we show the translationally averaged real-space two-body density matrix of the anisotropic Fermi gas for $v_0 = 0.05, 0.1, 0.2, 0.3, 0.35, 0.4, 0.5$ and $\delta = 0.1, 0.3, 0.5, 0.7$. Results are split into two figures: $v_0 = 0.05, 0.1, 0.2$ in Fig. (C.1), and $v_0 = 0.3, 0.35, 0.4, 0.5$ in Fig. (C.2). Each row in the figures has a fixed interaction strength, v_0, and each column has a fixed anisotropy δ. Each of the images are obtained from averaging the samples obtained from 1000 inference steps.

C.1.2 Pair-Momentum Density

In this section, we show the pair-momentum density of the anisotropic Fermi gas for $v_0 = 0.05, 0.1, 0.2, 0.3, 0.35, 0.4, 0.5$ and $\delta = 0.1, 0.3, 0.5, 0.7$. The figures are grouped by the interaction strength, with sub-captions stating the value of $\delta \in \{0.1, 0.3, 0.5, 0.7\}$. Similar to the main text in Chapter 7, the three sub-figures on each row in each of Figs. C.3 to C.9 represent the absolute values of the pair-momentum density with masked $\mathbf{k} = 0$ component (left), the unmasked version (middle) and the phase of the pair-momentum density (right), respectively. Each of the images is obtained by Fourier transforming the translationally averaged two-body density matrix in Figs. C.1 and C.2.

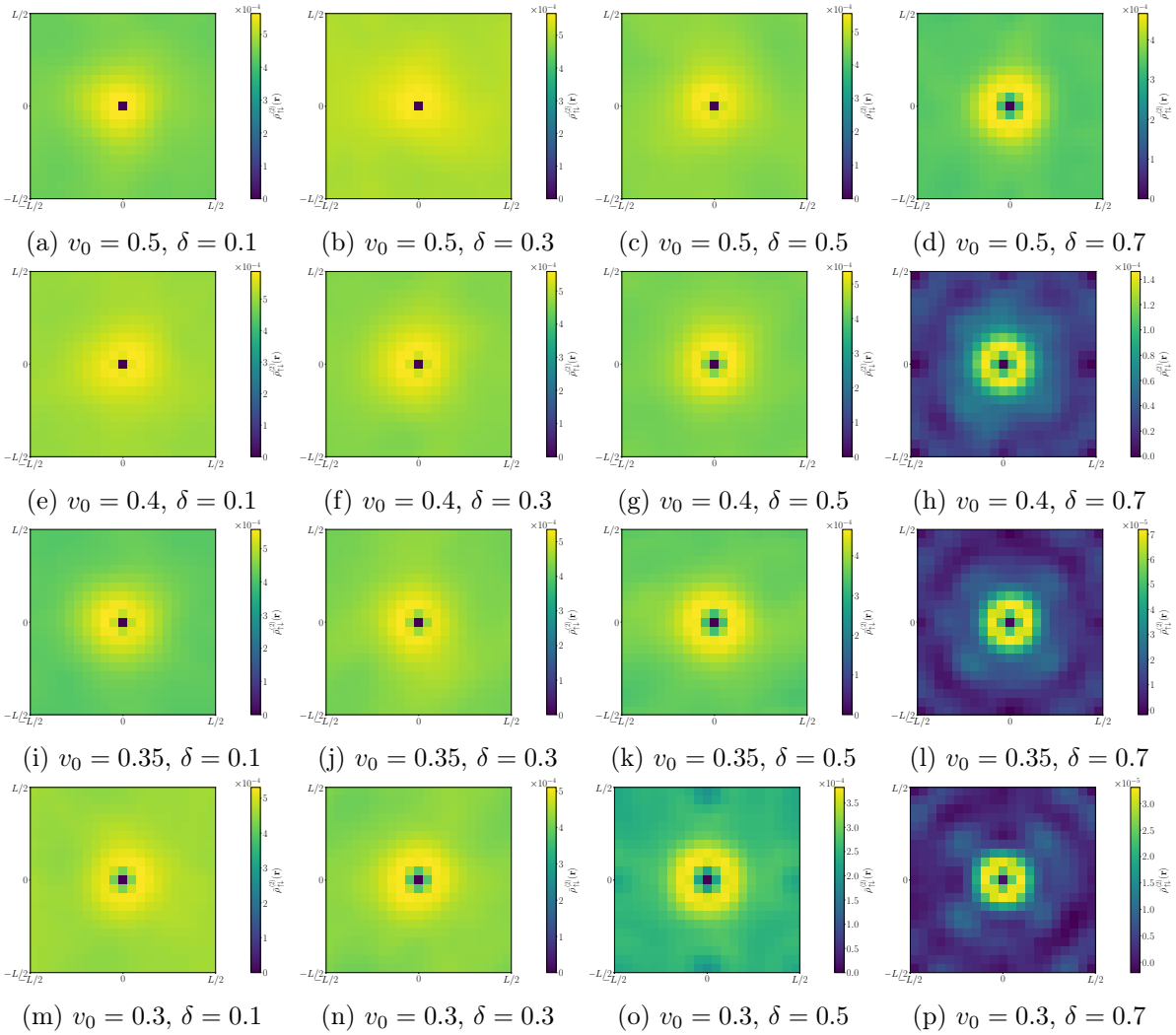


(a) $v_0 = 0.2$, $\delta = 0.1$ (b) $v_0 = 0.2$, $\delta = 0.3$ (c) $v_0 = 0.2$, $\delta = 0.5$ (d) $v_0 = 0.2$, $\delta = 0.7$

(e) $v_0 = 0.1$, $\delta = 0.1$ (f) $v_0 = 0.1$, $\delta = 0.3$ (g) $v_0 = 0.1$, $\delta = 0.5$ (h) $v_0 = 0.1$, $\delta = 0.7$

(i) $v_0 = 0.05$, $\delta = 0.1$ (j) $v_0 = 0.05$, $\delta = 0.3$ (k) $v_0 = 0.05$, $\delta = 0.5$ (l) $v_0 = 0.05$, $\delta = 0.7$

Figure C.1: The real space, translationally averaged two-body density matrix, $\bar{\rho}^{(2)}_{\uparrow\downarrow}(\mathbf{r})$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.05, 0.1, 0.2$ and $\delta = 0.1, 0.3, 0.5, 0.7$.



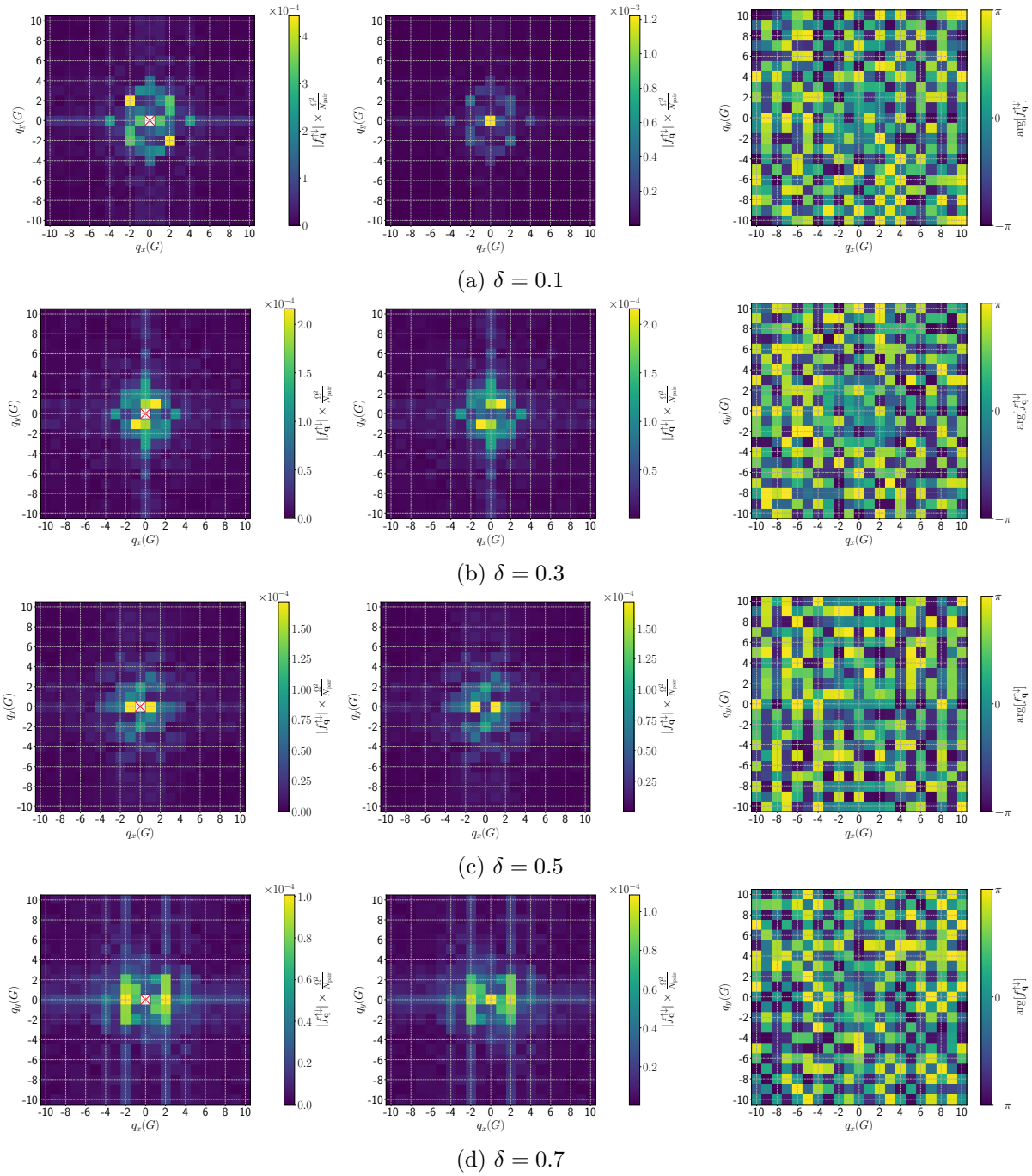
(a) $v_0 = 0.5$, $\delta = 0.1$ (b) $v_0 = 0.5$, $\delta = 0.3$ (c) $v_0 = 0.5$, $\delta = 0.5$ (d) $v_0 = 0.5$, $\delta = 0.7$

(e) $v_0 = 0.4$, $\delta = 0.1$ (f) $v_0 = 0.4$, $\delta = 0.3$ (g) $v_0 = 0.4$, $\delta = 0.5$ (h) $v_0 = 0.4$, $\delta = 0.7$

(i) $v_0 = 0.35$, $\delta = 0.1$ (j) $v_0 = 0.35$, $\delta = 0.3$ (k) $v_0 = 0.35$, $\delta = 0.5$ (l) $v_0 = 0.35$, $\delta = 0.7$

(m) $v_0 = 0.3$, $\delta = 0.1$ (n) $v_0 = 0.3$, $\delta = 0.3$ (o) $v_0 = 0.3$, $\delta = 0.5$ (p) $v_0 = 0.3$, $\delta = 0.7$

Figure C.2: The real space, translationally averaged two-body density matrix, $\bar{\rho}^{(2)}_{\uparrow\downarrow}(\mathbf{r})$, of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.3, 0.35, 0.4, 0.5$ and $\delta = 0.1, 0.3, 0.5, 0.7$.



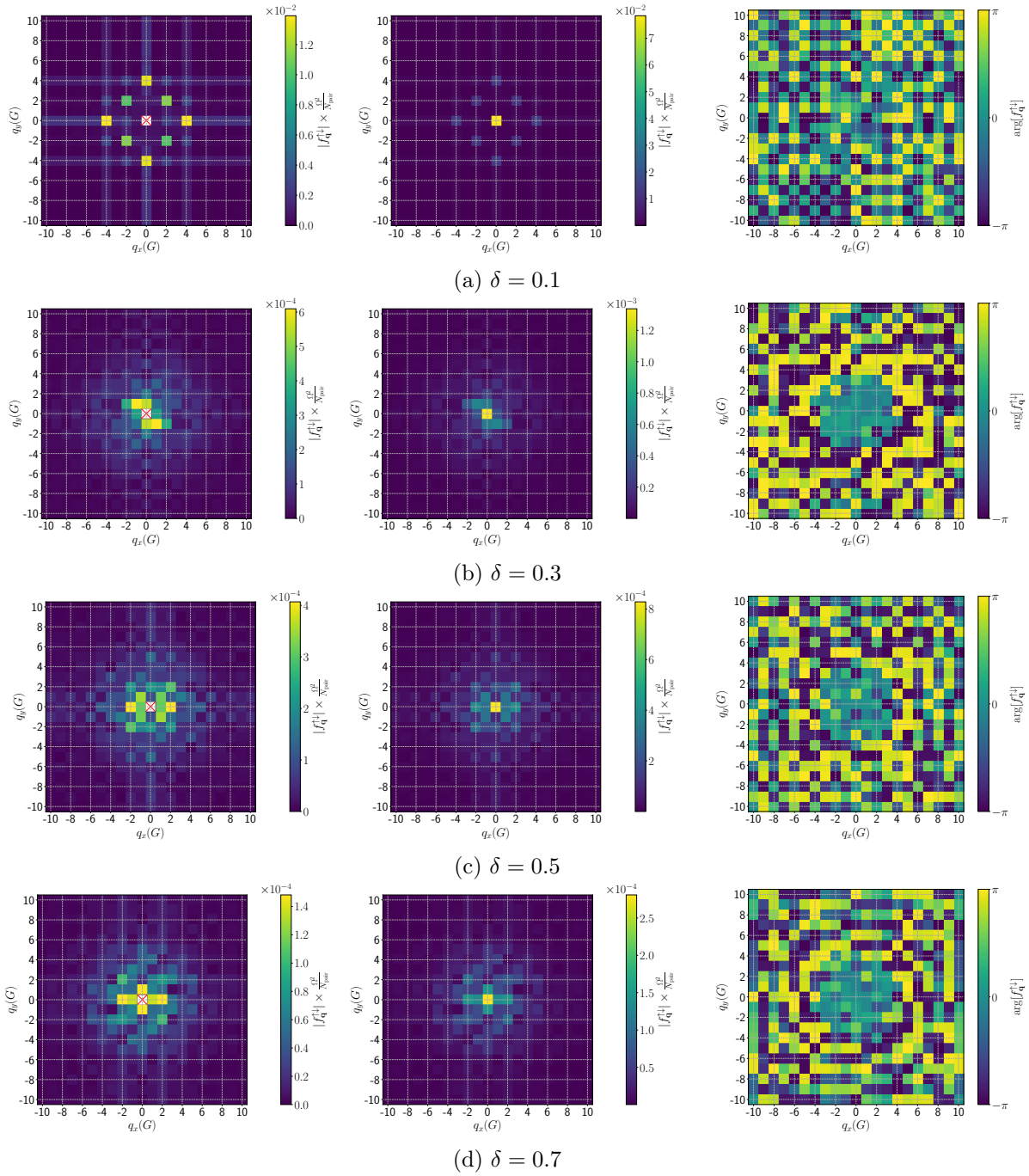
(a) $\delta = 0.1$

(b) $\delta = 0.3$

(c) $\delta = 0.5$

(d) $\delta = 0.7$

Figure C.3: The absolute values and the phases of the pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.05$ and $\delta = 0.1, 0.3, 0.5, 0.7$



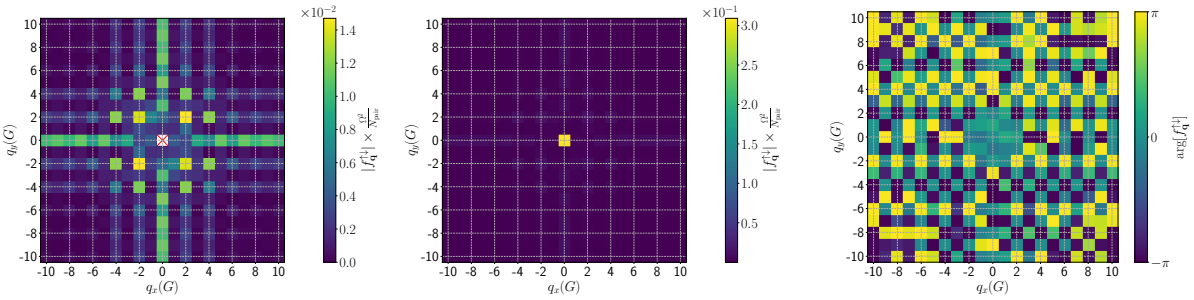
(a) $\delta = 0.1$

(b) $\delta = 0.3$

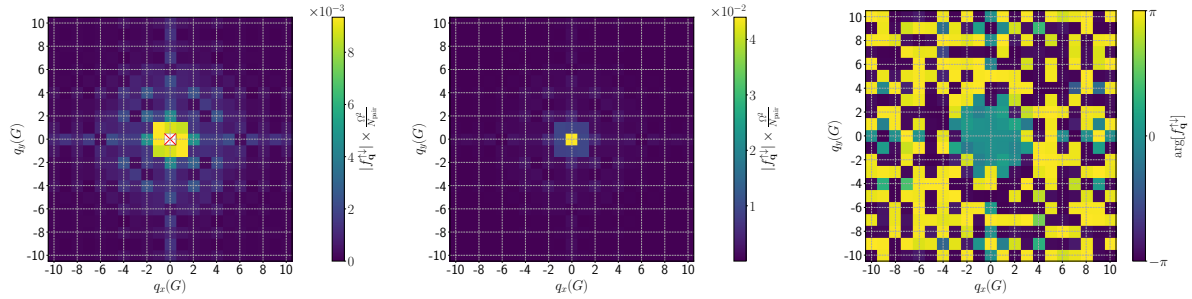
(c) $\delta = 0.5$

(d) $\delta = 0.7$

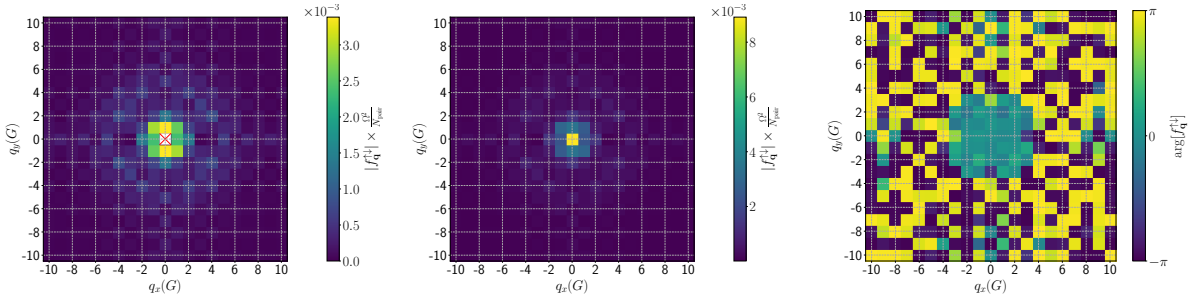
Figure C.4: The absolute values and the phases of the pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.1$ and $\delta = 0.1, 0.3, 0.5, 0.7$



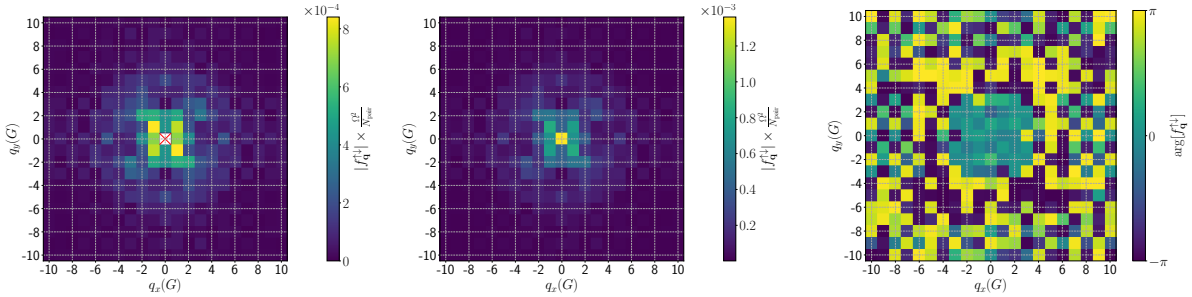
(a) $\delta = 0.1$



(b) $\delta = 0.3$

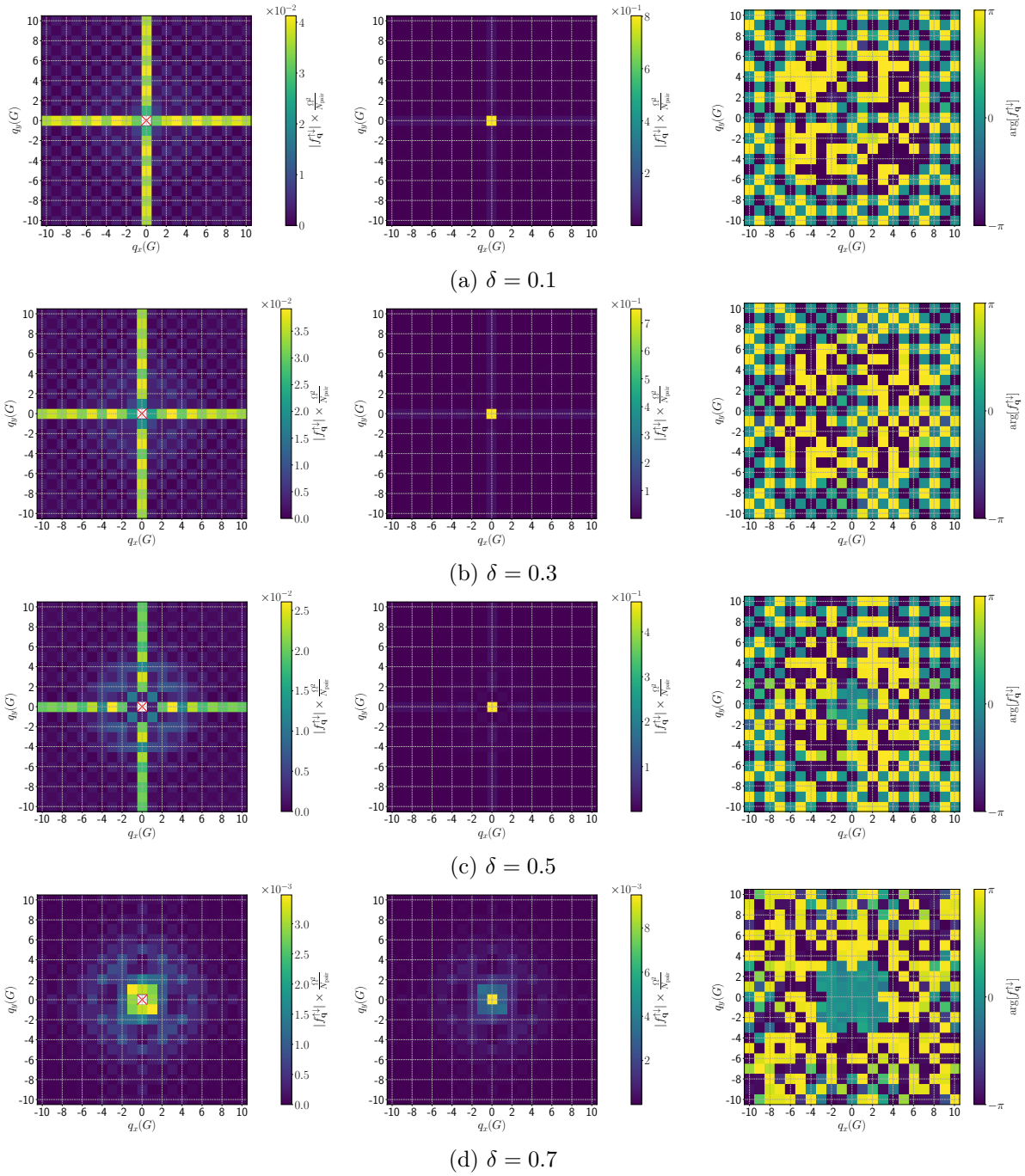


(c) $\delta = 0.5$



(d) $\delta = 0.7$

Figure C.5: The absolute values and the phases of the pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.2$ and $\delta = 0.1$



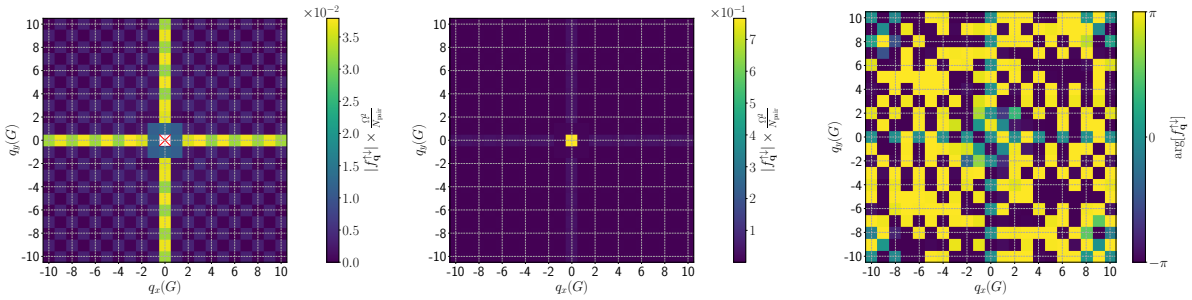
(a) $\delta = 0.1$

(b) $\delta = 0.3$

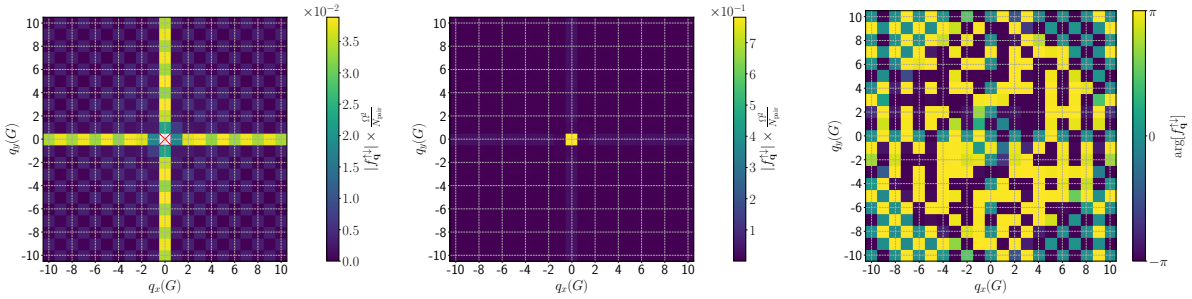
(c) $\delta = 0.5$

(d) $\delta = 0.7$

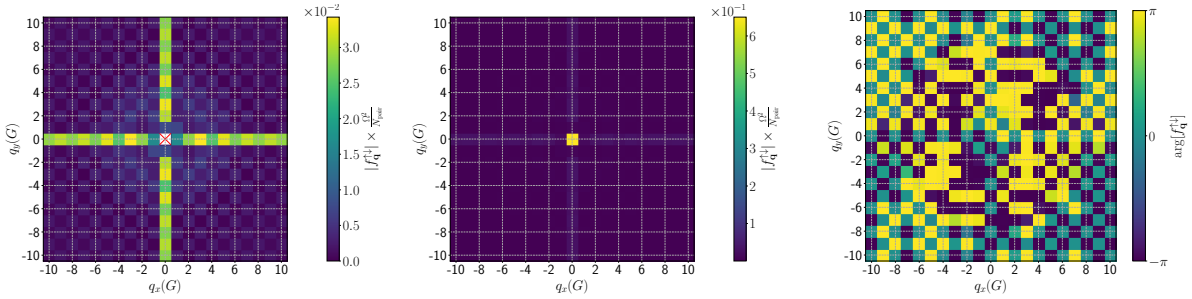
Figure C.6: The absolute values and the phases of the pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.3$ and $\delta = 0.1, 0.3, 0.5, 0.7$



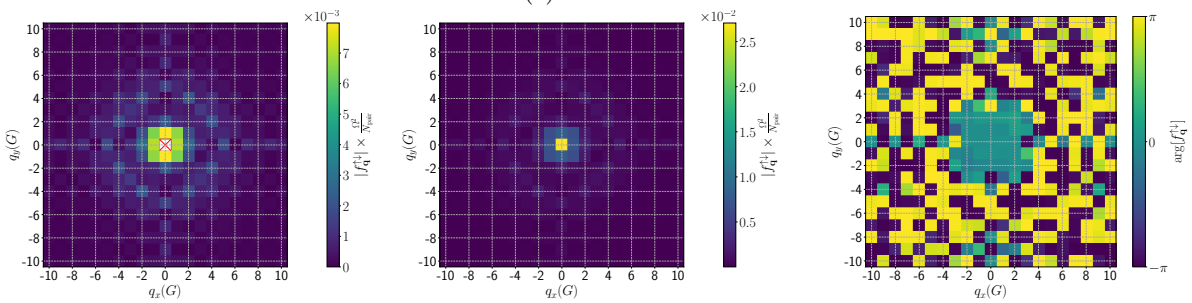
(a) $\delta = 0.1$



(b) $\delta = 0.3$

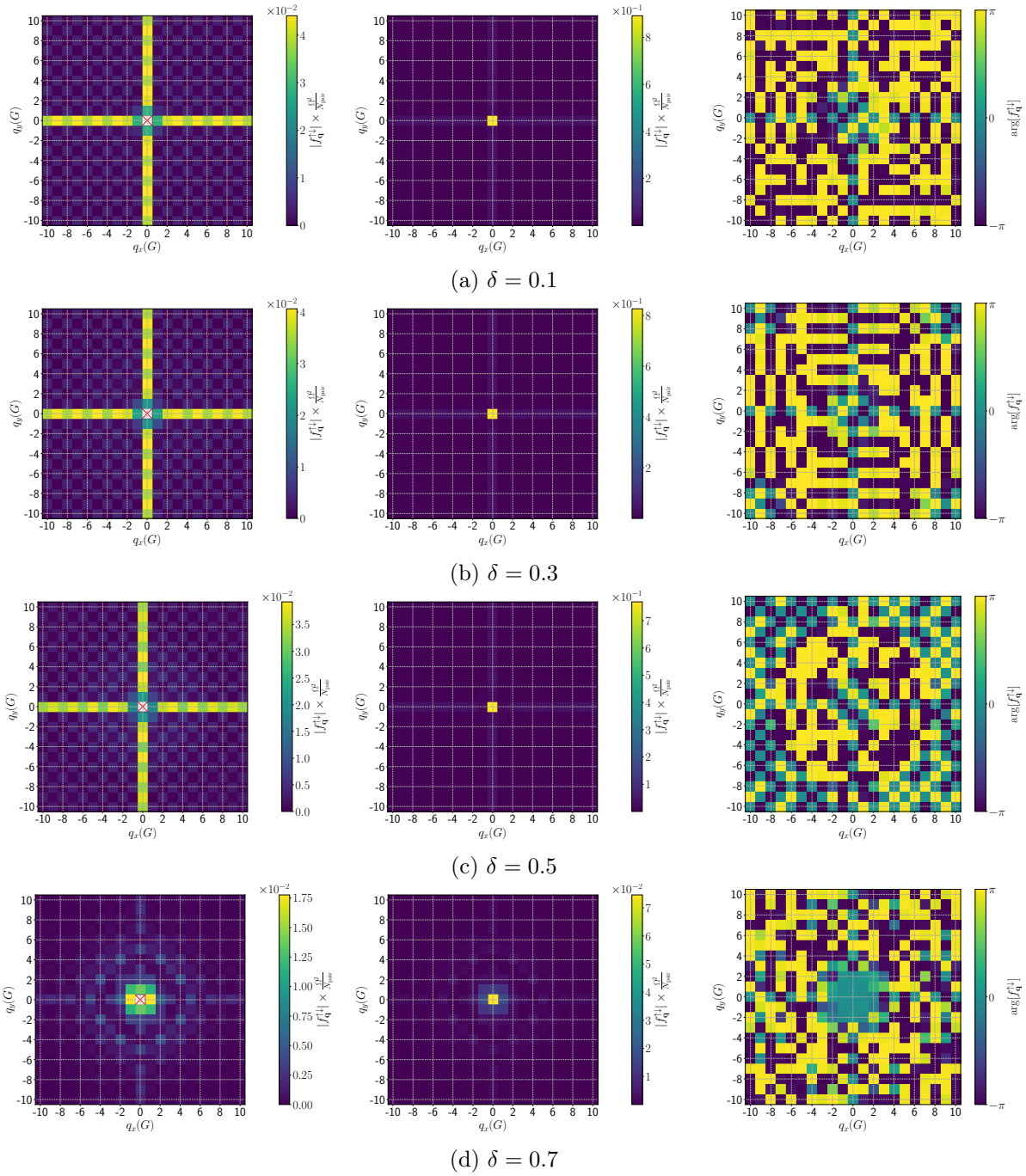


(c) $\delta = 0.5$



(d) $\delta = 0.7$

Figure C.7: The absolute values and the phases of the pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.35$ and $\delta = 0.1, 0.3, 0.5, 0.7$



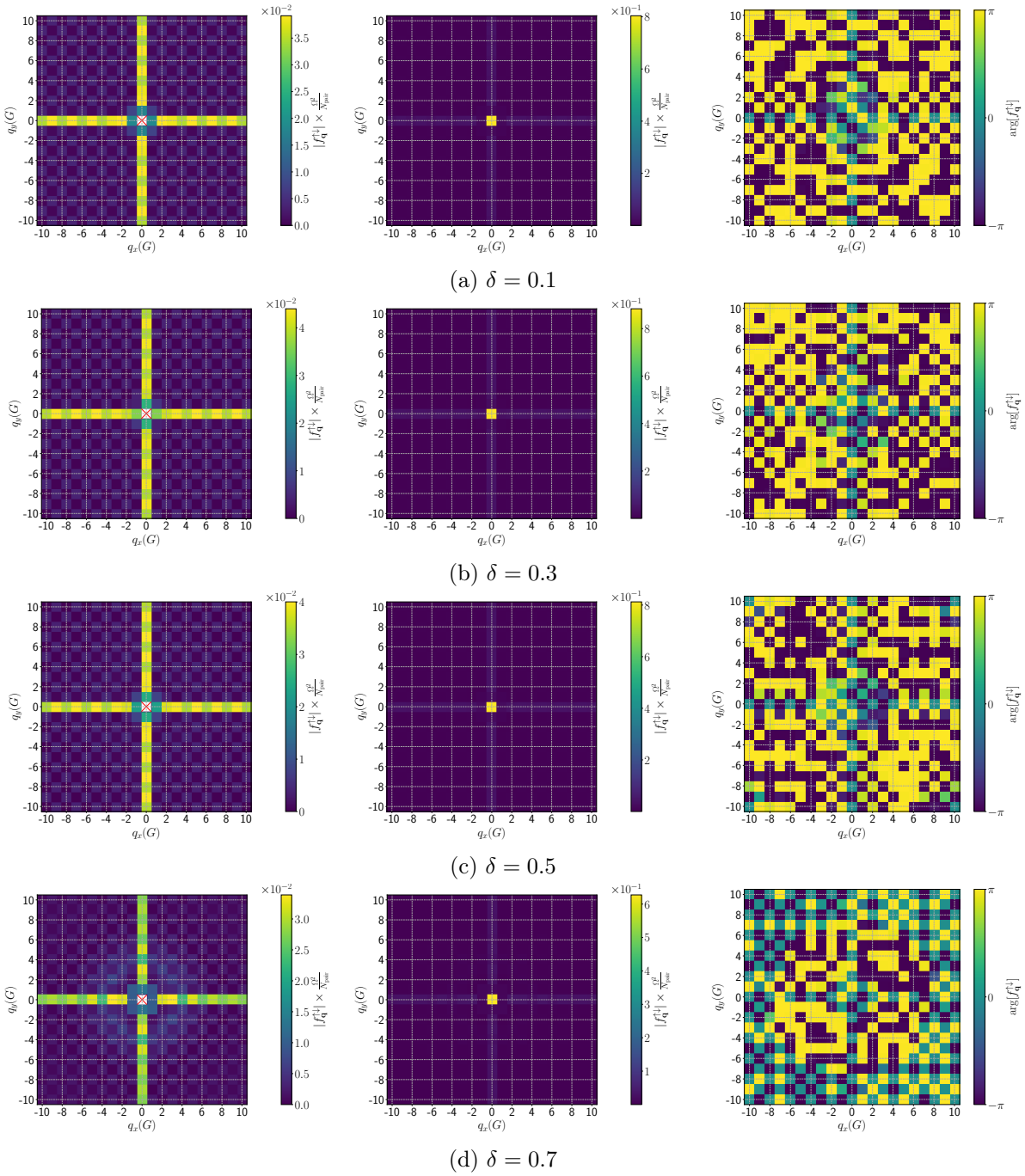
(a) $\delta = 0.1$

(b) $\delta = 0.3$

(c) $\delta = 0.5$

(d) $\delta = 0.7$

Figure C.8: The absolute values and the phases of the pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.4$ and $\delta = 0.1, 0.3, 0.5, 0.7$



(a) $\delta = 0.1$

(b) $\delta = 0.3$

(c) $\delta = 0.5$

(d) $\delta = 0.7$

Figure C.9: The absolute values and the phases of the pair-momentum density $f_{\mathbf{q}}^{\uparrow\downarrow} \times \frac{\Omega^2}{N_{\text{pair}}}$ of the $(N^{\uparrow}, N^{\downarrow}) = (11, 11)$ anisotropic Fermi gas with $v_0 = 0.5$ and $\delta = 0.1, 0.3, 0.5, 0.7$