Interference in Complex Quantum Systems
From localization in high-dimensional lattices to surface spin echo with molecules

by

Joshua Tyler Cantin
B. ASc., University of Waterloo, 2014

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Chemistry)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

September 2019

© Joshua Tyler Cantin, 2019
The following individuals certify that they have read, and recommend to the Faculty of Graduate 
and Postdoctoral Studies for acceptance, the dissertation entitled:

**Interference in Complex Quantum Systems: From localization in high-dimensional lattices**
to surface spin echo with molecules

submitted by **Joshua Tyler Cantin** in partial fulfillment of the requirements for the degree of
**Doctor of Philosophy** in **Chemistry**

**Examining Committee:**

Supervisor **Roman V. Krems, Chemistry**

Supervisory Committee Member **Grenfell N. Patey, Chemistry**

University Examiner **Takamasa Momose, Chemistry and Physics**

University Examiner **Marcel Franz, Physics**

**Additional Supervisory Committee Members:**

Supervisory Committee Member **Mona Berciu, Physics**

Supervisory Committee Member **Allan K. Bertram, Chemistry**
Abstract

This thesis considers (1) novel manifestations and applications of quantum interference in complex systems and (2) development of new approaches to study complex quantum systems. First, I examine a general model of particles with long-range hopping amplitudes. For at least 30 years, it has been widely accepted that these particles do not undergo Anderson localization in 3D lattices. We show that these particles do undergo Anderson localization in 3D lattices if their hopping amplitudes are isotropic. In contrast, particles with anisotropic long-range hopping appear to follow the widely held expectations. We show these results by demonstrating that cooperative shielding extends to 3D cubic lattices with isotropic long-range hopping, but not with anisotropic long-range hopping and by computing the scaling behaviour of participation ratios and energy level statistics.

Secondly, I develop a fully quantum mechanical model of molecular surface spin-echo experiments, which study surface properties and dynamics by scattering molecules off the sample surface. This model, based on the transfer matrix method, incorporates molecular hyperfine degrees of freedom, allows for the efficient calculation of the experimental signal given a molecule-surface scattering matrix, and permits us to begin addressing the inverse scattering problem. This fully quantum model is required to properly describe these experiments as the semi-classical methods used to describe experiments using helium-3 atoms do not take magnetic-field induced momentum changes into account. We apply our method to the case of ortho-hydrogen and then apply Bayesian optimization to determine the molecule-surface scattering matrix elements from a calculated signal, for a scattering matrix defined by three parameters. Our work sets the stage for Bayesian optimization to solve the inverse scattering problem for these experiments.

Finally, I propose using Bayesian model calibration to improve the convergence of Monte Carlo calculations in regions where the sign problem or critical slowing down are an issue. Specifically, Bayesian
model calibration would correct poorly converged Monte Carlo calculations with the information from a small number of well-converged Monte Carlo calculations. As a simple proof of principle demonstration, we apply Bayesian model calibration to a diagrammatic Monte Carlo calculation of the scattering length of a spherical potential barrier.
Lay Summary

For at least 30 years, it has been accepted that particles with long-range hopping are not localized by disorder in 3D lattices. However, we show that these particles can be localized by disorder if their hopping amplitudes are isotropic, while we find particles with long-range anisotropic hopping amplitudes follow these prior expectations. We also develop a fully quantum model of molecular hyperfine interferometry experiments. These experiments study surface properties and dynamics by scattering molecules from surfaces. Our model allows us to interpret these experiments by giving us an efficient tool needed to address the inverse scattering problem. We use Bayesian optimization to begin determining the molecule-surface scattering matrix elements from a calculated signal. Finally, we propose using Bayesian model calibration to improve the convergence of Monte Carlo calculations in regions where the sign problem or critical slowing down cause these calculations to be impractical.
Preface

Chapter 2 is based on the article by Joshua T. Cantin, Tianrui Xu and Roman V. Krems, *Effect of the anisotropy of long-range hopping on localization in three-dimensional lattices*, Physical Review B, 98, 014204 (2018). Specifically, the introduction of the article was expanded with background information to make sections Sections 2.2 – 2.4, Section 2.5 contains the body of the article, and the conclusion of the article was expanded into Sections 2.6 and 2.7. Also, a modified form of the abstract of the article was used in the thesis abstract and in Chapter 1. The original draft of the article was written by me and edited by Roman V. Krems, Tianrui Xu, and me. I wrote the data analysis code for this work as well as the wrapper for the diagonalization algorithm. I also wrote a Python version of the original C++ code to check the original code. I performed most of the data analysis. The project was identified and designed by Roman V. Krems, Tianrui Xu, and me. Copyrights to all sections of this chapter taken from the article are owned by the American Physical Society, 2018.

Chapter 3 is based on the article by Joshua T. Cantin, Gil Alexandrowicz, and Roman V. Krems, *Transfer matrix theory of surface spin echo experiments with molecules*, arXiv:1906.04846 (2019), which is to be submitted to a peer-reviewed journal. Specifically, the introduction of the article was modified to form Sections 3.1 and 3.2, Sections 3.3–3.8 contain the body of the article, and the conclusion of the article was modified to form Section 3.9. Also, a modified form of the abstract of the article was used in the thesis abstract and in Chapter 1. The original draft of the article was written by me and edited by Roman V. Krems, Gil Alexandrowicz, and me. I derived the formulae, with some conceptual aid by both Roman and Gil, and wrote the corresponding code in Python and Octave. I also performed the numerical calculations for the article and for the comparison between the semi-classical and quantum methods. The semi-classical code was provided to me by Helen J. Chadwick and was originally written by Oded Godsi. I made some
modifications to the semi-classical code to allow for a useful comparison. I performed most of the data analysis for this article. The project was identified and designed by Roman V. Krems, Gil Alexandrowicz, and me.

Chapter 3 also frequently references the article by Oded Godsi, Gefen Corem, Yosef Alkoby, Joshua T. Cantin, Roman V. Krems, Mark F. Somers, Jörg Meyer, Geert-Jan Kroes, Tsofar Maniv and Gil Alexandrowicz, *A general method for controlling and resolving rotational orientation of molecules in molecule-surface collisions*, Nature Communications 8, 15357 (2017). I developed and performed the transfer matrix analysis and the corresponding numerical calculations for this article. Roman V. Krems contributed to the development of this transfer matrix analysis. I also wrote the section titled *Time independent approach* of this article’s Supplementary Information, which Roman edited. My contribution to this article was identified and designed by Roman V. Krems, Gil Alexandrowicz, and me.

Chapter 4 is original, unpublished work. The project was identified and designed by Roman V. Krems and me. I wrote the Python code, in consultation of the code provided to me by Rodrigo A. Vargas-Hernández, and performed the numerical calculations and data analysis.

Chapter 5 is original, unpublished work. The project was identified and designed by Roman V. Krems and me. I wrote the Python code for the Bayesian model calibration and performed the numerical calculations and data analysis. The diagrammatic Monte Carlo Fortran code used is freely available on the website of Nikolay Prokof’ev [1].
## Contents

Abstract ................................................................. iii

Lay Summary ............................................................. v

Preface ................................................................. vi

Contents ................................................................. viii

List of Tables ........................................................... xii

List of Figures .......................................................... xiii

Acknowledgements ......................................................... xv

Dedication ............................................................... xvi

1 Introduction ........................................................... 1

1.1 Introduction and Overview ........................................ 1

1.1.1 Anderson Localization ......................................... 1

1.1.2 Surface Spin Echo Experiments with Molecules .............. 2

1.1.3 Monte Carlo Enhanced with Bayesian Model Calibration .... 4

1.2 Main Results ......................................................... 4

2 Anderson localization for particles with long-range hopping? .. 6

2.1 Short- vs. Long-range Hopping ..................................... 6
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2</td>
<td>Anderson Localization with Short-range Hopping</td>
<td>8</td>
</tr>
<tr>
<td>2.3</td>
<td>Anderson Localization with Long-range Hopping</td>
<td>13</td>
</tr>
<tr>
<td>2.4</td>
<td>Cooperative Shielding in One Dimension</td>
<td>17</td>
</tr>
<tr>
<td>2.5</td>
<td>Anderson Localization with Anisotropic Hopping</td>
<td>18</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Models</td>
<td>19</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Cooperative Shielding in Three Dimensions</td>
<td>21</td>
</tr>
<tr>
<td>2.5.3</td>
<td>Participation Ratios</td>
<td>27</td>
</tr>
<tr>
<td>2.5.4</td>
<td>Energy Level Statistics</td>
<td>34</td>
</tr>
<tr>
<td>2.6</td>
<td>Conclusion</td>
<td>38</td>
</tr>
<tr>
<td>2.7</td>
<td>Future Directions</td>
<td>39</td>
</tr>
<tr>
<td>2.8</td>
<td>Summary</td>
<td>41</td>
</tr>
<tr>
<td>3</td>
<td>Molecular hyperfine interferometry</td>
<td>42</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>42</td>
</tr>
<tr>
<td>3.2</td>
<td>From Helium-3 Spin Echo to Molecular Hyperfine Interferometry</td>
<td>43</td>
</tr>
<tr>
<td>3.3</td>
<td>Description of a Molecular Hyperfine Interferometry Experiment</td>
<td>44</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Molecular Hyperfine Hamiltonian</td>
<td>47</td>
</tr>
<tr>
<td>3.4</td>
<td>Impact of the Magnetic Lens on the Molecular States</td>
<td>50</td>
</tr>
<tr>
<td>3.5</td>
<td>Wavepacket Propagation and Signal Calculation</td>
<td>52</td>
</tr>
<tr>
<td>3.6</td>
<td>Transfer Matrix Formalism with Internal Degrees of Freedom</td>
<td>55</td>
</tr>
<tr>
<td>3.6.1</td>
<td>Propagation and Discontinuity Matrices</td>
<td>56</td>
</tr>
<tr>
<td>3.6.2</td>
<td>Rotation Matrices</td>
<td>61</td>
</tr>
<tr>
<td>3.6.3</td>
<td>Scattering Transfer Matrices</td>
<td>62</td>
</tr>
<tr>
<td>3.6.4</td>
<td>Calculation of Eigenstate Coefficients</td>
<td>64</td>
</tr>
<tr>
<td>3.7</td>
<td>Application to ortho-Hydrogen</td>
<td>65</td>
</tr>
<tr>
<td>3.7.1</td>
<td>ortho-Hydrogen Hyperfine Hamiltonian</td>
<td>65</td>
</tr>
<tr>
<td>3.7.2</td>
<td>Experiment and Observables</td>
<td>66</td>
</tr>
<tr>
<td>3.8</td>
<td>Comparison with a Semi-Classical Method</td>
<td>70</td>
</tr>
<tr>
<td>3.9</td>
<td>Conclusion and Future Directions</td>
<td>75</td>
</tr>
</tbody>
</table>
3.10 Summary ............................................. 77

4 Bayesian optimization to determine the molecule-surface scattering matrix .......... 78
  4.1 Introduction ......................................... 78
  4.2 The Inverse Scattering Problem ......................... 78
  4.3 Bayesian Optimization ................................. 79
     4.3.1 Gaussian Processes ............................... 80
     4.3.2 Gaussian Process Model ......................... 83
     4.3.3 Bayesian Optimization ......................... 85
  4.4 Preliminary Results .................................. 88
  4.5 Conclusion and Future Directions ...................... 93
  4.6 Summary ............................................ 95

5 Monte Carlo enhanced with Bayesian model calibration ............................... 96
  5.1 Introduction ......................................... 96
  5.2 Monte Carlo Calculations ............................. 96
  5.3 Bayesian Model Calibration ............................ 98
  5.4 Simplified Bayesian Model Calibration ................ 99
  5.5 Proof of Principle Demonstration ..................... 101
  5.6 Conclusion .......................................... 103
  5.7 Summary ............................................ 105

6 Conclusion .............................................. 106
  6.1 Summary of Results .................................. 106
  6.2 Outlook ............................................. 107
     6.2.1 Anderson Localization with Long-range Hopping .... 108
     6.2.2 Molecular Hyperfine Interferometry ............... 108
     6.2.3 Monte Carlo Enhanced with Bayesian Model Calibration .... 109

Bibliography ............................................ 110
List of Tables

Table B.1 Relative Probabilities of the State Selector and the Detector . . . . . . . . . . . . . . . . . . [131]
List of Figures

Figure 2.1  Localized Wavefunction ............................................................... 10
Figure 2.2  Energy Gap for 1D, 3D Isotropic and 3D Anisotropic Hopping Systems .......... 22
Figure 2.3  Energy Level Structures in the Case of Isotropic and Anisotropic Hopping ........ 26
Figure 2.4  Scaling of Participation Ratios for a 3D System with Isotropic Hopping and \( \alpha = 0 \) . . 28
Figure 2.5  Scaling of Participation Ratios for a 3D System with Anisotropic Hopping and \( \alpha = 0 \) . . 31
Figure 2.6  Scaling of Participation Ratios for a 3D System with \( \alpha = 1 \) for Isotropic and Anisotropic
             Hopping ................................ ............................................................... 32
Figure 2.7  Scaling of Participation Ratios for a 3D System with \( \alpha = 3 \) for Isotropic and Anisotropic
             Hopping ................................ ............................................................... 35
Figure 2.8  Scaling of Mean Energy Level Spacing Ratios for \( \alpha = 3 \) with Isotropic and Anisotropic
             Hopping ................................ ............................................................... 36
Figure 2.9  Scaling of Mean Energy Level Spacing Ratios for Isotropic Hopping with \( \alpha = 1,3 \) Near
             Zero Energy ................................ ............................................................. 37
Figure 2.10 Mean Energy Level Spacing Ratios for \( \alpha = 3 \) as a Function of Filling Fraction and
               Disorder Strength ................................ ................................................. 38
Figure 3.1  Schematic of a Generic Molecular Hyperfine Interferometer ......................... 48
Figure 3.2  Abstraction of a Generic Field Profile of a Molecular Hyperfine Interferometer .... 56
Figure 3.3  Approximate Magnetic Field Profile for an Experiment Using \( oH_2 \) .................... 66
Figure 3.4  Theoretical and Experimental Signals ................................ .................... 67
Figure 3.5  Theoretical and Experimental 2D Fourier Amplitude Plots ............................. 68
Acknowledgements

Christ! Tianrui! Mama! Papa! Roman! Krems group! Gil! Alexandrowicz group! Takuto! Gord! Ute!
Birgitta! Whaley group! Norm! Geoff! Michelle! Claude!

Je veux remercier mon père Joseph Réjean René Cantin, ma gand-mère Marie Marguerite Cantin (née Pom-
inville), et mon grand-père Joseph Ernest Cantin.

Auch, würde Ich danken meine Mutter Sylvia Anna Cantin (geb. Morell), meine Großmutter Christine
Morell (geb. Weirauch), und mein Großvater Anton Morell.

我也想给我妻子徐天睿说谢谢。她支持我支持了几年了。她也鼓励我鼓励了几年了。

I would also like to thank the Natural Sciences and Engineering Research Council of Canada, the Horizon
2020 Research and Innovation Programme, and the International Research Training Group (IRTG 2079)
“Cold Controlled Ensembles in Physics and Chemistry” for their financial support.

Thank you also to the members of IRTG 2079, the members of the Canadian Center for Research on Ultra-
Cold Systems, and to everyone else with whom I interacted during my PhD, the number of whom is always
too large to list.

S.D.G.
In memory of Ernest Cantin and Tony Morell
Chapter 1

Introduction

1.1 Introduction and Overview

Over the past several decades, researchers have achieved an unprecedented level of control of atomic and molecular single and many-body systems [2]. Researchers now routinely capture ensembles of atoms in optical dipole traps [3-5], magneto-optical traps [5, 6], and optical lattices [5 7-9] at milli- to microkelvin temperatures [5 10]. Ensembles of molecules can also be placed in various traps [11-14], though with greater difficulty. The large de Broglie wavelengths of these atoms and molecules gives rise to interesting quantum interference phenomena, from the formation of atomic [9, 15, 16] and molecular [17, 18] Bose-Einstein condensates to Anderson localization [19, 20] (the trapping of a quantum particle by disorder).

1.1.1 Anderson Localization

These new experiments with ensembles of atoms and molecules are stimulating the research field to revisit models that were not thoroughly studied. For example, recent experiments have shown the ability to tune the range of interactions in a one dimensional lattice [21, 22] and there is an experiment that can tune the anisotropy of the interactions of atoms dressed by Rydberg excitations [23]. Thus, one model with renewed interest describes the motion of particles on a disordered lattice that permits long-range hopping amplitudes with various forms of anisotropy. Interestingly, it has been widely accepted for at least 30 years that particles with such long-range hopping do not undergo Anderson localization in three dimensional lattices [24-26], despite a subtle hint otherwise [27]. However, several recent studies demonstrated localization of particles
with long-range hopping in one dimensional lattices [28–30]. In particular, it was recently shown that the effect of long-range hopping in one dimensional lattices can be mitigated by cooperative shielding, which makes the system behave effectively as one with short-range hopping [30].

In Chapter 2, we demonstrate that particles with long-range isotropic hopping amplitudes can undergo Anderson localization in three dimensional lattices and that whether the particle localizes depends on the anisotropy of the hopping amplitudes. We show that cooperative shielding extends to three dimensional lattices with isotropic long-range $r^{-\alpha}$ hopping, but not to three dimensional cubic lattices with anisotropic long-range hopping. We demonstrate the presence of localization in three dimensional lattices with uniform ($\alpha = 0$) isotropic long-range hopping and the absence of localization with uniform anisotropic long-range hopping by using the scaling behaviour of eigenstate participation ratios. We compare the scaling behaviour of participation ratios and energy level statistics to show that the existence of delocalized, non-ergodic extended, or localized states in the presence of disorder depends on both the exponents $\alpha$ and the anisotropy of the long-range hopping amplitudes.

1.1.2 Surface Spin Echo Experiments with Molecules

Greater control over cold atomic and molecular beams allows researchers to use these beams as tools to study problems in other areas of physics. Researchers now use these beams to investigate systems as different as the electric dipole moment of the electron [31] and the dynamics of material surfaces [32]. For instance, $^3$He spin-echo experiments use beams of $^3$He atoms to study surface morphology, molecular and atomic surface diffusion [33–39], phonon dispersions [40, 41], phason dispersions [42] and even phase transitions of ionic liquids [43]. However, the interactions between $^3$He atoms and surfaces or their adsorbates are typically isotropic and weak. To overcome these limitations, we propose to use generic closed shell molecules instead of $^3$He in surface spin-echo experiments. The molecular degrees of freedom, such as rotation, may be exploited to provide additional insight into surfaces and the behaviour of their adsorbates. Indeed, a recent experiment has shown that ortho-hydrogen can be used as a probe that is sensitive to the orientation of a Cu(115) surface [44]. However, the additional degrees of freedom offered by molecules also pose a theoretical challenge: a large manifold of molecular states and magnetic field-induced couplings between internal states.

To address this challenge, in Chapter 3, we develop a fully quantum mechanical approach to model
these molecular surface spin-echo experiments, which are also known as molecular hyperfine interferometry experiments. We thereby connect the experimental signal to the elements of the time-independent molecule-surface scattering matrix. We derive a one-dimensional transfer matrix method that includes the molecular hyperfine degrees of freedom and accounts for the spatial separation of the molecular wavepackets due to the magnetic control fields. This work gives us an efficient tool that we need to address the inverse scattering problem for these experiments and sets the stage for a framework that describes molecular surface spin-echo experiments to study dynamic surfaces. We apply the fully quantum method to the case of ortho-hydrogen, show that the calculated experimental signal is sensitive to the scattering matrix elements, and perform a preliminary comparison to experiment.

With this new framework, we also set the stage for machine learning techniques to determine the scattering matrix elements from experimental measurements. These machine learning techniques are gaining increasing popularity not only in the technology industry, but also in physics and chemistry [45–50]. Neural networks have been used in potential energy surface fitting [51], to identify phase transitions [52, 53], and for quantum state tomography [54]. Support vector machines have been used to identify order parameters [55] and Gaussian processes have been used in potential energy surface fitting [56–58], prediction of a phase transition [59], and scattering calculations [60–62]. Bayesian optimization has been used to tune density functionals [63] and for the inverse scattering problem in quantum reaction dynamics [64].

In Chapter 4, we apply Bayesian optimization to the inverse scattering problem for molecular hyperfine interferometry experiments. The goal of this inverse scattering problem is to determine the several dozen scattering matrix elements from the experimental signal. Such a large parameter space makes it difficult to use traditional optimization methods. Our approach to this problem is to efficiently minimize the difference between the experimental signal and the results of the transfer-matrix computations by using the Bayesian optimization algorithm to intelligently vary the scattering matrix elements. In this way, we apply Bayesian optimization to determine the scattering matrix corresponding to a target calculated signal for a diagonal matrix determined by three or fewer parameters. We also demonstrate that Bayesian optimization is significantly more efficient than grid searches of comparable resolution and lay the foundation for applying Bayesian optimization to determine scattering matrices defined by several tens of parameters.
1.1.3 Monte Carlo Enhanced with Bayesian Model Calibration

Given the complexity of new cold atomic and molecular experiments, such as those studying the two-dimensional Fermi-Hubbard model [65, 66], it is important to have a solid theoretical understanding of the physics involved. As such, there are several many-body numerical techniques that are used to study these systems. These include the density matrix renormalization group [67–69], dynamical mean field theory [70], path integral Monte Carlo [71, 72], and diagrammatic Monte Carlo [73–75]. However, there are still many systems, such as the two dimensional Hubbard model [76] just mentioned, that are not easily addressed by these numerical techniques. Because of these difficulties, researchers are beginning to use machine learning techniques to improve the efficiency of various numerical methods. For example, restricted Boltzmann distributions have been used to learn effective Hamiltonians in classical Monte Carlo calculations [77, 78] and Cui and Krems have used Bayesian model calibration to improve the efficiency of scattering calculations by combining quantum scattering calculations with classical scattering calculations [60].

In Chapter 5, we propose the idea to use Bayesian model calibration to improve the efficiency of Monte Carlo calculations in regions where critical slowing down or the sign problem inhibit the efficient calculation of observables. Particularly, we propose to perform several poorly converged calculations in these regions and then use Bayesian model calibration to correct these calculations with the information obtained from just a few well-converged calculations. Such a procedure would reduce the computational cost required to achieve the desired level of accuracy. As a simple proof of principle demonstration, we apply Bayesian model calibration to a simple diagrammatic Monte Carlo calculation of the scattering length of a spherical potential barrier.

In Chapter 6 of this thesis, I summarize the results of these different projects related to complex quantum systems and discuss their future directions.

1.2 Main Results

In this section, I list the main results of this thesis.

- Showed that particles with long-range isotropic hopping can localize in three dimensional lattices and that, generally, the presence or absence of delocalized, extended non-ergodic, or localized states depends on both the range and the anisotropy of a particle’s hopping amplitudes (Chapter 2)
• Showed that cooperative shielding extends to three-dimensional systems, if the hopping amplitudes are isotropic, while cooperative shielding appears absent if the hopping is anisotropic with the angular form of the interaction between dipoles aligned along one of the principal axes of the cubic lattice (Chapter 2)

• Proposed using generic closed shell molecules for molecular hyperfine interferometry experiments and developed an efficient fully quantum formalism needed to address the inverse scattering problem (Chapter 3)

• Applied the fully quantum formalism to the description of molecular hyperfine interferometry experiments using oH₂ molecules and showed that various features of the scattering matrix affect the experimental signal (Chapter 3)

• Applied Bayesian optimization to the inverse problem for molecular hyperfine interferometry experiments and determined a scattering matrix parametrized by three phases, from data generated from known target parameters (Chapter 4)

• Laid the foundation to efficiently determine, via Bayesian optimization, matrices defined by more than three parameters (Chapter 4)

• Proposed applying Bayesian model calibration to improve the efficiency of Monte Carlo calculations in regions where the sign problem or critical slowing down are problematic (Chapter 5)

• Provided a proof of principle demonstration that Bayesian model calibration can improve the accuracy of many poorly converged Monte Carlo calculations with the information from just a few well-converged calculations (Chapter 5)
Chapter 2

Anderson localization for particles with long-range hopping?

I begin this chapter by precisely defining the terms long- and short-range hopping in Section 2.1. I then discuss the original insights into Anderson localization in Section 2.2. Next, I describe one of the primary influential ideas related to Anderson localization of particles with long-range hopping in Section 2.3. I discuss the recently discovered phenomenon of cooperative shielding in Section 2.4. In Section 2.5, I discuss our results concerning the effect of anisotropy of long-range hopping amplitudes on Anderson localization and the connection with cooperative shielding. I then conclude the chapter and discuss future directions in Sections 2.6 and 2.7, respectively. Finally, I list the key results in Section 2.8.

2.1 Short- vs. Long-range Hopping

Before discussing Anderson localization with short-range hopping, I will first define short- and long-range hopping. A common model of electrons moving in a material is the tight-binding Hamiltonian with nearest neighbour hopping [79]. In this model, each atom is modelled as a lattice site, between two of which the electron can “hop”. This hopping is parametrized by a hopping amplitude that describes the probability amplitude of the electron (or other particle or quasiparticle) to move from one lattice site to another (i.e. from atom to atom). Nearest neighbour hopping describes the situation where the hopping amplitude between two sites is a constant for neighbouring lattice sites, and zero otherwise; this is the shortest possible hopping
range (excepting a range of zero, that is).

There can also be next-nearest neighbour hopping where the particle can hop two sites simultaneously. While longer-ranged than nearest-neighbour hopping, I also consider next-nearest neighbour hopping short-ranged for the purposes of this thesis. Extending this notion further, a particle could in principle move any number of sites in a single hop. Thus, the definition of “short-” or “long-” ranged hopping is, in a sense, arbitrary. Is a hop of 10 sites long? 100 sites?

Given these difficulties, we could instead look at the decay profile of the likelihood the particle hops a particular distance. That is, how does the hopping amplitude \( t(\vec{r}) \) decay as a function of \(|\vec{r}|\), where \( \vec{r} \) is the vector between the initial and final sites? Clearly, if \( t(\vec{r}) \) does not decay as a function of distance, such that the likelihood to hop to the next site or to a site infinitely far away are equal, the hopping can be considered long-ranged. Again, however, we are left with arbitrarily defining a cut-off between short- and long-ranged hopping. A common, but by no means universal, definition is to choose a cut-off in line with whether the integral over all space of the hopping amplitude diverges or converges \[24, 26, 28, 80\].

We examine the integral

\[
I = \int d^d \vec{r} t(\vec{r}),
\]

where \( d \) is the dimension of the lattice. We then declare the hopping to be short-ranged if the integral converges and long-ranged if the integral diverges. Such a definition aligns well with the notion of long- and short-ranged forces \[80, 81\].

For an exponentially decaying \( t(\vec{r}) \), \( I \) is clearly finite. However, if \( t(\vec{r}) \) follows the power-law profile \(|\vec{r}|^{-\alpha}\), the convergence of \( I \) depends on the value of \( \alpha \). In particular, if \( \alpha > d \), the integral converges and the hopping is short-ranged. If \( \alpha < d \), the integral diverges and the hopping is long-ranged. Finally, if \( \alpha = d \), the integral also diverges and the hopping is considered critically long-ranged. Taking this definition of long- and short-ranged hopping, I will now explain the Anderson localization of particles with short-ranged hopping.
2.2 Anderson Localization with Short-range Hopping

Disorder is present throughout materials, from dislocations in crystal lattices to amorphous mixes of organic molecules; it is critical that we understand the impact of this disorder on material physics and energy transport through these materials. In 1958, Anderson [24] was the first to show that a quantum particle remains localized if there is sufficient disorder present within the system. That is, a quantum particle placed at a particular location within a material will eventually diffuse away, unless sufficient disorder exists to trap it in or near its original location. In particular, the disorder scatters the wave-like particle to such an extent that the resultant interference pattern is constructive primarily near the particle’s initial location.

It was later realized that this localization phenomenon is not restricted to quantum particles, but to a whole array of classical wave phenomena. Experiments have shown the presence of Anderson localization in light [82, 83] (even with overhead transparencies! [84]) and sound [85–87], in addition to ultra-cold atoms [19, 20, 88]. Indeed, an optical fibre has been developed that does not use total internal reflection to trap the light within the fibre, but rather localization in the radial direction caused by engineered disorder in the fibre material [89].

Let us look in more detail at a particle, such as an electron, a quasi-particle, or an exciton, moving through a material. First, we model the system as a generic particle moving on a lattice. Such a model applies to electrons moving from atom to atom in a crystal, rotational excitations moving from molecule to molecule in an optical lattice, or to electronic excitations moving from molecule to molecule in an organic-semiconductor. For now, we focus on a particle that only has nearest neighbour hopping in the tight-binding model:

\[
H = \sum_{\langle i,j \rangle} t_{ij} c_j^\dagger c_i + \sum_{i} \epsilon_i c_i^\dagger c_i, \tag{2.2}
\]

where \( t_{ij} \) is the hopping amplitude of a particle to move from lattice site \( i \) to site \( j \), \( c_i^\dagger \) is the creation operator for a particle on site \( i \) and \( \epsilon_i \) is the energy of a particle on site \( i \). The sums run over all sites of a \( d \)-dimensional lattice, while the angle brackets indicate the inclusion of only nearest neighbours in the sum.

If the on-site energies \( \epsilon_i \) and the hopping amplitudes are all identical, \( H \) can be diagonalized in the momentum basis and a particle placed on a single lattice site will ballistically transport away from its origi-
inal site. However, we can introduce disorder into the model by randomizing either $\epsilon_i$, known as on-site or diagonal disorder, or by randomizing $t_{ij}$, known as off-diagonal disorder. Many different probability distributions can be used for this, from uniform, to Gaussian, to binary. Typically, the “strength” of the disorder is characterized by some measure of the width of the distribution; that is, the strength is characterized by how much variation is likely from site to site or hop to hop.

Typically, as disorder is added to the system, the transport changes from ballistic to diffusive. The diffusion coefficient in this regime is a function of the disorder strength. It is initially just renormalized to a smaller value (weak localization), but eventually becomes zero at a particular disorder strength (strong localization). Note, however, that not all systems have this diffusive regime. Particularly, while a three dimensional system exhibits the transition from ballistic to diffusive motion and then from diffusive motion to localization, one and two dimensional systems can be localized by any small amount of disorder [90]. Once the system is in the strong localization regime, the particle can no longer traverse the lattice but retains a profile that decays exponentially from the initial site. That is,

$$\left|\psi(x)\right|^2 \sim e^{-\frac{x-x_0}{\xi}} \text{ for large } x,$$

(2.3)

where $\psi(x)$ is the particle wavefunction as a function of position $x$, $x_0$ is the original location of the particle, and $\xi$ is termed the localization length [91].

There are two things to note. First, though the wavefunction decays exponentially at long distances, it does not do so smoothly but varies widely within an exponential envelope, see Figure 2.1. The wavefunction is also not truly exponentially decaying near $x_0$. Secondly, the localization length is rather loosely defined; differences of factors of two are common in the definition of $\xi$ and differences arise in practical calculation as the definition of “large” $x$ is rather arbitrary. Thus, the localization length is better viewed as a concept and a referential length scale than an actual material property.

At this point, I will discuss the importance of quantum coherence in Anderson localization and demonstrate one of the many subtleties of Anderson localization. Following the Les Houches lecture notes by Müller and Delande [91], let us consider a single particle in a continuous one dimensional system that has no potential except for delta-functions spaced apart at random intervals (i.e. a Dirac comb with random intervals instead of equal intervals).
Figure 2.1: Probability distribution $|\psi|^2$ of a particle initially placed on the centre site of a disordered one dimensional lattice. The Hamiltonian corresponds to Eqn. (2.2) with the hopping amplitudes constant and the on-site energies drawn from a uniform distribution of width $2.5t$. Both panels show the probability distribution at time $T = 50t$ and only differ in the scaling of the ordinate. Exponential, or near exponential, decay of the wave function is observed at large distances.

If we look at two neighbouring delta-functions, spaced apart by a distance $\Delta z$, the transmission $T_{12}$ of a quantum particle from the left of the first delta-function to the right of the second is given by [91]:

$$T_{12} = \frac{T_1 T_2}{|1 - \sqrt{R_1 R_2} e^{ik\Delta z}|^2},$$  \hspace{1cm} (2.4)

where $T_i$ is the transmission through the $i$th delta-function, $R_i$ is the reflection from the $i$th delta-function, and $k$ is the linear momentum of the particle.

If we decide to ignore any coherence between pairs of delta-functions, we can treat $T_{12}$ as the representative transmission of all the pairs and average the phase $\theta \equiv k\Delta z$ over the interval $[0, 2\pi)$ to get an ensemble average [91]:

$$\langle T_{12} \rangle = \frac{T_1 T_2}{1 - R_1 R_2},$$  \hspace{1cm} (2.5)

which is the classical solution. From this, we can get the resistance to be proportional to the system length and obtain a result consistent with Ohm’s law [91]. Clearly, ignoring the coherence between pairs of scatterers does not result in an answer consistent with Anderson localization (cf. Eqn. 2.3). This highlights the
necessity of coherence for Anderson localization; the phenomenon of localization is weakened or destroyed at high temperatures or by the presence of other degrees of freedom (such as phonons) that couple to the particle in question [91].

Thus, to obtain a result consistent with Eqn. 2.3, we would seemingly need to keep track of all of the phase relationships between scattering events and then perform a suitable averaging over all phases. This is no easy task. Intriguingly, there is a simpler route. Let us take the ensemble average of ln $T_{12}$ [91]:

$$\ln T_{12} = \ln T_1 + \ln T_2 + \ln \left| 1 - \sqrt{R_1 R_2} e^{i \kappa \Delta} \right|^2$$

$$\langle \ln T_{12} \rangle = \ln T_1 + \ln T_2 + \left\langle \ln \left| 1 - \sqrt{R_1 R_2} e^{i \kappa \Delta} \right|^2 \right\rangle$$

$$= \ln T_1 + \ln T_2,$$

where the third term on the right hand side of the second line is zero as the integrand is analytic inside the integration contour (i.e. a circle of radius less than 1 and centred at 1).

From the above, we can infer that the total log-averaged transmission $\ln T_{tot}$ through a set of $N$ delta-functions is given by [91]:

$$\langle \ln T_{tot} \rangle = \sum_i N \ln T_i$$

$$= N \ln T_1$$

$$= nL \ln T_1,$$

where it is assumed that all $T_i = T_1$ and that the delta-functions appear at a linear density $n$ in a system of length $L$, such that the number of delta-functions $N = nL$. Thus, the typical transmission $T_{typ}$ through a
region of length $L$ is given by [91]:

$T_{\text{typ}} \equiv e^{\langle \ln T_{\text{tot}} \rangle}
= e^{nL\ln T_1}
= e^{-nL|\ln T_1|}
= e^{-\frac{L}{\xi}}, \quad (2.9)$

where the third lines follows as $T_1 < 1$ and the localization length is identified as $\xi^{-1} \equiv n|\ln T_1|$.

This shows that while the transmission itself is a self-averaging quantity for a classical or incoherent quantum system, it is the log-transmission that is self-averaging in the coherent quantum system (or classical wave system). Indeed, the transmission through a disordered sample varies from sample to sample and if one measures the transmission of a set of disordered systems, the resultant probability distribution is log-normal, rather than normal, and the representative or “typical” transmission of the ensemble is the geometric mean $e^{\langle \ln T_{\text{tot}} \rangle}$ rather than the arithmetic mean $\langle T_{\text{tot}} \rangle$ [91].

The above analysis of the transmission through a disordered system not only highlights the importance of coherence for Anderson localization, but also the subtleties that can appear when trying to find and determine the representative quantities of a disordered system. These subtleties can hinder the analysis of these disordered systems, but also provide for a rich field that includes many different approaches and techniques. These techniques range from locator expansions [24], to scaling theory [90], to replica field theory [92].

Finally, it is also important to understand how Anderson localization manifests itself in different dimensions. In 1979, Abrahams et al. [90] used scaling theory to show that a single particle always localizes in one and two dimensional systems, assuming zero temperature, an infinite system, and no correlation in the disorder. With correlation in the disorder, however, delocalized states can still exist [93–96]. In contrast to these lower dimensions, Abrahams et al. also showed that in three dimensions there is a critical disorder strength that separates a diffusive regime from a localized regime [90]. It is also known that this diffusion-localization transition is a critical transition where the localization length $\xi$ and the diffusion coefficient $D$ have the relations $\xi \sim (w - w_c)^{-\nu}$ and $D \sim (w_c - w)^{\nu}$ with the critical exponent $\nu = 1.58 \pm 0.01$, the
disorder strength $w$, and the critical disorder strength $w_c$ [90, 97, 98].

I would like to emphasize that all of the above results are for systems whose particles have short-range hopping amplitudes or for systems of classical waves. If the hopping amplitudes are long-range, however, many of the above results can change drastically, while others remain the same. In the next section, I will discuss some of the known results and how well they actually apply to systems with long-range hopping.

### 2.3 Anderson Localization with Long-range Hopping

In Section 2.2, the presence or absence of Anderson localization in non-interacting systems with nearest neighbour hopping was examined for various dimensionalities. Here, I discuss situations where the hopping amplitude $t_{ij}$ of Eqn. (2.2) is beyond nearest-neighbour and decays as $r^{-\alpha}$, where $\alpha \in \mathbb{R}_{\geq 0}$ is the decay exponent or the hopping “range”. To reiterate, I consider the hopping “short-ranged” when $\alpha > d$, the system dimension, “long-ranged” when $\alpha \leq d$, and critically long-ranged when $\alpha = d$. There is increasing interest in disordered lattice models with long-range hopping, as excitons with hopping range $\alpha = 3$ are found in molecular crystals with topological disorder [99], organic semiconductors with impurities [100], diluted ensembles of atoms and molecules trapped in optical lattices [101–106], $J$-aggregates [107], photosynthetic complexes [108, 109], and ensembles of Rydberg atoms [110–112].

Systems whose hopping amplitudes exhibited such a power-law decaying profile were also already examined by Anderson in his seminal paper [24]. Through use of a locator expansion, Anderson concluded that localization is present when the hopping is short-ranged and absent otherwise. Later Levitov [25, 26, 113] approached the system with a resonance argument and came to the same conclusion. I will describe this argument in detail, with some slight generalizations.

In essence, a particle with long-range hopping can diffuse away from its initial lattice site because of resonances between sites that are an arbitrary distance away from each other [25]. Such resonances occur whenever the hopping amplitude $t_{ij}$ between sites $i$ and $j$ is similar or larger in magnitude to the difference in on-site energy between the two sites. That is, a resonance occurs when

$$t_{ij} \gtrless |E_j - E_i|,$$

(2.10)

where $E_i$ is the energy of the particle at site $i$ [25].
Let’s assume that the particle is initially localized to some single lattice site $i$. To prove that this particle does not remain localized, Levitov [25] showed that site $i$ is in resonance with another lattice site $j$ such that the distance between the two sites $|\vec{r}_j - \vec{r}_i| > R$, where $R$ is an arbitrarily large distance. We now follow the main ideas of Levitov’s proof, but generalized to a $d$-dimensional system [25]. We begin by taking concentric $(d-1)$-spheres centred on the site $i$ such that [25]

\[
2^k R < |\vec{r}_j - \vec{r}_i| < 2^{k+1} R, \tag{2.11}
\]

where $k \in \mathbb{Z}$ and $\vec{r}$ is some distance vector. Now, the volume $V_{k,d}$ of the $k$th region is

\[
V_{k,d} = \text{vol}_d \left( 2^{k+1} R \right) - \text{vol}_d \left( 2^k R \right) \\
\propto \left( 2^{k+1} R \right)^d - \left( 2^k R \right)^d \\
\propto \left( 2^k R \right)^d, \tag{2.12}
\]

where $\text{vol}_d (r)$ gives the volume of a $(d - 1)$-sphere of diameter $r$. Taking the lattice site density $n$, the number of sites between the concentric spheres is given by $nV_{k,d} [25]$.

Now, assuming that the on-site energies are identically and independently drawn from a uniform distribution centred about zero and of width $W$, the probability distribution of the difference in energy between two sites $\Delta \equiv |E_j - E_i|$ is triangular and given by

\[
P(\Delta) = \frac{2}{W} \left( 1 - \frac{\Delta}{W} \right). \tag{2.13}
\]

Taking the hopping amplitude $t_{ij} = \frac{1}{|\vec{r}_{ij}|}$, where $r_{ij} \equiv |\vec{r}_j - \vec{r}_i|$, the probability that sites $i$ and $j$ are in resonance
is given by

\[ P_{\text{res}} = P\left( \Delta \leq t_{ij} \right) \]

\[ = \int_0^{t_{ij}} d\Delta P(\Delta) \]

\[ = \frac{2t_{ij}}{W} - \frac{t_{ij}^2}{W^2} \]

\[ = \frac{2}{W r_{ij}^\alpha} + O\left( \frac{1}{r_{ij}^\alpha} \right) \]

\[ \approx \frac{2}{W \left( 2^k R \right)^\alpha}, \quad (2.15) \]

where \( t_{ij} = \frac{1}{\bar{r}_{ij}} \) is substituted in the penultimate line and the final line arises as \( \bar{r}_j \) is taken to be in the region between the \( k \)th and \((k+1)\)th concentric \((d-1)\)-spheres, see Eqn. (2.11). Now, the number of resonances \( N_k \) in the \( k \)th shell is given by

\[ N_k = nV_{k,d}P_{\text{res}} \]

\[ = \frac{2n \left( 2^k R \right)^d}{W \left( 2^k R \right)^\alpha} \]

\[ = \frac{2n}{W \left( 2^k R \right)^{\alpha-d}}, \quad (2.16) \]

From the above, we can see that three particular cases arise:

(i) If \( \alpha > d \), then \( N_k \to 0 \) as \( k \to \infty \). That is, there are resonances only a finite distance away and the system remains localized.

(ii) If \( \alpha < d \), then \( N_k \to \infty \) as \( k \to \infty \). That is, there are an infinite number of resonances infinitely far away that cause the system to delocalize.

(iii) If \( \alpha = d \), then \( N_k = \text{const.} \) as \( k \to \infty \). That is, there are still at least some resonances infinitely far away and the system delocalizes.

We note that the neglected \( O\left( \frac{1}{r_{ij}^\alpha} \right) \) term does not change the above conclusions as, when \( \alpha \geq d \), the neglected term approaches zero as \( k \to \infty \) and, when \( \alpha < d \), the neglected term cannot change the fact that
$N_k$ diverges as $k \to \infty$.

The above analysis thus predicts that systems with short-range hopping localize, while those with long-range hopping delocalize. A similar analysis has also been recently applied to study many-body localization of particles with long-range interactions \cite{114}. The analysis by both Anderson \cite{24} and Levitov \cite{25, 26, 113} have led to it becoming widely accepted that non-interacting particles with long-range hopping do not undergo Anderson localization. However, this is despite the fact that, in the same year as Levitov’s work \cite{25}, Burin and Maksimov \cite{27} use a renormalization group analysis to show that one may expect a system with long-range hopping to localize under certain conditions. Burin and Maksimov indicate that their differing prediction from Levitov’s analysis arises as they treat a system whose hopping amplitudes are isotropic with respect to angle, while the hopping amplitudes in Levitov’s system have a dipolar-like angular dependence.

Interestingly, the analysis I presented above, as distilled from Ref. \cite{25}, does not appear to depend on the angular dependence of the hopping amplitudes. This calls into question why the paper by Levitov \cite{25} and that of Burin and Maksimov \cite{27} come to different conclusions. Unfortunately, this difference has not been appreciated by the community. Indeed, the paper by Burin and Maksimov \cite{27} has had only a single citation \cite{115} until 2018, when it received two more \cite{116, 117}.

Furthermore, an increasing number of papers have identified situations where long-range hopping (or interactions) does not result in delocalization. Deng et al. \cite{28} have seen algebraic localization of a particle with long-range hopping in a 1D system with off-diagonal disorder. Nandkishore and Sondhi \cite{118} have shown using field-theory techniques that many-body systems with long-range interactions in one and two dimensions can localize and hypothesized that the same is true in three dimensions. Nandkishore and Sondhi even go so far as to question the validity of the locator expansion and resonance arguments for systems with long-range hopping.

Furthermore, Santos et al. \cite{119} and Celardo et al. \cite{30} have discovered a phenomenon, cooperative shielding, that causes effective short-range behaviour in a system with long-range interactions or hopping. Cooperative shielding precludes the resonance arguments used to demonstrate delocalization in the presence of on-site disorder and allows localization to take place. Celardo et al. \cite{30} have in particular demonstrated localization in one dimension for particles with long-range hopping. Ossipov has also demonstrated local-
ization of particles with uniform hopping on a $d$ simplex, via a similar mechanism [29]. In the next section, I discuss the phenomenon of cooperative shielding in one dimension and how it impacts localization behaviour.

### 2.4 Cooperative Shielding in One Dimension

Cooperative shielding in one dimensional systems was first discovered by Santos et al. [119] and Celardo et al. [30] in 2016. In the non-interacting case, cooperative shielding describes the situation where the long-range terms of the Hamiltonian cause the Hilbert space to split into a long-range and a short-range subspace. The short-range subspace can be described by an effective short-ranged Hamiltonian. In the thermodynamic limit, if a wavefunction begins in the short-range subspace, it will be described by the effective short-ranged Hamiltonian for all time. Adding disorder to this short-ranged effective Hamiltonian can then cause the particle to localize.

Following Ref. [119], one can take the Hamiltonian

\[ \hat{H} = \hat{H}_0 + \hat{V}, \]  

(2.17)

which consists of a short-ranged part $\hat{H}_0$ and a long-ranged part $\hat{V}$ such that $[\hat{H}_0, \hat{V}] = 0$. If $\hat{V}$ has a degenerate eigensubspace $\mathcal{V}$ such that $\hat{V} | v_k \rangle = \nu | v_k \rangle \forall | v_k \rangle \in \mathcal{V}$, then

\[ | \psi(t) \rangle = e^{-i \frac{\hat{H}_0 t}{\hbar}} | \psi(0) \rangle \]

\[ = e^{-i \mathcal{V} t} e^{-i \frac{\hat{H}_0 t}{\hbar}} | \psi(0) \rangle, \]

(2.18)

if $| \psi(0) \rangle = \sum_k c_k | v_k \rangle$, where $c_k$ are the expansion coefficients of the initial wavefunction $| \psi(0) \rangle$. In this situation, where the wavefunction is initially contained within the eigensubspace $\mathcal{V}$, one can see that the only effect of the long-range part of the Hamiltonian is to add a global phase; the effective Hamiltonian is $\hat{H}_0$ and the wavefunction is “shielded” from the long-range terms of $\hat{H}$ [119].

Now, however, if $[\hat{H}_0, \hat{V}] \neq 0$, but is such that the Hilbert subspace $\mathcal{V}$ is still quasi-degenerate, the shielding effect still occurs, albeit for finite time [119]. The shielding time is positively related to the gap between the Hilbert subspaces and, if the gap grows with the system size, diverges in the limit of infinite
system size. This divergence with the system size results in cooperativity $[30]$ $[119]$.

Thus, we see that systems with long-range hopping can, in principle, localize, at least in one-dimension. Indeed, Celardo et al. $[30]$ demonstrate single particle localization for a one dimensional system with long-range hopping. In the next section, I discuss how we have extended the notion of cooperative shielding to three dimensions and demonstrate that the presence of cooperative shielding is dependent on the angular profile of the hopping amplitudes.

### 2.5 Anderson Localization with Anisotropic Hopping

As mentioned in prior sections, the effect of anisotropy of long-range hopping on localization in high-dimensional lattices has not been examined. We examine this here by studying the impact of anisotropy on cooperative shielding in 3D lattices. Specifically, we show that the cooperative shielding demonstrated previously for 1D lattices extends to particles with isotropic long-range hopping in 3D cubic lattices. However, this cooperative shielding does not extend to particles in 3D cubic lattices with the anisotropic hopping that corresponds to the dipole-dipole interaction between dipoles aligned along one of the lattice axes. We also provide numerical and analytical evidence for the localization of particles with long-range isotropic hopping in 3D lattices.

The remainder of this chapter is organized as follows: Following the description of the models, we discuss the phenomenon of cooperative shielding in 3D systems. We analytically diagonalize the Hamiltonian with isotropic hopping $t \propto 1/r^\alpha$ for 3D cubic lattices for arbitrary $\alpha$ and illustrate that it exhibits the same energy level structure as that of a 1D system with cooperative shielding. We find this to be the case for both periodic and open boundary conditions. The presence of cooperative shielding suggests the presence of localization. Contrastingly, we find no clear evidence for cooperative shielding at most values of $\alpha$ for anisotropic dipolar hopping, for either periodic or open boundary conditions.

We use the scaling behaviour of participation ratios to demonstrate the existence of localized states for isotropic hopping when $\alpha = 0$ (i.e. uniform infinite-range hopping), even for weak disorder. We show that the anisotropic hopping case we consider contains only non-ergodic extended and delocalized states, even for strong disorder. This provides further evidence that cooperative shielding does not exist for the anisotropic hopping considered here, as cooperative shielding is expected to be strongest when $\alpha = 0$.  

18
We then examine finite \( \alpha \), and show that anisotropic hopping again only supports delocalized and non-ergodic extended states for \( \alpha = 1 \). Interestingly, for anisotropic hopping with \( \alpha = 3 \) (i.e. dipolar interactions), we observe no delocalized states given sufficient disorder. Whether the states are localized or non-ergodic extended is indeterminate, however. The isotropic hopping case supports delocalized states and either non-ergodic extended or localized states (or both) for \( \alpha = 1 \) and 3.

Following this, we use the scaling behaviour of energy level statistics to investigate the physically important case of \( \alpha = 3 \). Our results indicate the presence of at least some localized states near zero energy for isotropic hopping, while the anisotropic case is inconclusive. For \( \alpha = 1 \) and isotropic hopping, energy level statistics also indicate the presence of, at least, some localized states near zero energy. To connect with currently attainable experiments, we also examine the phase diagram for both diagonal and binary off-diagonal disorder (as per the quantum percolation model) when \( \alpha = 3 \). This provides a basic map that can guide experiments with relevant finite systems, such as polar molecules on an optical lattice [105].

### 2.5.1 Models

We consider a single particle in a disordered and diluted cubic lattice with \( N \) sites per dimension, with the hopping amplitude \( \propto \frac{1}{r^\alpha} \), where \( \alpha \) sets the hopping range. The angular dependence of the hopping amplitude is described below. The Hamiltonian we consider has the following general structure:

\[
\hat{H} = \sum_i \omega_i \hat{c}^\dagger_i \hat{c}_i + \sum_i \sum_{j \neq i} t_{ij} \hat{c}^\dagger_i \hat{c}_j, \tag{2.19}
\]

where the operator \( \hat{c}_i \) removes the particle from site \( i \), \( \omega_i \) is the energy of the particle in site \( i \), and \( t_{ij} \) is the amplitude for particle tunnelling from site \( j \) to site \( i \). We introduce disorder by randomizing both the values of \( \omega_i \) and \( t_{ij} \), which makes Eqn. (2.19) relevant for both disordered lattices and amorphous systems.

We randomize the values \( \omega_i \in [-\omega/2, \omega/2] \) by drawing them from a uniform distribution. We define the disorder strength as

\[
W \equiv \frac{\omega}{t_{\text{max}}}, \tag{2.20}
\]

where \( t_{\text{max}} \equiv \max|t_{ij}| \), in order to allow a direct comparison between the isotropic and anisotropic models.
defined below by normalizing the disorder amplitude to the largest hopping amplitude present.

We randomize the values $t_{ij}$ as in the site percolation model. To do this, we define $t_{ij} = d_{ij} \tau_{ij}$, where $d_{ij}$ is the dilution parameter, and divide the lattice sites into two subsets $P$ and $Q$, with $pN^3$ sites in the $P$ subset and $(1 - p)N^3$ in the $Q$ subset. For a given value $p$, the lattice sites are assigned to the subsets at random. The dilution parameter is then defined as:

$$d_{ij} = \begin{cases} 
1, & i \in P \text{ and } j \in P \\
0, & i \in Q \text{ and/or } j \in Q 
\end{cases} \quad (2.21)$$

With $t_{ij}$ and $d_{ij}$ thus defined, the Hamiltonian (2.19) describes a generic particle in a disordered, diluted lattice with $pN^3$ sites. The angular dependence of the hopping amplitude is determined by the magnitudes of $\tau_{ij}$. We consider two types of models for $\tau_{ij}$.

### Isotropic Hopping

For the case of isotropic hopping, we define

$$\tau_{ij}^I = \frac{\gamma}{|\vec{r}_{ij}|^\alpha}, \quad (2.22)$$

where $\vec{r}_{ij} = \vec{r}_i - \vec{r}_j$ is the distance between sites $i$ and $j$ in the 3D lattice. With $\alpha > 3$ the hopping is short-range, with $\alpha \leq 3$, long-range. The value of $\gamma$ is chosen such that $\tau_{ij}^I \equiv \bar{t} = 1$ for nearest neighbour (NN) sites. Thus, $t_{\max}$ is 1 for isotropic hopping.

### Anisotropic Hopping

We choose the tensorial form of the anisotropic hopping to be the same as the dipole-dipole interaction between polar molecules subjected to an electric field along the $z$-direction, see, for example, Ref. [120]:

$$\tau_{ij}^A = \frac{\gamma \left(1 - 3 \cos^2 \theta_{ij}\right)}{|\vec{r}_{ij}|^\alpha}, \quad (2.23)$$

where $\theta_{ij}$ is the angle between $\vec{r}_{ij}$ and the $z$-axis. We choose $\gamma$ as defined above, which makes $\tau_{ij}^A = 1$ for NN sites in the $x$-$y$ plane and $\tau_{ij}^A = -2$ for NN sites along the $z$-axis. Thus, $t_{\max} = 2$ for the anisotropic hopping.
considered here.

Note that, for the remainder of this chapter, the term *anisotropic* refers specifically to the form in Eqn. (2.23), where \( \theta_{ij} \) is the angle between \( \vec{r}_{ij} \) and the z-axis, with the z-axis chosen to lie along one of the principal axes of the cubic lattice. Changing the alignment of the dipoles to be along another direction (*i.e.* changing the direction of the electric field) may result in different localization behaviour. Furthermore, many other forms of anisotropy are possible (*i.e.* quadrupolar-like, octopolar-like, etc.), but are beyond the scope of this chapter.

In this section (2.5), the system parameters of interest are \( p, W, N, \alpha \) and whether the hopping is iso- or anisotropic. For all of the numerical results, we perform exact diagonalization to obtain *all* of the eigenvalues and eigenstates of the dense Hamiltonian matrix. These eigenvalues and eigenstates are needed to perform the energy level statistics without any approximations and to examine the participation ratios across the entire spectrum. This restricts the size of the system to be smaller than what can be achieved with state-of-the-art Jacobi-Davidson algorithms, which only obtain a small subset of the total number of eigenvalues and eigenstates [121, 122]. However, we are still able to observe scaling behaviour for some properties.

### 2.5.2 Cooperative Shielding in Three Dimensions

Recent work has shown the presence of cooperative shielding in 1D single- and many-body systems with long-range hopping or interactions [30, 119]. Cooperative shielding allows the dynamics of a Hamiltonian with long-range features to be described by an effective short-range Hamiltonian for a finite time, *i.e.* the dynamics are effectively “shielded” from the long-range components for a finite time.

This phenomenon occurs because of the formation of short- and long-range subspaces in the Hilbert space, in turn caused by the long-range terms in the Hamiltonian. In many instances, the short-range subspaces have many more states than, and are separated by an energy gap from, the long-range subspace. When this gap increases with the system size, it increases the “shielding” time and makes the shielding cooperative. In the infinite size limit, the dynamics of a wavefunction initially in the short-range subspace are then completely governed by the effective short-range Hamiltonian. Cooperative shielding allows for the localization of particles with long-range hopping [30] and provides an explanation as to why localization is observed in various systems with long-range hopping or long-range interactions [27, 28, 118].
Here, we show that the energy level structure and gap behaviour conducive to cooperative shielding in 1D are also present in 3D cubic lattices, if the long-range hopping is isotropic. If the hopping is of the anisotropic form (2.23), the energy level structure becomes different and does not exhibit well-separated Hilbert subspaces, precluding the cooperative shielding described in [30].

**Figure 2.2:** The energy gap at the top of the spectrum as a function of the hopping range exponent $\alpha$. Panel A is for long-range hopping in a 1D lattice with periodic boundary conditions. $\Delta_{1D}^{\alpha}$ is determined analytically; see Eqn. (2.29) and associated text. Panel B is for a 3D lattice with isotropic hopping. $\Delta_{3D}^{\alpha}$ is determined analytically under periodic boundary conditions; see Eqn. (2.27). Panels C and D show the energy gap for the Hamiltonian with anisotropic hopping. The gap energy is determined numerically and under open boundary conditions. Panel D is an expanded view of panel C. $N$ is the system side-length (total of $N^3$ sites in the 3D case). In all plots, $p = 1$ and $W = 0$. Note that we examine the top of the spectrum as the effective mass is negative.

**Isotropic Hopping**

To illustrate cooperative shielding in 3D, we first diagonalize analytically the Hamiltonian for an ideal lattice with hopping of arbitrary range. The model with the isotropic hopping (2.22) can be written, in the absence
of disorder and with periodic boundary conditions, as

$$\hat{H}_\alpha^I = \sum_{i,j,h,l,m,n=-\frac{N-1}{2}}^{\frac{N-1}{2}} \frac{\zeta}{r^\alpha} \left( 1 - \delta_{il} \delta_{jm} \delta_{hn} \right) \hat{c}^\dagger_{lmn} \hat{c}_{ijh}, \quad (2.24)$$

where $\zeta > 0$ is the hopping parameter, the lattice side-length $N$ is assumed odd for simplicity, $\hat{c}_{ijh}$ is the annihilation operator on site $(i,j,h)$, $r = \sqrt{(l-i)^2 + (m-j)^2 + (n-h)^2}$ is the unit-less distance between sites, and the term $\frac{1-\delta_{il} \delta_{jm} \delta_{hn}}{r^\alpha}$ is understood to be zero when $i = l$, $j = m$, and $h = n$ (i.e. the particle can hop anywhere except to its original site). Note that $\zeta > 0$ implies a negative effective mass, such as is seen for holes in a semi-conductor. Given that we examine only non-interacting particles, whether the effective mass is positive or negative does not change the final results, except for where one would expect to find the relevant shielding gap (i.e. at the top or bottom of the spectrum).

The case of $\alpha = 0$ produces the uniform, isotropic Hamiltonian $\hat{H}_0^I$, which can be diagonalized in the momentum basis to yield

$$\hat{H}_0^I = \zeta N^3 \hat{c}^\dagger_{000k} \hat{c}_{000k} - \zeta \sum_{k_1k_2k_3} \hat{c}^\dagger_{k_1k_2k_3} \hat{c}_{k_1k_2k_3}, \quad (2.25)$$

where the subscript of $\hat{c}_{000k}$ means $k_1 = k_2 = k_3 = 0$. The energy gap between this specific state $|000k\rangle$ and the $(N^3 - 1)$-fold degenerate states $|k_1k_2k_3\rangle$, with one or more of $k_1, k_2, k_3$ not equal to zero, is thus $\Delta = \zeta N^3$ for the uniform, isotropic Hamiltonian. This is the energy gap responsible for cooperative shielding. In the 1D case [30], this gap is $\Delta = \zeta N$, which allows us to surmise that in the 2D case, $\Delta = \zeta N^2$, and that for any lattice geometry or dimension, $\Delta = \zeta M$, where $M$ is the total number of sites. This occurs because the Hamiltonian (2.24) with $\alpha = 0$ for an arbitrary lattice with $M$ sites is proportional to the adjacency matrix of a complete graph with $M$ nodes. The spectrum of this adjacency matrix is known [123] to contain one $(M - 1)$-fold degenerate eigenvalue and one non-degenerate eigenvalue, which are separated by a gap proportional to $M$. This further implies that all lattice dimensions and geometries are equivalent for a particle with uniform isotropic hopping. In agreement with this, the above energy level structure (2.25) has been observed by Ossipov, who studied the Anderson model on the $d$-simplex [29].

The limiting case of $\alpha \to \infty$ produces the tight-binding model, also diagonal in momentum space. Given that the eigenstates of these two limits are momentum states and that the $\frac{1}{r^\alpha}$ factor does not break any
additional symmetries, the eigenstates of \( \hat{H}_\alpha \) with finite \( \alpha \) must also be momentum states. Evaluating the matrix elements of \( \hat{H}_\alpha \) in momentum space, we find

\[
E_{k_1,k_2,k_3}^I(\alpha) = \langle k_1 k_2 k_3 | \hat{H}_\alpha^I | k_1 k_2 k_3 \rangle = \\
2\zeta \sum_{l=1}^{N-1} \left( 1 - \frac{l}{N} \right) \cos \left( \frac{2\pi k_l}{N} \right) + \cos \left( \frac{2\pi l}{N} \right) \\
+ 4\zeta \sum_{l=1}^{N-1} \sum_{m=1}^{N-1} \left( 1 - \frac{l}{N} \right) \left( 1 - \frac{m}{N} \right) \cos \left( \frac{2\pi k_l}{N} \right) \cos \left( \frac{2\pi k_m}{N} \right) + \cos \left( \frac{2\pi l}{N} \right) \cos \left( \frac{2\pi m}{N} \right) \\
+ 8\zeta \sum_{l=1}^{N-1} \sum_{m=1}^{N-1} \sum_{n=1}^{N-1} \left( 1 - \frac{l}{N} \right) \left( 1 - \frac{m}{N} \right) \left( 1 - \frac{n}{N} \right) \cos \left( \frac{2\pi k_l}{N} \right) \cos \left( \frac{2\pi k_m}{N} \right) \cos \left( \frac{2\pi k_n}{N} \right) \\
\left( l^2 + m^2 + n^2 \right)^{\alpha/2},
\] (2.26)

where \( k_i \in \{ k | k \in \mathbb{Z} \wedge k \in [-\frac{N-1}{2}, \frac{N-1}{2}] \} \) refers to the component of the reciprocal lattice vector in the direction \( i \), such that the crystal momentum is \( 2\pi k_i/Na \), \( a \) is the lattice constant, and \( |k_1 k_2 k_3\rangle \) is a momentum eigenstate. Details about the derivation can be found in Appendix A. It can be shown that Eqn. (2.26) reduces to \( \langle k_1 k_2 k_3 | \hat{H}_0^I | k_1 k_2 k_3 \rangle \), see Eqn. (2.25), in the limit where \( \alpha \to 0 \) and to the tight-binding model result in the limit where \( \alpha \to \infty \) and \( N \to \infty \).

The gap for generic \( \alpha \) can be defined as

\[
\Delta_\alpha = E_{0,0,0}^I(\alpha) - E_{0,0,1}^I(\alpha)
\] (2.27)

Given Eqn. (2.26) and \( \zeta > 0 \), it is clear that \( E_{0,0,0}^I(\alpha) \) is the largest eigenvalue and that the next highest energy level has states with a single unit of momentum. This energy level is six-fold degenerate (i.e. \( E_{0,0,\pm 1}^I = E_{\pm 1,0,0}^I \forall \alpha \)), as can be seen from the symmetries of \( E_{k_1,k_2,k_3}^I(\alpha) \) in \( k_i \); the choice of \( E_{0,0,1}^I(\alpha) \) in the definition of \( \Delta_\alpha \) is arbitrary. Also, since \( \zeta > 0 \), \( E_{0,0,0}^I \geq E_{0,0,1}^I \), and \( \Delta_\alpha \geq 0 \).

For comparison, in 1D the Hamiltonian is given by:

\[
\hat{H}_\alpha^{1D} = \sum_{i,j=-\frac{N-1}{2}}^{\frac{N-1}{2}} \zeta |i-j|^{\alpha} \hat{c}_i \hat{c}_j.
\] (2.28)
The corresponding eigenvalues are:

$$
E^{I,1D}_k(\alpha) = \langle k | \hat{H}^{I,1D}_\alpha | k \rangle = 2 \zeta \sum_{l=1}^{N-1} \left( 1 - \frac{l}{N} \right) \cos \left( \frac{2\pi kl}{N} \pi \right),
$$

(2.29)

where $k \in \{ k | k \in \mathbb{Z} \land k \in [-\frac{N-1}{2}, \frac{N-1}{2}] \}$ is the component of the reciprocal lattice vector. A corresponding gap can be defined as $\Delta^{I,1D}_\alpha = E^{I,1D}_0(\alpha) - E^{I,1D}_1(\alpha)$ and is shown in Figure 2.2 (A). This 1D Hamiltonian (2.28) has been shown to exhibit cooperative shielding in Ref. [30].

Figure 2.2 shows $\Delta_\alpha$ as a function of $\alpha$ for various system sizes. Panel (A) of Figure 2.2 plots the results for the 1D lattice. Panel (B) shows that, for a 3D lattice with isotropic hopping, the dependence of the gap separating the long-range and short-range subspaces on $\alpha$ is qualitatively the same as in 1D, except that the gap becomes independent of the system size at $\alpha = 3$, instead of $\alpha = 1$. We have confirmed by numerical calculations that this behaviour is the same for even and odd $N$ and with open boundary conditions. Based on the 1D and 3D results of Figure 2.2, it should be expected that a 2D system with isotropic hopping likely has a similar gap behaviour, exhibiting the transition at $\alpha = 2$.

The Anderson transition is known to exist in 3D for $\alpha > 3$ [24], but questions arise for the case with $\alpha \leq 3$. Figure 2.2 (B) shows that cooperative shielding can be expected for isotropic hopping with $\alpha < 3$, suggesting that the Anderson transition could also occur in this region. As a side note, the localization may not be exponential, as seen in [28].

Note that the curves in Figure 2.2 (B) corresponding to different lattice sizes all converge at the transition point. Thus, as in 1D [30], the gap is independent of lattice size when $\alpha$ equals the dimension (in our case, three). If the disorder strength is smaller than the gap, then the gap between the long- and short-range subspaces remains open. A finite gap would seemingly indicate a finite shielding time, perhaps precluding localization at infinite times for $\alpha = 3$. However, consider a quantum particle placed at an individual lattice site. The contribution of the delocalized state $|000_k\rangle$ to the wave packet of this particle is proportional to $\frac{1}{N}$, indicating a zero contribution in the infinite lattice size limit. This leaves the possibility of cooperative shielding and localization open in the infinite size limit, but does not confirm them. In the following sections, we examine the effect of cooperative shielding on localization in 3D lattices.
**Anisotropic Hopping**

The dependence of the gap separating the short- and long-range subspace is markedly different for the model with anisotropic hopping (see Figure 2.2 (C) and (D)). It is more difficult to obtain the analytical results for the model with anisotropic hopping because of the $\cos^2 \theta$ term, even if the eigenstates of the system with periodic boundary conditions are still momentum states (rotational symmetry is broken, but not translational). Here, we study the behaviour of the gap at the top of the spectrum as a function of $\alpha$ numerically, using open boundary conditions. The results shown in Figure 2.2 (C) illustrate the qualitative difference from the isotropic case. In particular, for the anisotropic model, there are several values of $\alpha < 3$ where the gap becomes zero. At these values of $\alpha$, cooperative shielding may not occur, thus precluding localization.

From examining the energy level structure, it can be concluded that the zero-gap points arise from level crossings. It is thus clear that there is no separation of subspaces at the top of the spectrum, in contrast to the case of isotropic hopping. While there is a gap that grows with the system size for $\alpha \lesssim 2$, the state at the top of the spectrum is not an equal superposition of all sites, as it is when cooperative shielding is known to occur. Furthermore, the top energy level is shown to be a doublet under periodic boundary conditions,
indicating that this particular gap actually disappears in the infinite size limit when translational invariance is restored.

There is, however, a gap between the highest-energy doublet and the next doublet. We examine this gap for \( N = 11 \) and periodic boundary conditions. The curve (not shown) appears similar to those in Figure 2.2 (C) and (D), though the number of zero-gap points is reduced. This is likely because the higher symmetry permits the formation of doublets, which produce fewer level crossings (i.e. one for each doublet instead of one for each state). Given these level crossings, the same conclusions can be reached for the periodic boundary case as for the open boundary case: there is no clear separation of subspaces at the top of the spectrum and no evidence for cooperative shielding. While there is a gap at small values of \( \alpha \), the states at the top of the spectrum are not equal superpositions of all sites and not necessarily fundamentally different from the rest of the eigenstates. Cooperative shielding, of the form discussed in [30], thus appears precluded.

This is investigated more fully further below by examining the scaling behaviour of the participation ratios.

In addition to the gap behaviour, the overall energy level structure appears different between the two cases, as exemplified in Figure 2.3 for ideal lattices of side-length \( N = 30 \). Here, the top several eigenstates are shown for both isotropic (left panel) and anisotropic (right panel) hopping with open boundary conditions. The values are obtained numerically. Both the degeneracies of the levels and the overall behaviour as a function of \( \alpha \) are very different. The energy level crossings in the anisotropic hopping case are in the region \( 2 \lesssim \alpha \lesssim 3 \). Their exact locations are best observed as the zero-gap points in Figure 2.2 (D). Note that the differences between Figure 2.3 (left panel) and Eqn. (2.26), particularly in terms of the degeneracy layout, are due to the differing boundary conditions (open vs. periodic).

### 2.5.3 Participation Ratios

We have established the presence of cooperative shielding in 3D systems with isotropic long-range hopping, which suggests the possibility of localization. We now examine this possibility, for both isotropic (2.22) and anisotropic (2.23) hopping with open boundary conditions and \( \alpha = 0, 1, \) and 3, from a participation ratio point of view. In a later section, we examine the energy level statistics for \( \alpha = 3 \).

The participation ratio (PR) is an effective count of the number of sites occupied by a wavefunction. It
Isotropic, $\alpha = 0$

**Figure 2.4:** Participation ratio (PR) as a function of energy for the eigenstate of a 3D system with isotropic hopping, $W = 16.5$, $p = 1$, and $\alpha = 0$. The number of disorders included ranges from 148 to 2100, as required to obtain sufficiently small error bars. *Left Panel:* Plot of the PR of each eigenstate vs its eigenvalue for several disorder realizations for $N = 11$, 21, and 31 (green ■, orange ■, and blue ■, respectively). Note the symmetric logarithmic scaling of the abscissa, where the linear portion is within $\pm 10\bar{t}$. Large black circles merely highlight the location of the states at the top of the band. *Right Panel:* PR averaged within energy bins and over disorder realizations. The plot is focused on the band near zero energy. Note the linear scale of the abscissa. Error bars for the average PR are 95% confidence intervals and smaller than the marker size where not seen.

is defined as

$$\text{PR} = \frac{1}{\sum_i |\psi(\bar{x}_i)|^4}, \quad (2.30)$$

where the sum is over all lattice sites, $\psi$ is a normalized wavefunction and $\bar{x}_i$ is the position of lattice site $i$. Typically, $\text{PR} \propto N^\beta$ [124–126]. This scaling behaviour gives information on the nature of the eigenstates. In particular, if $\beta = 0$, the eigenstates are localized, if $\beta = 3$, delocalized and, if $0 < \beta < 3$, the states can be labelled as extended non-ergodic [124–126]. Formally, extended non-ergodic states are states where $\text{PR}^{-1} \not\to \langle \text{PR}^{-1} \rangle$ as $N \to \infty$, where $\langle \cdots \rangle$ indicates the average over disorder realizations. Informally and in the context of this thesis, extended non-ergodic states span the system, but do not fill the entirety of the accessible space. Extended non-ergodic states are typically multi-fractal, though this property is definitively identified by examining the generalized participation ratios [125–127]. We focus on the scaling behaviour of the PR (2.30), rather than the generalized participation ratios, as the limited system sizes accessible when obtaining the entire spectrum via exact diagonalization makes a full quantitative analysis difficult.
We analyze the scaling behaviour of the PR of the eigenstates via two approaches. The first obtains a quantitative estimate of $\beta(E)$ by fitting the above scaling equation to the behaviour of $\langle \text{PR} \rangle$, where $\langle \text{PR} \rangle$ is the PR averaged within a particular energy window and over multiple disorder realizations. The second approach obtains a qualitative understanding of the system by examining the scaling behaviour of the distributions of both PR and $\text{PR}/N^3$ at all energies. Three cases are discerned:

(i) If the PR does not change with the system size, but $\text{PR}/N^3$ does, then $\beta = 0$.

(ii) If the PR scales with the system size, but $\text{PR}/N^3$ does not, then $\beta = 3$.

(iii) If both PR and $\text{PR}/N^3$ change with the system size, then $0 < \beta < 3$.

No other cases arise as $\beta \in [0, 3]$.

We note that the spectra of the systems we examine have an interesting dual-scale behaviour: there is a large density of states near zero energy, but also many states at energies very far away from zero. Thus, the spectra typically have one or more “wings”. To capture the behaviour of the PR across all of these energy scales, a symmetric logarithmic scale is used for the abscissa, where the portion of the axis near zero is scaled linearly, while the portions far from zero are scaled logarithmically. The size of the linear portion is denoted on the abscissa of the graph and in the captions.

**Scaling of Participation Ratios when $\alpha = 0$: Isotropic Hopping**

Figure 2.4 (left panel) shows the PR of every eigenstate, for multiple disorder realizations, plotted against the energy of the eigenstate for $W = 16.5$. The band at low energy is of width $16.5\tilde{t}$ and arises from the $(N^3 - 1)$-fold degenerate ground state of the ideal lattice. The points at the top of the spectrum arise from the $|000_k\rangle$ state of the ideal system. Figure 2.4 (right panel) shows the average PR as a function of energy for the band at low energies. There is clearly no scaling of $\langle \text{PR} \rangle$ with the system size. The fit to the $\langle \text{PR} \rangle$ of the high energy $|000_k\rangle$ state shows the scaling exponent to be 3. Identical behaviour has been observed for $W = 1$ and 35.

Evidently, the band at low energy is composed of localized states, while the $|000_k\rangle$ state remains delocalized, even in the presence of strong disorder. Such behaviour is easily understood from the energy level structure of the ideal system (described in Section 2.5.2). The extreme degeneracy of the ground state allows any small amount of disorder to strongly couple the states, leading to localization [29]. The $|000_k\rangle$ state
remains delocalized, however, as the shielding gap is significantly larger than the disorder strength, minimizing coupling. Given that the shielding gap diverges with the system size, the $|000\rangle$ state must remain delocalized in the thermodynamic limit for any finite disorder strength, while the low energy states remain localized. The same behaviour is observed for particles with isotropic uniform hopping on the $d$-simplex [29].

Given that cooperative shielding allows the system to be described by an effective short-range Hamiltonian [30], one could expect to see the localization transition near $W = 16.5$, the critical disorder strength of the Anderson transition for nearest-neighbour hopping in 3D lattices. Indeed, the dynamics of the system studied by Celardo et al. [30] is governed by an effective nearest-neighbour hopping Hamiltonian. However, they explicitly include a nearest-neighbour hopping term in their Hamiltonian, in addition to the long-range term. As it would be inconsistent with our generalization of the dipolar hopping amplitude to arbitrary $\alpha$, we do not include an additional nearest-neighbour hopping term in our Hamiltonian (2.19). Thus, the ground state of our system when $\alpha = 0$ is extremely degenerate (essentially flat-band) and localizes given any finite disorder strength.

Scaling of Participation Ratios when $\alpha = 0$: Anisotropic Hopping

The upper panels of Figure 2.5 show the PR vs. energy for $W = 1$ and 16.5 (left and right, respectively), while the lower panels show the averaged PR at the same disorder strengths. One can clearly see scaling with the system size for both disorder strengths. Given that the bandwidth of the system is a function of the system size, the scaling behaviour cannot be quantitatively determined for all band energies. Clearly acceptable choices of energies for comparison include the top and bottom states of the band and the states near the centre. The wing regions, however, do not overlap well enough for a quantitative analysis. Fits to the scaling equation reveal that the top and bottom states at both disorders scale as $N^3$. The states in the energy region $[-1, 1]$ for the system with $W = 1$ also scale as $N^3$. Interestingly, the states in the centre portion of the system with $W = 16.5$ have energy-dependent scaling exponents that vary from 2.6 to 2.8. We also examine the system $W = 35$, but find it qualitatively identical to the $W = 16.5$ case: the scaling exponents of the top and bottom states are 3, while those of the states near the band centre vary from 2.4 to 2.7.

We note that the errors on the scaling exponents can be as large as 0.2. We obtain this upper bound on
Anisotropic, $\alpha = 0$

$W = 1$

$W = 16.5$

Figure 2.5: Participation ratio (PR) as a function of energy for the eigenstate of a 3D system with anisotropic hopping; $W = 1$ and 16.5; $p = 1$; $\alpha = 0$; and $N = 11, 21, \text{ and } 31$ (green ■, orange ■, and blue ■, respectively). The number of disorders included ranges from 50 to 2350, as required to obtain sufficiently small error bars. Note the symmetric logarithmic scaling of the abscissa, where the linear portion is within $\pm 1 \tilde{t}$ or within $\pm 16 \tilde{t}$. Top Panels: Plot of the PR of each eigenstate vs its eigenvalue for several disorder realizations. Lower Panels: PR averaged within energy bins and over disorder realizations for $W = 1$ and 16.5. Error bars for the average PR are 95% confidence intervals and smaller than the marker size where not seen.

the error from one fit to the scaling equation that resulted in a value of 3.2, while $\beta$ cannot be larger than 3. Given that we only have three system sizes that span a relatively small range of sizes, our scaling exponents have low, but not negligible, precision. Qualitatively, an examination of $\text{PR}/N^3$ (not shown) shows that the wings, for both disorder strengths, scale as $N^3$. The central states also scale as $N^3$ for $W = 1$, while they scale with $0 < \beta < 3$ for the central states at $W = 16.5$. These qualitative results support the accuracy of the
quantitative results discussed above, despite their low precision.

The above results show that the states in the wings are delocalized, even for strong disorder. Furthermore, the states near the centre of the band become extended non-ergodic for sufficient disorder strength. We also see that there are no localized states in this system, in great contrast to the isotropic case.

\[ \alpha = 1 \]

**Isotropic**

**Anisotropic**

### Figure 2.6: Top Panels
Plot of the Participation ratio (PR) of each eigenstate vs its eigenvalue for several disorder realizations. **Lower Panels:** Plot of the PR/N^3 of each eigenstate vs its eigenvalue for several disorder realizations. Plots are for a 3D system with isotropic and anisotropic hopping; \( W = 16.5; p = 1; \alpha = 1; \) and \( N = 11, 21, \) and \( 31 \) (green ■, orange ■, and blue ■, respectively). The number of disorders included ranges from 100 to 350, as required to obtain well-formed distributions. Note the symmetric logarithmic scaling of the abscissa, where the linear portion is within \( \pm 10\tilde{t} \) or within \( \pm 16\tilde{t} \).
Scaling of Participation Ratios when $\alpha = 1, 3$

We examine the scaling behaviour of the PR for isotropic and anisotropic hopping with $\alpha = 1$ and $\alpha = 3$ by examining the qualitative scaling behaviour of PR and $\text{PR}/N^3$. We do not examine the quantitative behaviour as the limited system sizes accessible causes fits to the scaling equation to be even less reliable than when $\alpha = 0$. Figure 2.6 demonstrates that, for $\alpha = 1$ and $W = 16.5$, there is a single wing for the isotropic case, while there are two for the anisotropic case. These wings contain PR values that increase with the system size, while the $\text{PR}/N^3$ values are constant: the values scale as $N^3$ and the states are delocalized.

In the isotropic case near zero energy, however, the behaviour of the PR is not as clear. A portion of the PR values in this energy range does scale with the system size, but the distribution asymmetrically increases in width instead of shifting vertically. This obfuscates the interpretation. However, given that the $\text{PR}/N^3$ values all reduce in magnitude as the system size increases, it can be concluded that none of these central states are delocalized. Whether the states are extended non-ergodic or localized is not clear, however.

In the anisotropic case with $\alpha = 1$, the states near zero energy are clearly extended non-ergodic. This is seen as both the PR and the $\text{PR}/N^3$ distributions shift vertically as the system size increases, in the appropriate directions. Additionally, note that the rate of scaling with the system size near zero energy is different between the isotropic and anisotropic cases. The isotropic case has PR distributions that scale slowly, while the $\text{PR}/N^3$ distributions scale quickly. The reverse is true for the anisotropic case. We can thus conclude that the isotropic hopping scaling exponent is smaller than the anisotropic hopping scaling exponent for the states near zero energy.

Figure 2.7 demonstrates that, for $\alpha = 3$ and $W = 16.5$, there is a single wing for the isotropic case, while there are none for the anisotropic case. Also note that the bandwidths are much smaller, allowing for a linear scaling of the x-axis. The behaviour of the isotropic case is similar to that seen for $\alpha = 1$ and points to the same conclusion: the wing contains delocalized states, while the energy region near zero contains states that are not delocalized.

At this level of disorder, the spectrum of the anisotropic case is dominated by the disorder, producing a rather uniform distribution without wings. Interestingly, the scaling behaviour is similar to the near-zero-energy portion of the isotropic case: the PR distributions widen asymmetrically but do not shift in location as the size increases, while the $\text{PR}/N^3$ distributions shift downward as the system size increases. There are
thus no delocalized states, though there may be a combination of localized and extended non-ergodic states. This is not necessarily in contrast with the prior work of Levitov [25, 26, 113] as the extended non-ergodic states still span the system, allowing weak transport.

The above analysis of the participation ratios for various values of $\alpha$ and for different isotropies shows that single-particle systems with long-range hopping display a rich landscape of states, from delocalized to extended non-ergodic to localized. We expect that the extended non-ergodic states we typically find near zero energy in this work are likely multi-fractal. To find and quantify the multi-fractal dimensions and multi-fractal spectra [125] of these states, however, will require more sophisticated numerical or analytical techniques capable of accessing larger system sizes. There are algorithms that can reach system sizes of $N \approx 300$ for targeted energy ranges [121, 122], but it is not clear whether even these would be enough to properly study the scaling behaviour of the eigenstates.

### 2.5.4 Energy Level Statistics

Given the ambiguity observed in the scaling behaviour of the participation ratios for the physically important case of $\alpha = 3$, we now examine the system for both isotropic (2.22) and anisotropic (2.23) hopping and open boundary conditions from an energy level statistics point of view.

We use the mean level spacing ratio, defined by Oganesyan and Huse [128] as

$$\langle r \rangle = \langle \min(r_n', \frac{1}{r_n'}) \rangle$$  \hspace{1cm} (2.31)

where

$$r_n' = \frac{E_{n+1} - E_n}{E_n - E_{n-1}},$$

$E_n$ is the $n$th eigenvalue, sorted in order of increasing magnitude and $\langle \cdot \rangle$ denotes the average over all eigenvalues and over many instances of disorder. For a system whose energy level spacing distribution is Poissonian, $\langle r \rangle = 2 \ln 2 - 1 \approx 0.38629$. If, however, the system belongs to the Gaussian Orthogonal Ensemble (GOE), $\langle r \rangle = 4 - 2\sqrt{3} \approx 0.53590$ [129]. As discussed in Refs. [128, 130], Poissonian level statistics are characteristic of localization and GOE statistics are characteristic of diffusion.
Figure 2.7: *Top Panels*: Plot of the Participation ratio (PR) of each eigenstate vs its eigenvalue for several disorder realizations. *Lower Panels*: Plot of the PR/$N^3$ of each eigenstate vs its eigenvalue for several disorder realizations. Plots are for a 3D system with isotropic and anisotropic hopping; $W = 16.5$; $p = 1$; $\alpha = 3$; and $N = 11, 21, \text{ and } 31$ (green ■, orange ■, and blue ■, respectively). The number of disorders included ranges from 99 to 350, as required to obtain well-formed distributions. Note the linear scaling of the abscissa.

Figure 2.8 presents $\langle r \rangle$ vs $W$ for the isotropic and dipolar Hamiltonians with $\alpha = 3$ and open boundary conditions. It can be seen in both cases that $\langle r \rangle$ drops as $W$ increases, though not completely to the Poissonian value. However, in the isotropic case, $\langle r \rangle$ drops with the increasing lattice size towards the Poissonian limit. As illustrated in Figure 2.8 (left panel), we observe a significant change with the system size. In contrast, in the anisotropic case, the curves display no discernible scaling with the system size, even in the case of extremely strong disorder. There is thus no indication that the Poissonian limit will be attained.

It is noteworthy that the scaling behaviour in the isotropic case begins at $W \approx 15$, which is the expected
location, given cooperative shielding. Cooperative shielding causes the Hamiltonian to be effectively short-ranged, such as the tight-binding model. For the tight-binding model, the diffusion-to-localization transition occurs at \( W \approx 16.5 \) [131]. At disorder strengths between 5 and 15, the curves in the isotropic case collapse onto one another, preventing the possibility of scaling to the infinite size limit. At \( W = 1 \), it is clear that the system is approaching the GOE value as the system size increases.

Given that the scaling behaviour of the PRs reveals two clear regimes in the isotropic case, one where the states are delocalized and one where they are localized or extended non-ergodic (see Figure 2.7), it is useful to focus on the region near zero energy. Figure 2.9 (left panel) demonstrates that the scaling behaviour observed previously is enhanced when the energy level spacing ratio is averaged only over the energy region \(-\bar{t} \leq E \leq \bar{t}\). Figure 2.9 (right panel) demonstrates that the same scaling behaviour also occurs when \( \alpha = 1 \). These results provide numerical evidence for localization in the infinite size limit for the case with isotropic long-range hopping and \( \alpha = 1 \) and 3. This, combined with the likely presence of cooperative shielding and the observed scaling behaviour of the participation ratios, strongly suggests that some, to many, of the states at low energy are localized.

**Full Spectrum**

![Figure 2.8](image-url)

**Figure 2.8:** Mean energy level spacing ratio \( \langle r \rangle \) as a function of disorder strength \( W = \frac{\omega_0}{t_{\text{max}}} \) for the isotropic and anisotropic hopping Hamiltonians with open boundary conditions. The hopping range exponent is \( \alpha = 3 \) and the filling fraction is \( p = 1 \) in both cases. \( N \) denotes the lattice side-length (\( N^3 \) sites). The horizontal dashed lines denote the values of \( \langle r \rangle \) for the Poisson distribution (\( \sim 0.38629 \)) and the GOE (\( \sim 0.53590 \)). The error bars are 95% confidence intervals based on 96 disorders and are smaller than the marker size where not seen.
Near Zero Energy

Figure 2.9: Left panel: Same as Figure 2.8 (left panel), but with the mean energy level spacing ratio $\langle r \rangle$ including only the eigenvalues in the energy range $-\tilde{t} \leq E \leq \tilde{t}$. The results are averaged over 1071 disorders for $N^3 = 11^3$ and $21^3$, and 96 disorders for $N^3 = 31^3$. Right panel: Same as left panel, but for $\alpha = 1$. The results are averaged over 1100 disorders for $N^3 = 11^3$ and $21^3$, and 100 disorders for $N^3 = 31^3$.

To connect with realistic experimental conditions, we also examine the effect of a diluted lattice. This is relevant for experiments with polar molecules in an optical lattice [104, 105] or for amorphous molecular solids [100]. A diagram of the mean level spacing ratio, $\langle r \rangle$, as a function of both $p$ and $W$ for both isotropic and dipolar hopping with $\alpha = 3$ and open boundary conditions is presented in Figure 2.10.

While the two plots are qualitatively similar, $\langle r \rangle$ falls off more quickly with the disorder strength and filling fraction in the isotropic case. This can be seen by comparing the areas of the two diagrams where $\langle r \rangle > 0.49$. The value of $\langle r \rangle$ remains high along the $p$-axis, which is in agreement with the results of Deng et al. [124], who studied the case of dipolar hopping in a diluted lattice.

Figure 2.10 (right panel) shows that in experiments with polar molecules in optical lattices one should not expect to see localization of single rotational excitations due to the dilution alone, if it exists at all. Additional on-site disorder, such as from an optical speckle potential, is required. Exploring this phase diagram is within reach of current experiments. For example, Figure 2.10 (right panel) shows that for molecules on an optical lattice with $N \approx 30$ and a lattice population of 30% [105] (a typical filling fraction aimed at in current experiments), the region where $\langle r \rangle$ drops significantly can be explored by varying the
optical speckle potential from below to above $W = 5$. Scaling behaviour of $\langle r \rangle$ can also be investigated as optical lattices can have $N$ up to 60 (a typical size of an optical lattice).

**Figure 2.10:** Mean energy level spacing ratio $\langle r \rangle$ as a function of disorder strength $W = \frac{\omega}{t_{\text{max}}}$ and filling fraction $p$ for the isotropic and anisotropic hopping Hamiltonians with open boundary conditions. The hopping range exponent is $\alpha = 3$ and the lattice size is $N^3 = 31^3$ in both cases. The colour varies from the value of $\langle r \rangle$ for the Poisson distribution ($\sim 0.38629$) to that for the GOE ($\sim 0.53590$). The white circles denote the values obtained via exact diagonalization of the Hamiltonian; piecewise cubic interpolation is used for the rest of the plot. The black lines indicate where $\langle r \rangle = 0.49$. The results are averaged over 96 realizations of disorder, giving 95% confidence intervals that are at most $\pm 0.002$. The green data points mark the location of the Anderson transition for the 3D tight-binding model in the infinite size limit, as determined by Root and Skinner [131].

## 2.6 Conclusion

We have illustrated that the cooperative shielding discovered for 1D lattices with long-range hopping [30, 119] also occurs in 3D lattice models with isotropic long-range hopping. This suggests the possibility of Anderson localization in 3D systems with long-range hopping (i.e. $\alpha \leq 3$). Given the form of the results in three dimensions, the same is likely to be true of a 2D system with isotropic long-range hopping ($\alpha \leq 2$). We have also presented evidence indicating the lack of cooperative shielding in models with anisotropic hopping. The specific form of the anisotropy we consider here has the angular dependence of the dipole-dipole interaction between dipoles aligned along one of the principal axes of the lattice.

We have demonstrated that there are fundamental differences between disordered lattice systems with isotropic long-range hopping and those with anisotropic long-range hopping, particularly with regards to
the impact of disorder. In addition to the difference in cooperative shielding, we have shown that the energy level structures of systems with isotropic hopping are qualitatively and quantitatively different from those of systems with dipolar anisotropic hopping (whether with periodic or open boundary conditions).

We have used the scaling behaviour of the eigenstate participation ratios to demonstrate the presence of localized states in the isotropic case with uniform hopping ($\alpha = 0$) and the absence of localized states in the anisotropic case with uniform hopping ($\alpha = 0$). We have also demonstrated that energy level statistics support the presence of localized states in 3D cubic lattices with isotropic hopping that varies as $r^{-3}$, while they are inconclusive for systems with the anisotropic dipolar hopping considered here. We have further shown that energy level statistics indicate the presence of localized states for Coulomb-like ($\alpha = 1$) isotropic hopping, while the scaling behaviour of the eigenstate participation ratios illustrates the absence of localized states when $\alpha = 1$ and the hopping is of the anisotropic form (2.23).

We have shown that the localization properties (or lack thereof) found for systems with dipolar hopping [25, 26, 113, 114] cannot be assumed to apply to systems with isotropic hopping, in accordance with the suggestion of Burin and Maksimov [27]. More generally, we have demonstrated that the presence or absence of delocalized, extended non-ergodic, or localized states depends on not only the range of the hopping amplitudes, but also their anisotropy.

### 2.7 Future Directions

Future studies that can access significantly larger system sizes are required to fully characterize the shape of the localized states, to determine whether the extended non-ergodic states are truly multi-fractal, and to determine how the multi-fractal dimensions and spectra change with hopping range and isotropy. Future theoretical and experimental studies can also probe the dependence of localization on the direction of the quantization axis (as defined by the direction of an external field) relative to the principal axes of the underlying cubic lattice. Indeed, it would be interesting to investigate whether cooperative shielding in 3D lattices can be induced by tilting the quantization axis. Additionally, it would be interesting to explore the effect of the lattice geometry on cooperative shielding. Finally, this work opens up the question of how different types of anisotropy, such as quadrupolar-like, can impact localization behaviour.

Furthermore, while some of our results indicate that cooperative shielding likely also exists in two
dimensional systems, it is important to directly confirm this and to determine the impact of anisotropy on cooperative shielding and localization in this case. Examining the impact of different lattice geometries and of stacked two dimensional lattices on cooperative shielding and localization would also be interesting. This may give insights into the localization behaviour (or lack thereof) of excitons in stacked monolayers [132].

Additionally, it is not currently known why anisotropic hopping seems to preclude cooperative shielding. It may be because the average of the hopping amplitudes over angle is zero, which may indicate disruption of the constructive interference required to induce localization. That is, the two paths on a loop may no longer interfere in the same manner as in the isotropic case. This possibility needs to be explored further, potentially by taking a diagrammatic approach or simply examining the impact of a hopping amplitude anisotropy that does not average to zero on cooperative shielding and localization.

Also, a recent paper published in 2019 by Nosov, Khaymovich and Kravtsov [133] discusses the notion of correlation-induced localization where increasing correlation in the long-range hopping amplitudes decreases the disorder strength needed to localize the system, to the point where localized states exist for any finite disorder strength. Indeed, it may be that changing the anisotropy of the hopping amplitudes is effectively changing the correlations present in the hopping amplitudes and thus the constructive interference pattern required for localization. This is a direction worth pursuing.

This now concludes my discussion on Anderson localization and how it relates to many-body molecular systems. Namely, ensembles of molecules can have excitations that effectively behave as particles with long-range hopping. The nature of this long-range hopping can greatly impact the localization properties of these quasi-particles and thus energy transport within the molecular ensemble.


2.8 **Summary**

In this section, I list the key results presented in Chapter 2. More detailed results can be found in Section 2.6.

- The presence or absence of delocalized, extended non-ergodic, or localized states depends on both the range and the *anisotropy* of the hopping amplitudes.

- Cooperative shielding extends to three-dimensional systems, if the hopping amplitudes are isotropic (Section 2.5.2).

- Cooperative shielding is also likely to exist in two dimensional systems, if the hopping amplitudes are isotropic (Section 2.5.2).

- Cooperative shielding appears absent if the hopping is anisotropic with the angular form of the dipole-dipole interaction between dipoles aligned along one of the principal axes of the cubic lattice (Section 2.5.2).

- Isotropic hopping on a disordered cubic lattice permits the formation of localized states, even if the hopping is long-ranged ($\alpha \leq 3$), as evidenced by the scaling of eigenstate participation ratios and of energy level statistics (Sections 2.5.3 and 2.5.4).

- The anisotropic hopping case we consider appears to prevent the formation of localized states when the hopping is long-ranged ($\alpha < 3$), as evidenced by the scaling of participation ratios and apparent lack of cooperative shielding (Sections 2.5.2 and 2.5.3).

- The evidence provided by the scaling of participation ratios and of energy level statistics is inconclusive regarding whether the states are localized or extended non-ergodic for the anisotropic case when the hopping is critically long-ranged ($\alpha = 3$). Extended non-ergodic states would be consistent with the literature. (Sections 2.5.3 and 2.5.4)
Chapter 3

Molecular hyperfine interferometry

3.1 Introduction

This chapter discusses our proposal to use closed shell molecules as probes in surface spin echo experiments and the fully quantum framework we have developed to model these experiments. Particularly, I discuss $^3$He spin echo experiments and their extension to molecular hyperfine interferometry experiments at the start of Section 3.2. In Section 3.3, I describe a generic molecular hyperfine interferometry experiment. I then discuss, in Section 3.4, the molecular state after the state-selecting magnetic lens. In Section 3.5, we time-evolve the molecular state and integrate the result over the length of the detection window to obtain the relationship between the system eigenstates and the detector current. To obtain the system eigenstates, we derive and apply, in Section 3.6, a transfer matrix formalism that includes internal degrees of freedom. I also discuss the rotation and scattering transfer matrices used to account for the apparatus geometry and the molecule-surface interaction, respectively. In Section 3.7, we demonstrate the application of this theoretical framework to the case of $\text{H}_2$, illustrate the need to integrate over the velocity distribution, illustrate the sensitivity of the calculated signal to various features of the scattering transfer matrix and perform a preliminary comparison with experiment. We compare the method of the present chapter to the semi-classical method discussed by Godsi et al. [44] in Section 3.8. I conclude the chapter in Section 3.9 and give a bulleted summary of the chapter results in Section 3.10.
3.2 From Helium-3 Spin Echo to Molecular Hyperfine Interferometry

Understanding the dynamics of molecule-surface interactions is of central importance for many applications in science and technology. One of the most powerful experimental techniques for studying surface dynamics is $^3$He spin echo (HeSE) [134, 135]. In a HeSE experiment, a beam of spin-polarized $^3$He atoms is passed through two regions of strong magnetic field before and after the $^3$He atoms meet a sample surface. The dependence of the spin-echo signal on one or both of the magnetic fields encodes information about the atom-surface scattering event. HeSE experiments and related helium atom scattering experiments fulfil a unique role amongst other experimental techniques for studying surfaces as they can probe both fast time-scales (pico- to nanosecond) and small distances (angstroms to tens of nanometers) and they are explicitly surface sensitive [134].

HeSE experiments [134, 135] have been used to study surface morphology [136], molecular and atomic surface diffusion [33–39], inter-adsorbate forces [37, 137], phonon dispersions [40, 41], phason dispersions [42], structures and phase transitions of ionic liquids [43] and friction between adsorbates and surfaces [138, 139]. HeSE experiments have improved our understanding of potential energy surfaces [140–142] and surface-adsorbate interactions [143–146] and are frequently combined with microscopic calculations to both test theory and gain insight into surface-adsorbate interactions [147, 148]. However, the interactions between $^3$He and surfaces or their adsorbates are typically isotropic and relatively weak.

One can potentially obtain further information about surface dynamics by using molecules instead of $^3$He atoms in a surface spin-echo experiment. Molecules offer rotational degrees of freedom, which could be exploited to gain new insights into surface dynamics. Indeed, Godsi et al. [44] have recently shown that ortho-hydrogen ($o$H$_2$) can be used as a sensitive probe of surface morphology: their experiment was able to discern how the interaction between an $o$H$_2$ molecule and a Cu(115) surface depends on the orientation of the rotational plane of the hydrogen molecule relative to the surface. In addition, one could exploit the transfer of rotational energy from the probe molecules to surface adsorbates (or vice versa) in order to study the relative effects of the rotational and translational motion on the dynamics of the adsorbates. However, the increased complexity of molecules (compared to $^3$He atoms) makes the analysis of the spin-echo experiments complicated and requires one to account for the interplay of the translational, nuclear spin and rotational degrees of freedom in strong magnetic fields of differing orientations, in addition to the
molecule-surface scattering event.

In this chapter, I present a general and efficient theoretical framework that can be used to analyze surface interferometry experiments that use closed shell molecules to study static surfaces. The large manifold of molecular states and magnetic field-induced couplings between internal states preclude a simple semi-classical description for these molecular hyperfine interferometry (MHI) experiments, in contrast to HeSE experiments. We develop a fully quantum mechanical model of surface-sensitive MHI experiments by deriving a one-dimensional transfer matrix method that includes the internal hyperfine degrees of freedom of the probe molecules. This allows us to efficiently solve the time-independent Schrödinger equation with a multiplication of matrices. We account for the experimental geometry with rotation matrices and describe the molecule-surface interaction with a scattering transfer matrix (a transformed version of the standard scattering matrix).

We apply our method to an $\text{oH}_2$ hyperfine interferometry experiment and illustrate the importance of integrating over the velocity distribution of the molecules. We further show that information about the scattering matrix elements is encoded in the experimental signal. In particular, we demonstrate that the experimental signal is sensitive to the magnitude and phase of the diagonal elements of the scattering transfer matrix. We also show that the signal is sensitive to inelastic scattering events that change the projection quantum numbers of the molecular hyperfine states, described by scattering transfer matrices with non-zero diagonal and off-diagonal matrix elements. This sets the stage for determining, in part or in whole, the scattering transfer matrix elements of a particular molecule-surface interaction by comparing the computed and experimentally-measured signals. Finally, through comparison with the semi-classical method discussed by Godsi et al. [44], we demonstrate that our method opens a path towards an even more powerful framework that can analyze surface interferometry experiments using closed shell molecules to study dynamic, instead of static, surfaces.

3.3 Description of a Molecular Hyperfine Interferometry Experiment

A surface-sensitive molecular hyperfine interferometer uses a beam of molecules to probe various surface properties. To do this, a set of magnetic fields are used to simultaneously manipulate the internal hyperfine states of the probe molecules and create a spatial superposition of molecular wavepackets. These wavepack-
ets sequentially impact the sample surface and scatter in all directions. A second set of magnetic fields collects the molecules scattered in a narrow solid angle. This second set of magnetic fields further manipulates the molecular wavepackets, partially recombining them and allowing for molecular self-interference. Wavepackets with particular hyperfine states are then passed into a detector. A schematic of the experiment is depicted in Figure 3.1. We now discuss the different stages of the experiment in more detail.

The beam source must produce a continuous (or pulsed) beam of molecules with a sufficiently narrow velocity profile, mean velocity suitable for a particular experiment, sufficiently high flux, and a density low enough to ensure that the molecules are non-interacting. One current apparatus [44] uses a supersonic expansion to produce such a beam. One can also envision experiments with slow molecular beams produced by extraction (sometimes with hydrodynamic enhancement) from a buffer-gas cooled cell [149] or with molecular beams controlled by electric-field [150] or magnetic-field deceleration [151]. Deceleration provides control over the mean velocity and narrows the velocity spread [152], which could be exploited for novel interferometry-based applications.

The experiment selects molecules in particular hyperfine states by employing a magnetic lens whose magnetic field has a gradient in the radial direction. A cylindrically-symmetric field gradient is used to ensure sufficient molecular flux. The lens focusses molecules with low-field seeking states and defocusses molecules with high-field seeking states, allowing for purification of the molecular beam. After the lens, a unique quantization axis for the internal states is developed by using an auxiliary field that adiabatically rotates all magnetic moments until they lie along a single direction perpendicular to the beam path. The end of this auxiliary field is a strong dipolar field aligned along the $z$ direction that defines the quantization axis. Hexapole magnets can be used as a magnetic lens as their magnetic field gradients are sufficiently cylindrically symmetric [44, 153, 154]. More details about the internal states of the molecules immediately after the magnetic lens can be found in Section 3.4.

Solenoids whose magnetic fields are parallel to the beam propagation path are used to manipulate the molecular hyperfine states. These solenoids are labelled as the control fields in Figure 3.1. Arbitrary magnetic field profiles can be obtained by changing the solenoid winding patterns and/or using multiple successive solenoids.

The hyperfine states of a molecule change energy as the molecule enters a magnetic field. These changes
to the hyperfine energy levels cause simultaneous changes in the molecular momenta, as the total energy is conserved. That is, when molecules enter a solenoid, molecules in low-field seeking states slow down and those in high-field seeking states speed up. Furthermore, because the direction of the magnetic field in a control field is not along the $z$ axis, the molecules are in a superposition of hyperfine states, with respect to the quantization axis defined by the magnetic field. Thus, the differences in momenta cause the different components of each molecular wavepacket to spatially separate as the wavepacket traverses the solenoid. Upon exiting the solenoid, the components of each wavepacket return to their original momenta, but remain spatially separated. That is, each wavepacket is now in an extended spatial superposition.

Each of these spatially separated wavepacket components comprise a superposition of the field-free hyperfine states. The exact superpositions of each wavepacket component, as well as the spatial separations between the components, depend on the magnetic field profile of the first branch. Each of the wavepacket components sequentially impacts the sample surface and scatters in all directions. However, the experiment only captures those molecules that pass through a particular solid angle. While a current experiment [44] fixes the angle between the two branches, one can in principle explore many different scattering geometries by varying both the angle between the two branches of the apparatus and the orientation of the sample.

After scattering, the collected molecules enter another set of control fields in the second branch of the apparatus. The hyperfine states again change in energy and momenta. In a helium-3 spin echo experiment, if the second magnetic field profile is identical but opposite in direction to the magnetic field profile of the first branch, the spatially separated wavepacket components realign (to first order) as they traverse the magnetic field(s), producing a spin echo. This allows the wavepacket components to interfere with each other. Interestingly, it has recently been shown [155] that echoes are also produced when the device operates with the fields in the same direction. With an arbitrary hyperfine Hamiltonian, such a realignment is only partial, though still useful. Experiments can be performed that explore either this spin-echo region or different relationships between the two magnetic field profiles, which may allow for a variety of insights about the sample surface. For example, the two field profiles can be different or the field magnitudes can be varied simultaneously, keeping $B_1 = -B_2$. These different regimes of operation may produce different echoes, which can be collectively analyzed to provide more insight into molecule-surface interactions.

Additionally, as the spatially separated wavepacket components hit the surface sequentially, rather than
simultaneously, any temporal changes in the surface that are on the time scale of the impact-time separation can differentially impact the phases of each wavepacket component. This may result in different interference patterns or even loss of coherence. This loss of coherence is the basis for the sensitivity of HeSE measurements to surface motion [41]. Here, as in the recent experiment by Godsi et al. [44], we focus on surfaces whose dynamics are either much faster or much slower then the molecule-surface or wavepacket-surface interaction times. Note, however, that the current framework is suitable for extension to interaction regimes where the surface dynamics are comparable to these time scales.

After leaving the last solenoid of the second branch, the wavepackets pass through another auxiliary field that begins with a strong dipolar field in the $z'$ direction. The auxiliary field then adiabatically connects magnetic moments aligned along the quantization axis to the radial direction of the final hexapole lens. This hexapole lens then focusses wavepackets with low-field seeking hyperfine states into the ionization detector and defocusses the rest. Finally, the ionization detector produces a current that is proportional to the molecular flux into the detector port. We describe how to calculate the molecular flux that enters the detector port in Section 3.5 and the related transfer matrix formalism in Section 3.6.

Analyzing the detector current as a function of the magnetic field profiles, the apparatus geometry, and the sample orientation can provide information about the interaction of the molecules with the sample surface. We discuss one possible analysis scheme in Section 3.7.

3.3.1 Molecular Hyperfine Hamiltonian

In principle, the only requirement for a molecular species to be suitable for molecular hyperfine interferometry is that the molecule have internal degrees of freedom whose energies are magnetic-field dependent. Such a requirement could be fulfilled by the presence of a nuclear spin, a rotational magnetic moment, or even an electronic spin. In practice, however, if the energy dependence on the magnetic field is too weak relative to the kinetic energy, state selection and state manipulation is difficult. On the other hand, if the dependence is too strong, the molecules may be difficult to control. Given these restrictions, we deem molecules that have a closed shell and are in an electronic state with zero orbital angular momentum to be most suitable for molecular hyperfine interferometry. In this case, the dominant interactions induced by magnetic fields are due to the nuclear magnetic spins of the molecules.

The hyperfine states of such a closed shell molecule with zero orbital angular momentum arise from
Figure 3.1: A generic molecular hyperfine interferometer consists of a beam source (green), magnetic lenses (dark blue), auxiliary fields (light blue), control fields (purple), the sample (hatched rectangle) in an ultra-high vacuum chamber, and the detector (red). See Section 3.3 for more details on each component. The arrows and dashed line indicate the direction and path of the molecular beam, which is initially along the $+x$ direction and then along the $-x'$ direction after scattering. The two branches of the apparatus are separated by an angle $\theta$. $z$ and $z'$ denote the direction of the quantization axes before and after scattering, respectively. This definition of the quantization axes has been chosen to match the experiment by Godsi et al. [44] and to simplify rotating the quantization axes in the transfer matrix method. The $y$ and $y'$ axes are identical and point into the page.

coupling between the nuclear spin and the rotational degrees of freedom. Interactions of these hyperfine states with a magnetic field arise from the response of the nuclear and rotational magnetic moments to the external magnetic field. We assume that the hyperfine Hamiltonian, also referred to here as the Ramsey Hamiltonian [156], is of the following form:

$$\hat{H}^R(\vec{B}) = U \left( \hat{I}^2, \hat{J} \cdot \hat{J}, I, J \right) + V \left( \hat{I}^2, \hat{I} \cdot \vec{B}, I, \vec{B}^2 \right) + Q \left( \hat{J}^2, \hat{J} \cdot \vec{B}, J, \vec{B}^2 \right),$$

(3.1)

where $\vec{B}$ is the vector of the external magnetic field, assumed to be uniform across the molecule; $\hat{I}$ and $\hat{J}$ are the nuclear spin and rotational angular momentum operators, respectively; $I$ and $J$ are the nuclear spin and rotational angular momentum quantum numbers, respectively; $U$ contains all spin-rotational couplings (such as $\hat{I} \cdot \hat{J}$ or $\hat{I}^2 \hat{J}^2$); $V$ contains all interactions of the nuclear spins with the magnetic field (such as $\hat{I} \cdot \vec{B}$); and $Q$ contains all interactions of the rotational angular momentum with the magnetic field (such as $\hat{J} \cdot \vec{B}$). Both $V$ and $Q$ are assumed to be proportional to positive powers of $|\vec{B}|$.

At large magnetic fields, $V$ and $Q$ dominate, making the eigenbasis $|I m_I J m_J\rangle$, where $m_{I/J}$ is the projec-
tion of the angular momentum $\mathbf{I}/\mathbf{J}$ onto the external magnetic field direction. At zero field, $\hat{H}^R$ is diagonalized by $|IJFM\rangle$, where $\hat{F} = \hat{I} + \hat{J}$ is the total angular momentum operator and $M$ is the projection of $\hat{F}$ onto a chosen quantization axis. At intermediate fields, the eigenbasis is a function of the magnetic field and can be represented as a superposition of either $|IJFM\rangle$ or $|Im_jm_j\rangle$ states. Note that $M$ is a good quantum number at all field strengths. We call an eigenstate of $\hat{H}^R$ a Ramsey state, which we denote as $|R\rangle$ and which has the energy $E_R$. The number of eigenstates of $\hat{H}^R$ is $N_R$, such that $1 \leq R \leq N_R$.

We treat the apparatus as a one dimensional system and account for the actual geometry by rotating the basis of the hyperfine states at the appropriate locations (see Section 3.6.2). The total Hamiltonian can thus be written as

$$\hat{H}(x) = \frac{\hat{p}^2}{2m} + \hat{H}^R(\mathbf{B}(x))$$

(3.2)

where $\hat{p}$ is the centre of mass momentum operator, $m$ is the molecular mass, and $x$ is the position of the molecule in the apparatus. The magnetic field $\mathbf{B}(x)$ is now spatially dependent, reflecting the magnetic field profiles of the two branches of the apparatus. In principle, the total Hamiltonian should incorporate molecule-surface interaction terms. However, we include these molecule-surface interactions through the use of a scattering transfer matrix (see Section 3.6.3). This allows for the coherent propagation of the molecular wavepacket through the apparatus to be determined analytically and for the molecule-surface interactions to be described phenomenologically.

The system eigenstates $|ER\rangle$ are defined by the total Hamiltonian (3.2) through $\hat{H} |ER\rangle = E |ER\rangle$. Note that the system eigenstate $|ER\rangle$ is $N_R$ degenerate and that any linear combination of these states with the same label $E$ is also an eigenstate of $\hat{H}$. This degeneracy occurs as, while the $N_R$ different Ramsey states may have different energies, the kinetic energy can always be selected to maintain the same total energy. For the sake of convenience, we choose the orthonormal basis to be that defined by $\hat{H}^R(\mathbf{B}(x)) |ER\rangle = E_R |ER\rangle$ for $x \leq 0^-$. The zero of $x$ is defined to be immediately after the magnetic lens, while $y^\pm \equiv \lim_{\delta \to y^\pm} \delta$. We use these limit definitions as we will deal with discontinuities in the magnetic field when working with the transfer matrix formalism (Section 3.6). As an example of the use of this notation, the statement that both one-sided limits are equal at the point $x$ (i.e. $\lim_{a \to x^-} f(a) = \lim_{b \to x^+} f(b)$) can be written as $f(x^-) = f(x^+)$. The above definition of $|ER\rangle$ produces, for all $x$, a unique labelling of the system eigenstate $|ER\rangle$ by
the total energy $E$ and the internal state $R$, where $R$ is a Ramsey state in the high magnetic field located immediately after the magnetic lens (i.e. at $x = 0^-$). Note that, because of this definition, $\hat{H}^R (\vec{B}(x)) |ER\rangle \neq E_R |ER\rangle$ for $x \geq 0^+$; that is, the system eigenstates are superpositions of the local Ramsey states for $x \geq 0^+$. This unique labelling of the system eigenstates is valid for all $x$ as the eigenstate wavefunctions have a well-defined phase relationship throughout the entire apparatus. See Section 3.6.1 for more details on the specifics of this phase relationship.

### 3.4 Impact of the Magnetic Lens on the Molecular States

The magnetic lenses are designed to focus molecules with certain hyperfine states either onto the sample or into the detector. The remaining molecules are either defocussed or insufficiently focussed and contribute significantly less to the experimental signal. Roughly, high-field seeking states are defocussed, some of the low-field seeking states are well focussed and the rest of the low-field seeking states are partially focussed. The actual proportions of each hyperfine state in the molecular beam must be measured or calculated from simulation. These magnetic lenses typically use large magnetic fields and large magnetic field gradients to perform this focussing [153, 154].

In general, magnetic lenses may take different forms, but we will consider lenses that have one key feature: the internal degrees of freedom of the outgoing molecular wavepackets are decohered in the high-magnetic field basis (i.e. $|Im_IJm_j\rangle$). More precisely, we assume that the wavepacket exiting the magnetic lens is a mixed state of the form:

$$
\rho_0 = \sum_{R_0} P_{R_0} |\Psi_{R_0k_0}\rangle \langle \Psi_{R_0k_0}| ,
$$

where

$$
|\Psi_{R_0k_0}\rangle = \int dr \psi_{R_0k_0}(r) |rR_0\rangle ;
$$

$\psi_{R_0k_0}(r) \equiv \langle r |\Psi_{R_0k_0}\rangle$ is the wavefunction of a molecule in state $|R_0\rangle$; $\rho_0$ is the initial (time $t = 0$) density matrix; $|rR_0\rangle \equiv |r\rangle |R_0\rangle$; $|r\rangle$ is an eigenstate of the position operator; $|R_0\rangle$ is an eigenstate of $\hat{H}^R (\vec{B}_{\text{lens}})$; $\vec{B}_{\text{lens}}$ is a high magnitude, $z$-aligned magnetic field; $k_0$ is the experimentally-determined mean wavenumber of the
wavepacket; and \( P_{R_0} \) is the probability that the hyperfine statevector of the molecule is \( |R_0\rangle \). Note that \( \rho_0 \) is diagonal in \( |R_0\rangle \) but not in \( |r\rangle \) (or \( |k\rangle \), the momentum basis). Also, \( \vec{B}_{\text{lens}} = \vec{B} (x = 0^-) \) corresponds to the final portion of the auxiliary field (i.e. a strong, \( z \) aligned, dipolar field), not the field inside the hexapole magnet itself (see Section 3.3).

That such a form of the wavepacket is valid follows from the work by Utz et al. [157]. The authors show that the two wavepackets arising from a spin–\( \frac{1}{2} \) particle passing through a Stern-Gerlach apparatus are quickly decohered with respect to one another, even before they separate spatially. That is, the quantum dynamics themselves cause decoherence between the spin degrees of freedom (but not the spatial); a measurement or coupling to an external bath is not required. This decoherence occurs as the large magnetic field gradients cause a rapid oscillation in the off-diagonal terms of the extended Wigner distribution. That is, the phase relationship between the spin-up and spin-down components oscillates heavily in both the position and momentum bases, destroying coherence.

Given that the magnetic lenses we consider act like a Stern-Gerlach apparatus for the molecular hyperfine states, it is reasonable to assume that the internal hyperfine degrees of freedom will also decohere. Thus, we need only determine the values of \( P_{R_0} \) for a specific magnetic lens. These can be found via semi-classical calculations [44, 158], may be measured experimentally [158] or may potentially be determined by solving the full 3D Schrödinger equation within the lens.

The mean velocity \( v_0 \) and velocity spread \( \sigma_v \) of the molecules in the molecular beam can be measured experimentally [44]. Both of these values are determined from the position and profile of scattering peaks obtained from the scattering of the probe molecules by appropriate sample surfaces [44]. We assume that the initial wavefunction of a molecule \( \psi_{R_0k_0}(r) \) is Gaussian and is characterized by \( k_0 \equiv mv_0/\hbar \) and \( \sigma_k \equiv m\sigma_v/\hbar \), where \( m \) is the mass of the molecule. More precisely,

\[
\psi_{R_0k_0}(r) = \int dk \frac{1}{(2\pi\sigma_k^2)^{\frac{3}{2}}} e^{-\frac{(k-k_0)^2}{4\sigma_k^2}} \frac{e^{ikr}}{\sqrt{2\pi}} = \sqrt{\sigma_k} \left( \frac{2}{\pi} \right)^{\frac{3}{4}} e^{ik_0r} e^{-r^2\sigma_k^2} \tag{3.5}
\]

where \( k_0^{R_0} \) is taken to be \( k_0 \). Though \( k_0^{R_0} \) may in fact depend slightly (on the order of ppm) on \( R_0 \), we show
later that the experimental signal is insensitive to small changes in $k_0^R$.

### 3.5 Wavepacket Propagation and Signal Calculation

The primary measured value of the experiment is a current that is proportional to the molecular flux entering the detector. This measured current is a function of the magnetic fields, the scattering geometry, and the surface properties. The molecular flux entering the detector can be calculated as the product of the molecular flux incident to the apparatus and the probability that a molecule entering the apparatus will successfully pass through the apparatus and be detected. It is this probability of detection $P_{\text{detection}}$ that is sensitive to the experimental parameters and surface properties. Note that the incident molecular flux could be either continuous or pulsed, as long as the density is low enough that the molecules can be considered non-interacting.

As the detector has a finite time-response, the probability of detection is given by

$$P_{\text{detection}} = \frac{1}{\tau} \int_{t_1}^{t_2} dt \langle \hat{C}(t) \rangle, \quad (3.6)$$

where $t_1$ and $t_2$ are the initial and final times of the detection window $\tau = t_2 - t_1$, and $\langle \hat{C}(t) \rangle$ is the expectation value of the detector measurement operator $\hat{C}$. This expectation value is given by

$$\langle \hat{C}(t) \rangle = \text{Tr} \left( \hat{\rho}(t) \hat{C} \right), \quad (3.7)$$

where $\hat{\rho}(t) \equiv \hat{U} \rho_0 \hat{U}^\dagger = \sum_{R_0} P_{R_0} |\Psi_{R_0k_0}(t)\rangle \langle \Psi_{R_0k_0}(t)|$ is the time evolved density matrix, $\hat{U} \equiv e^{-i\hat{H}t}$ is the time evolution operator, $\rho_0$ is the density matrix $(3.3)$ at $t = 0$, and $|\Psi_{R_0k_0}(t)\rangle \equiv \hat{U} |\Psi_{R_0k_0}\rangle$.

Given that the detector consists of a magnetic lens that focusses molecules with particular states into a measuring apparatus, such as an ionization detector [44], and that the internal degrees of freedom of these molecules are decohered by the second magnetic lens (see Section 3.4), we can model the detector with a diagonal operator

$$\hat{C} = \sum_{R_D} \int dx \; c_{R_D}(x) |xR_D\rangle \langle xR_D| \quad (3.8)$$

The matrix elements of $\hat{C}$ are the probabilities $c_{R_D}(x)$ of detecting, at position $x$, a molecule whose internal
state is a high-field eigenstate $|R_D\rangle$ of $\hat{H}^R$. Note that $c_{R_0}(x) = 0$ for $x < x_D$, the detector position.

Using the time evolution operator, we determine the time-dependence of the density matrix $\rho(t)$ to be

$$\rho(t) = \sum_{R_0,R_D} \int dE \int dE' P_{R_0} e^{-\frac{i}{\hbar}(E-E')t} \alpha^E_{R_0} \Phi^{E'R}_R \langle E' | R \rangle \langle E | R' \rangle,$$

(3.9)

where $\alpha^E_{R_0} = \int dr \psi_{R_0}(r) \Phi^{E}_R(r)$ is the overlap between the initial wavefunction $\psi_{R_0}(r)$ and the system eigenstate wavefunction $\Phi^{E}_R(r) \equiv \langle rR_0 | E \rangle$.

We can evaluate $\langle \hat{C}(t) \rangle$ by inserting a resolution of the identity $\sum_{R_D} \int dr \langle rR_D \rangle \langle rR_D | \hat{C} | r'R_D \rangle$, where $\hat{H}^R(\vec{B}(x)) | R_D \rangle = E_{R_D} | R_D \rangle$ for $x \geq x_D^+$ and $x_D$ is the starting location of the detector (see Figure 3.2). In other words, $|R_D\rangle$ is a Ramsey state in the strong dipolar magnetic field of the detector auxiliary field. The result is

$$\langle \hat{C}(t) \rangle = \sum_{R_D,R_{D'}} \int dr \int dr' \langle r' R_D' \rangle \rho(t) |r R_D\rangle \langle r R_D | \hat{C} | r' R_D' \rangle$$

(3.10)

where we have evaluated the trace in the $|r' R_D'\rangle$ basis and

$$\langle r' R_D' \rangle \rho(t) |r R_D\rangle = \sum_{R_0,R'} \int dE \int dE' P_{R_0} e^{-\frac{i}{\hbar}(E-E')t} \alpha^E_{R_0} \alpha^{E'R}_{R'} \Phi^{E'R}_R \Phi^{E'R}_R(r).$$

(3.11)

We emphasize that $R$ and $R_D$ are indices of different sets of Ramsey states, i.e. $\langle R | R_D \rangle \neq \delta_{R,R_D}$, unless the magnetic fields at the first magnetic lens ($x = 0^-$) and the detector magnetic lens ($x = x_D^+$) happen to be identical.

We also have

$$\langle r R_D | \hat{C} | r' R_D' \rangle = \sum_{R_D} \int dz c_{R_D}(z) \delta(r-z) \delta_{R_0 R_D} \delta(r'-z) \delta_{R_0 R_D}$$

$$= c_{R_D}(r) \delta(r'-r) \delta_{R_D}$$

(3.12)

which, when inserted with Eqn. (3.11) into Eqn. (3.10), results in

$$\langle \hat{C}(t) \rangle = \sum_{R_0,R_D} \int dE \int dE' P_{R_0} e^{-\frac{i}{\hbar}(E-E')t} \alpha^{E'R}_{R_0} \alpha^{E'R}_R \left( \sum_{R_D} \int dr \Phi^{E'R}_R(r) \Phi^{E'R}_R(r) c_{R_D}(r) \right).$$

(3.13)
The wavepacket is almost entirely confined to the region $r \leq 0^-$, as $\psi_{R, R_0}(r)$ has a Gaussian profile (3.5) with spatial width on the order of 10 Å (as determined from the measured velocity distribution for $\mathrm{oH}_2$ [44]). Thus, we can evaluate $\alpha_{k, R_0}^{ER} \equiv \int dr \psi_{R, R_0}(r) \Phi^{ER}_R(r)$ if we know $\Phi^{ER}_R(r)$ for $r \leq 0^-$. Given the definition of the eigenstate $|ER\rangle$, discussed in Section 3.3.1, we show in Section 3.6.1 that $\Phi^{ER}_R(r) = A_R e^{ik R} \delta_{RR_0}$ for $r \leq 0^-$ (cf. Eqn. (3.34)), where $k^{ER_0}$ is defined in Eqn. (3.23). Combined with the definition (3.4) of $\psi_{R, R_0}(r)$,

$$\alpha_{k, R_0}^{ER} \approx \int dr \delta_{RR_0} A_R^2 \sqrt{\sigma_k} \left( \frac{2}{\pi} \right)^{\frac{1}{2}} e^{i(k^{R_0} - k R)^2} e^{-r^2 \sigma_k^2}$$

$$= \delta_{RR_0} \Gamma_{k, R_0}^{ER}$$

(3.14)

where $\Gamma_{k, R_0}^{ER} = A_R^2 \left( \frac{2\pi}{\delta k} \right)^{\frac{1}{2}} - \frac{(k^{ER_0})^2}{4\delta^2}$. Thus,

$$\langle \hat{C}(t) \rangle = \sum_{R_0} \int dE \int dE' P_{R_0} e^{-\frac{i}{\hbar}(E-E') \tau} \Gamma_{k, R_0}^{ER} \Gamma_{k, R_0}^{ER'} \left( \sum_{R_D} \int dr \Phi^{ER}_R(r) \Phi^{ER'}_R(r) c_{R_D}(r) \right),$$

(3.15)

where we have performed the sums over $R$ and $R'$.

If the detection window $t_2 - t_1$ is large enough that the entire wavepacket passes through the detection region defined by $c_{R_D}(z)$ we have

$$P_{\text{detection}} = \frac{1}{\tau} \int_{t_1}^{t_2} \int dE \left. \int dE' P_{R_0} \frac{2\pi \hbar}{\tau} \delta \left( E - E' \right) \Gamma_{k, R_0}^{ER} \Gamma_{k, R_0}^{ER'} \left( \sum_{R_D} \int dr \Phi^{ER}_R(r) \Phi^{ER'}_R(r) c_{R_D}(r) \right) \right|_{t_1}^{t_2}$$

$$= \sum_{R_0} \int dE P_{R_0} \left| \Gamma_{k, R_0}^{ER} \right|^2 \left( \sum_{R_D} \int dr \left( \frac{2\pi \hbar}{\tau} \right) \Phi^{ER}_R(r) \right|_1^t c_{R_D}(r)$$

(3.16)

where $\frac{2\pi \hbar}{\tau} \delta \left( E - E' \right) = \frac{1}{\tau} \int_{-\infty}^{\infty} dr e^{-\frac{i}{\tau}(E-E')t}.$

Physically, one can see that the probability of detection is proportional to the overlap $\left| \Gamma_{k, R_0}^{ER} \right|^2$ of the initial wavepacket and a system eigenstate multiplied by the overlap $\int dr \left( \frac{2\pi \hbar}{\tau} \right) \left| \Phi^{ER}_R(r) \right|^2 c_{R_D}(r)$ of the same system eigenstate and the detection region, as expected.
Substituting for $|t^{ER_0}_{k_0 R_0}|^2$ and given that

$$\Phi^{ER_0}_{RD}(r) \equiv \langle r R_0 | E R \rangle$$

$$= e^{ik_{RD} r} \langle R_D | ER_0 \rangle$$

$$\equiv e^{ik_{RD} r} \beta^{ER_0}_{RD}$$

for $r \geq x^+_D$ (cf. Eqn. (3.34)), we have

$$P_{\text{detection}} = \sum_{R_0} \rho_{R_0} |A_{R_0}|^2 \int dE \left( \frac{2\pi}{\sigma_k} \right) \frac{1}{2} e^{\frac{(E_{R_0} - E)^2}{2\sigma^2_k}} \sum_{R_D} c_{R_D} \frac{2\pi \hbar}{\tau} \beta^{ER_0}_{RD} \right)^2$$

(3.17)

where $c_{R_D} \equiv \int dr \ c_{R_D}(r)$ and $\beta^{ER_0}_{RD} \equiv \langle R_D | ER_0 \rangle$, the projection of the system eigenstate $|ER_0\rangle$ onto the detector eigenstate $|R_D\rangle$ at $x^+_D$. For the purposes of comparing to experiment, only the dependence of $P_{\text{detection}}$ on the experimental parameters is needed, not its absolute value. Also, the value of $A_{R_0} = 1$ as $A_R e^{irk_{ER}} \equiv \langle r R_0 | E R_0 \rangle = e^{irk_{ER}}$ (for $r \leq 0^-$) because of the specific definition of the system eigenstates (see Section 3.3.1). Additionally, one can see that $P_{\text{detection}}$ is not sensitive to minor (on the order of ppm) changes in $k_{R_0}$ as $\sigma_k \propto k_0$ in experiment [44]. Finally, in Eqn. (3.17), only $\beta^{ER_0}_{RD}$ is dependent on the magnetic fields, the scattering geometry, and the surface properties. It is thus sufficient to work with the following equation:

$$P_{\text{detection}} \propto \sum_{R_0} \rho_{R_0} \int dE \ e^{\frac{(E_{R_0} - E)^2}{2\sigma^2_k}} \sum_{R_D} c_{R_D} \left| \beta^{ER_0}_{RD} \right|^2$$

(3.18)

To determine the values of $\beta^{ER_0}_{RD}$, we derive and apply the transfer matrix method with internal degrees of freedom (Section 3.6).

### 3.6 Transfer Matrix Formalism with Internal Degrees of Freedom

The transfer matrix method turns the solution of the time-independent Schrödinger equation of a 1D system into a product of matrices [159]. The present problem has two unique features: (i) the propagating molecules have many internal degrees of freedom which may be mixed as the molecule transitions from one local field to another and (ii) molecules change their propagation direction after scattering by the surface. Problem
(i) is addressed in Section 3.6.1 while (ii) is addressed in Section 3.6.2. The impact of scattering on the internal degrees of freedom is accounted for by using a scattering transfer matrix (Section 3.6.3).

Figure 3.2: Generic field profile of a molecular hyperfine interferometry experiment. The actual magnetic field profiles of the experiment are approximated by \(N+2\) regions of length \(L_i\) and constant magnetic field \(\vec{B}_i\) (black line). The true field profile is asymptotically approached as \(N \to \infty\). We assume large magnetic fields in the regions of the state selector (large arrow) and the detector (eye), which, when combined with the dephasing discussed in Section 3.4, allows us to neglect propagation in the selector and detector regions. That is, the exact locations of \(x_0\) and \(x_D\) are unimportant as long as \(x_0\) is in the high-field region of the state selector, \(x_D\) is in the high-field region of the detector, and all propagation is treated coherently between the two points. The initially Gaussian wavepacket propagates from \(x_0\) along the first branch to the sample surface (cross) at \(x_S\), then, after scattering, propagates along the second branch to \(x_D\). The two branches are separated by an angle \(\theta\). The vertical axis indicates the magnitude of the magnetic field \(|\vec{B}|\) (the direction is not depicted for clarity), with \(|\vec{B}| = 0\) indicated by the grey solid line. \(|R_i\rangle\) denotes the set of eigenstates of \(\hat{H}^{R}(\vec{B}_i)\), Eqn. (3.1), in each region.

### 3.6.1 Propagation and Discontinuity Matrices

We first break up the arbitrary magnetic field profiles of the apparatus into rectangular regions of constant field, as shown in Figure 3.2. We then solve the Schrödinger equation for a single eigenstate in a single region of constant field. Subsequently, we determine the impact of the boundary conditions that exist at the discontinuity between two regions of constant field. Using these solutions, we determine matrices that describe the spatial dependence of the eigenstate wavefunction coefficients within a region of constant field (propagation matrices) and matrices that describe how these coefficients change across the discontinuity between two regions of constant field (discontinuity matrices). Note that while we derive these matrices for
molecules whose internal degrees of freedom are described by the Ramsey Hamiltonian \((3.1)\), the formalism is not limited to this Hamiltonian.

Within a region of uniform magnetic field, the Ramsey Hamiltonian \(\hat{H}^R\) is constant, which allows us to derive the propagation matrix that includes the internal degrees of freedom. We begin by expanding a system eigenstate \(|E\tilde{R}\rangle\) as

\[
|E\tilde{R}\rangle = \sum_R \int dx \Phi_{E\tilde{R}}^R(x) |xR\rangle, \tag{3.19}
\]

where \(\Phi_{E\tilde{R}}^R(x) \equiv \langle xR|E\tilde{R}\rangle\), we define \(|xR\rangle \equiv |x\rangle |R\rangle\), and \(|R\rangle\) is one of the \(N_R\) Ramsey states of a molecule in some magnetic field \(\vec{B}\). Note that \(\vec{B}\) is not necessarily the local magnetic field \(\vec{B}_{loc}\) of the current region and thus \(|R\rangle\) is not necessarily an eigenstate of \(\hat{H}^R(\vec{B}_{loc})\) at this point. Also, the eigenstates \(|E\tilde{R}\rangle\) are labelled by their energy \(E\) and a particular Ramsey index \(\tilde{R}\), such that \(\hat{H}^R(\vec{B})|E\tilde{R}\rangle = E_{\tilde{R}}|E\tilde{R}\rangle\), with \(\vec{B}\) an arbitrarily chosen magnetic field.

Using Eqn. \((3.19)\), the Schrödinger equation with the total Hamiltonian \((3.2)\) can be shown to be (Appendix B.1):

\[
-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \Phi_{E\tilde{R}}^R(x) = \Phi_{E\tilde{R}}^R(x)E - \sum_R H_{R0R}^R \Phi_{E\tilde{R}}^R(x), \tag{3.20}
\]

where \(H_{R0R}^R = \langle R_0|\hat{H}^R(\vec{B}_{loc})|R\rangle\). Eqn. \((3.20)\) is in general difficult to solve because of the coupling of the internal degrees of freedom by \(\hat{H}^R(\vec{B}_{loc})\). However, if we choose the eigenbasis of the internal degrees of freedom to satisfy \(\hat{H}^R(\vec{B}_{loc})|R\rangle = E_R |R\rangle\) (that is, \(|R\rangle\) is now a Ramsey state of a molecule in the local magnetic field \(\vec{B}_{loc}\)), the equations decouple and we obtain

\[
\frac{\partial^2}{\partial x^2} \Phi_{E\tilde{R}}^R(x) = -\frac{2m}{\hbar^2} (E - E_R) \Phi_{E\tilde{R}}^R(x). \tag{3.21}
\]

The solution is

\[
\Phi_{E\tilde{R}}^R(x) = A_R e^{ikRx} + B_R e^{-ikRx}, \tag{3.22}
\]
where $A_R$ and $B_R$ are $R$-dependent coefficients and

$$ k_R \equiv \frac{\sqrt{2m(E - E_R)}}{\hbar}. \quad (3.23) $$

As per the single channel transfer matrix method [159], given that

$$ \Phi^{E_R}(x + \Delta x) = A_R e^{ik_R x} e^{ik_R \Delta x} + B_R e^{-ik_R x} e^{-ik_R \Delta x}, $$

we can collect the $A_R$ and $B_R$ coefficients into a $2N_R$-dimensional coefficient vector

$$ \vec{\phi}_x = (A_1, A_2, ..., A_{N_R}, B_1, B_2, ..., B_{N_R})^T $$

and write

$$ \vec{\phi}_{x_2} = \Pi_{x_2 - x_1} \vec{\phi}_{x_1}, \quad (3.24) $$

where $\Pi_x$ is the $2N_R \times 2N_R$ propagation matrix

$$ \Pi_x \equiv \left[ \bigoplus_{R} e^{ik_R x} \right] \oplus \left[ \bigoplus_{R} e^{-ik_R x} \right] $$

$$ = \begin{pmatrix}
    e^{ik_1 x} & \cdots & 0 \\
    \cdots & e^{ik_{N_R} x} \\
    e^{-ik_1 x} & \cdots & e^{-ik_{N_R} x}
\end{pmatrix}, \quad (3.25) $$

where $\oplus$ denotes the direct sum.

Following the derivation of Ref. [159], we can determine how the coefficients transform across a step discontinuity in the magnetic field. Using the propagation matrix (3.25) and a relabelling of the coordinate system, we can always set the discontinuity to appear at $x = 0$. Given that Eqn. (3.20) applies everywhere, the coefficients $\Phi^{E_R}(x)$ and their derivatives are continuous across the discontinuity (i.e. $\Phi^{E_R}(x) \in C^1(x)$), for each value of $R$. However, the coefficients are only known when $|R\rangle$ is an eigenstate of $\hat{H}_R(\vec{B}_{\text{loc}})$, which differs on each side of the discontinuity (that is, $\vec{B}(0^-) \neq \vec{B}(0^+)$). Note that the wavevector $|E\bar{R}\rangle$ is the same everywhere in the system. Thus, by writing the wavevector $|E\bar{R}\rangle$ in the two different bases corresponding to the eigenstates of $\hat{H}_R$ on each side of the field, we see that the coefficients at a specific value of $x$ are related...
by a basis transformation:

\[ |E\vec{R}^-| = |E\vec{R}^+| \]

\[
\sum_R \int dx \Phi_{R}^{E\vec{R}}(x) |xR^-| = \sum_R \int dx \Phi_{R}^{E\vec{R}}(x) |xR^+| 
\]

\[
\sum_{R,R^+} \int dx \Phi_{R}^{E\vec{R}}(x) \langle R^+|xR^- \rangle = \sum_{R,R^+} \int dx \Phi_{R}^{E\vec{R}}(x) |xR^+ \rangle 
\]

\[
\implies \Phi_{R}^{E\vec{R}}(x) = \sum_R \Phi_{R}^{E\vec{R}_R}(x) \langle R^+|xR^- \rangle, \tag{3.26}
\]

where \(|E\vec{R}^\pm|\) is the wavevector written in the basis of \(|R^\pm|\), \(|R^\pm|\) are the eigenstates of \(\hat{H}^R(\vec{B}(0^\pm))\) on the left (-) and right (+) sides of the discontinuity at \(x = 0\), respectively, and \(\sum_{R,R^+} |R^+ \rangle \langle R^+| \) was inserted in the third line (recall that \(|xR^- \rangle \equiv |x \rangle |R^- \rangle\)). The values \(|R^- \langle R^+| \) are recognized as the matrix elements \(S_{R-R^+}\) of the matrix \(\hat{S}_R^{R^+}\) whose columns are the eigenstates of \(\hat{H}^R(\vec{B}(0^+))\) written in the \(|R^-\rangle\) basis.

Since \(\Phi_{R}^{E\vec{R}}(x) \in C^1(x)\) for each value of \(R\) separately, we can equate the two limits \(\lim_{x^\pm \to 0^\pm} \Phi_{R}^{E\vec{R}}(x)\) and the two limits of the derivative \(\lim_{x^\pm \to 0^\pm} \frac{\partial}{\partial x} \Phi_{R}^{E\vec{R}}(x)\). Solving the resultant equations for the coefficients \(A_{R^+}\) and \(B_{R^+}\), we obtain (Appendix B.2):

\[
A_{R^+} = \sum_{R^-} S_{R^-R}^{R^+} \Delta_{R^+R^-} A_{R^-} + \sum_{R^-} S_{R^-R}^{R^+} \Delta_{R^+R^-} B_{R^-} \tag{3.27}
\]

\[
B_{R^+} = \sum_{R^-} S_{R^-R}^{R^+} \Delta_{R^+R^-} A_{R^-} + \sum_{R^-} S_{R^-R}^{R^+} \Delta_{R^+R^-} B_{R^-} \tag{3.28}
\]

where \(S_{R^-R}^{R^+} \equiv \langle R^+|xR^- \rangle, \Delta_{R^+R^-} \equiv \frac{1}{2} \left( 1 \pm \frac{k_{R^+}}{k_{R^-}} \right), \quad k_{R^\pm} = \frac{\sqrt{2m(E_{R^\pm})}}{\hbar} , \) and \(E_{R^\pm} \equiv \langle R^\pm|\hat{H}^R(\vec{B}(0^\pm))|R^\pm \rangle\). There are \(N_R\) such sets of equations, one for each value of \(R^+\). Working again with \(\vec{\phi}_x = (A_1,A_2,...,A_{N_R},B_1,B_2,...,B_{N_R})^T\), one can write the matrix equation

\[
\vec{\phi}_{x^+} = \mathbf{K}\vec{\phi}_{x^-}, \tag{3.29}
\]

where \(x^\pm\) indicates the location just before (-) or just after (+) the discontinuity located at \(x\) and \(\mathbf{K}\) is the
2NR × 2NR discontinuity matrix

\[
K \equiv \begin{pmatrix}
S_{R-}^{\dagger} \circ \Delta^+ & S_{R-}^{\dagger} \circ \Delta^- \\
S_{R+}^{\dagger} \circ \Delta^- & S_{R+}^{\dagger} \circ \Delta^+
\end{pmatrix}
\]  \hspace{1cm} (3.30)

where \(\circ\) denotes the element-wise Hadamard product, such that \((S_{R-}^{\dagger} \circ \Delta^\pm)_{R'^-R^-} \equiv S_{R-}^{\dagger} \Delta_{R'^-R^-}^\pm \). This matrix allows one to calculate the coefficients of the wavefunction as one moves from one region of constant magnetic field to another through a discontinuity. Thus, if one breaks up any magnetic field profile into a series of constant regions separated by discontinuities, one can systematically approach a perfect description of the propagation of a molecule with internal degrees of freedom through a magnetic field of arbitrary profile through repeated application of \(K\) and \(\Pi_x\). Furthermore, this approach is not restricted to molecules moving through magnetic fields. Many other types of quantum objects moving in a single dimension with internal degrees of freedom that couple to an external static potential can also be analyzed in this way.

The above analysis indicates that one needs to keep track of 2NR components to build up the eigenstates of the system exactly. However, for the current application in mind, one only needs NR components as the magnetic fields typically change the linear molecular momentum by such a small amount that the amplitudes \(B_R\) of the reflected part of the wavefunction are negligible.

For example, a typical velocity of the \(oH_2\) molecules in the experiment of Ref. [44] is \(v_{H_2} = 1450\text{m/s}\). This corresponds to the kinetic energy \(E_{H_2} = \frac{1}{2}m_{H_2}v_{H_2}^2 = 5.31 \times 10^9\text{kHz}\). Ramsey [156] indicates that the maximum energy change for the hyperfine states of \(oH_2\) at 500G is approximately -2550 kHz. The experiment of Ref. [44] has magnetic fields up to about 1000G. For such fields, the energy changes are approximately linear, so we expect the maximum change in energy to be \(\Delta E \approx -5100\text{ KHz}\). In the field-free region before the discontinuity, \(k_{R^-} \approx m_{H_2}v_{H_2}/\hbar\) and after the discontinuity in the field, \(k_{R^+} \approx \sqrt{2m_{H_2}(E - \Delta E)}/\hbar\), as per Eqn. (3.23). Then, \(|\Delta_{R^-R^+}^-| \approx 2.4 \times 10^{-7}\) and \(|\Delta_{R^+R^-}^+| \approx 1\), making \(K\) approximately diagonal and illustrating the decoupling of the forward and backward channels under typical experimental conditions.

We thus only need to keep track of the \(A_R\) components and can define a new coefficient vector

\[
\bar{\psi}_x \equiv (A_1, A_2, \ldots, A_{N_R})^T.
\]  \hspace{1cm} (3.31)

60
The corresponding $N_R \times N_R$ propagation $P_x$ and discontinuity $D$ matrices are

\[ P_x \equiv \bigoplus_R e^{i k_R x} \]
\[
\begin{pmatrix}
e^{i k_1 x} & 0 \\
. & . \\
0 & e^{i k_{N_R} x}
\end{pmatrix}
\]

(3.32)

\[ D \equiv S_{R-}^{R+} \cdot \Delta^+ \]
\[
\approx S_{R-}^{R+},
\]

(3.33)

where the matrix elements of $S_{R-}^{R+}$ are $S_{R-}^{R+} = \langle R^+ | R^- \rangle$, $\hat{H}^R (\vec{B}(0^\pm)) | R^\pm \rangle = E_{R^\pm} | R^\pm \rangle$, $0^\pm$ indicates the position just to the left (−) or right (+) of the discontinuity, and $k_R$ is defined as in Eqn. (3.23). Specifically, $D$ changes the basis of the vector $\vec{\psi}_x$ from $| R^- \rangle$ to $| R^+ \rangle$. That is, $\vec{\psi}_x$ is always in the eigenbasis of $\hat{H}^R (\vec{B}(x))$. Finally, given that $B_R \approx 0$, the eigenstate coefficients are now

\[ \Phi^E_R (x) = A_R e^{i k_R x}. \]

(3.34)

Given that a generic transfer matrix $M$ has the property $M \sigma z M^\dagger = \sigma_z [159]$, the decoupling of the forward and backward channels implies that the forward channel matrix $M_F$ (composed of a product of $P_x$ and $D$ matrices) is now unitary.

### 3.6.2 Rotation Matrices

Scattering off of the sample surface changes both the propagation direction and the internal states of the molecule. To take into account the change in the direction of the propagation path when applying the transfer matrix formalism, we need only change the orientation of the quantization axis. To do so coherently, we apply an $N_R \times N_R$ rotation matrix $R(\phi, \Theta, \chi)$ to $\vec{\psi}_x$, where $\phi$, $\Theta$, and $\chi$ are the Euler angles in the $YZY$ convention. In particular, given the definition of the axes shown in Figure[3.1] a passive rotation of the state vector about the $y$ axis by the angle $\theta$ is required to account for a change of angle $\theta$ in the propagation
direction. This passive rotation modifies the basis of $\vec{\psi}_x$, but leaves the physical state unchanged. We perform this rotation by applying the equivalent active rotation of angle $-\theta$ to $\vec{\psi}_x$.

The matrix elements of $\mathbf{R}(0, -\theta, 0)$, when written in the $|R\rangle$ eigenbasis of $\hat{H}^R(\vec{B}_\text{loc})$ and where $\vec{B}_\text{loc}$ is the local magnetic field, are:

$$
\mathbf{R}(0, -\theta, 0) \equiv \left[ \langle R' | \hat{R}(0, -\theta, 0) | R \rangle \right]
= \left[ \sum_{FMF'M'} \langle R' | F'M' \rangle \langle F'M' | \hat{R}(0, -\theta, 0) | FM \rangle \langle FM | R \rangle \right]
= \mathbf{S}^R_{FM} \mathbf{R}_{FM}(0, -\theta, 0) \mathbf{S}^R_{FM},
$$

(3.35)

where $|FM\rangle \equiv |IJFM\rangle$ is an angular momentum state with total angular momentum $F$, $z$ axis projection $M$, total nuclear spin angular momentum $I$ and total rotational angular momentum $J$; the subscripts of $\mathbf{R}_R$ and $\mathbf{R}_{FM}$ denote the basis of the matrix representation, $|R\rangle$ and $|FM\rangle$, respectively; $\mathbf{S}^R_{FM}$ is the matrix whose columns are the eigenstates $|R\rangle$ written in the $|FM\rangle$ basis, $\hat{R}(\alpha, \beta, \gamma)$ is the rotation operator (also in the ZYZ convention) and

$$
\mathbf{R}_{FM}(0, \beta, 0) = \left[ \delta_{FF'} D^F_{M'M}(0, \beta, 0) \right]
= \left[ \delta_{FF'} d^F_{M'M}(\beta) \right],
$$

(3.36)

where $D^F_{M'M}(\alpha, \beta, \gamma)$ are the Wigner D-matrices and $d^F_{M'M}(\beta)$ are the Wigner small d-matrices [160]. Note that $\mathbf{R}_{FM}(0, -\theta, 0)$ is diagonal in $F$, because of conservation of angular momentum, but not diagonal in $M$. Thus, one must be careful to also perform a passive rotation on the local magnetic field vector if rotations are performed in a region with non-zero field. Typically, however, the sample chamber is magnetically shielded.

### 3.6.3 Scattering Transfer Matrices

Scattering by the sample surface can involve many complex phenomena that may change the internal state, the momentum, and the total energy of the scattering molecule. For this thesis, we focus on scattering processes that conserve the total energy of the molecules. The interactions of the molecules with the sample surface can be phenomenologically described with the total scattering transfer matrix. This matrix is the
\(2N_R \times 2N_R\) matrix \(|\tilde{\Sigma}\rangle\) that relates the wavefunctions on the “left” side of the scattering event to those on the “right” (as opposed to the scattering matrix, which relates the incoming wavefunctions to the outgoing). However, because the initial wavepacket (3.5) does not contain any negative momentum states, the magnetic fields of the solenoids do not cause significant backscattering (Section 3.6.1), and the detector only detects molecular flux in the forward scattering direction, we need only work with the \(N_R \times N_R\) matrix \(\Sigma \equiv P_{\text{fwd}} \tilde{\Sigma} P_{\text{fwd}}^\dagger\), where \(P_{\text{fwd}}\) is an \(N_R \times 2N_R\) projection matrix onto the forward scattering states. We define \(\Sigma\) in the \(|Im_l J m_J, I J m_J\rangle\) basis, where the \(|Im_l J m_J\rangle\) states are themselves defined with respect to the quantization axis that is normal to the surface sample. We choose this basis to relate to scattering calculations, which are frequently carried out in the \(|J m_J\rangle\) basis with a quantization axis normal to the sample surface.

In general, the scattering transfer matrix elements \(\Sigma_{Im_l J m_J I J m'_J, I J m'_J}\) are functions of the incident energy \(E\), the outgoing energy \(E'\), the incident momentum \(\vec{k}\), and the outgoing momentum \(\vec{k}'\). As we are restricting ourselves to iso-energetic processes, \(E = E'\). Also, Eqn. (3.23) defines the magnitudes of the momentum before and after the scattering event. This leaves the scattering transfer matrix elements as functions of only energy and the four angles that define the scattering geometry. These angles are: (1) the angle between the two branches, (2) the angle between the surface normal and the scattering plane, (3) the angle between the first branch and the projection of the surface normal on the scattering plane, and (4) the azimuthal angle of the sample. The scattering plane is the plane defined by the two branches of the apparatus.

Given that the experiment only probes a single scattering direction at a time, the scattering transfer matrix will not, in general, be unitary. This incorporates state-dependent loss channels into the formalism. Additionally, the scattering transfer matrix is, in general, time-dependent. Here, we assume that the time-scales of the surface dynamics are significantly different from the molecule-surface or wavepacket-surface interaction time-scales and assume \(\Sigma\) to be time-independent.

Because \(\Sigma\) is defined with respect to the surface normal, we use rotation matrices to appropriately change the basis of \(\vec{\Psi}\) before and after applying the scattering transfer matrix. We ensure that the total rotation corresponds to the change in propagation direction induced by scattering off of the sample surface and that the quantization axis is again coplanar with the two branches of the apparatus.

The scattering transfer matrix elements for a specific molecule-surface interaction can be determined from scattering calculations [44, 161]. Alternatively, they can be treated as free parameters and determined
from the experimental measurements by solving the inverse scattering problem. Such a problem can potentially be solved efficiently using machine learning based on Bayesian optimization [45, 64].

### 3.6.4 Calculation of Eigenstate Coefficients

To determine the dependence of the probability of detection (3.18) on the magnetic fields and the surface properties, we must determine the coefficients \( \beta_{ERD} \). This can be done by multiplying the initial coefficient vector \( \psi_{ERD}^{\text{R}_0} \) of a system eigenstate \(|ER_0\rangle\) by a succession of transfer matrices to obtain the final coefficient vector \( \psi_{ERD}^{\text{R}_D} \equiv \left( \beta_1^{ERD}, \beta_2^{ERD}, \cdots, \beta_N^{ERD} \right)^T \):

\[
\psi_{ERD}^{\text{R}_D} = S_{R_D}^{R_0} M_2 M_1 \psi_{ERD}^{\text{R}_0},
\]

where \( M_{1/2} \) describes the propagation through the first/second branch of the apparatus, \( M_\Sigma \) describes the scattering, and \( S_{R_D}^{R_0\dagger} \) changes the basis of the coefficient vector to the eigenbasis \(|R_D\rangle\) of \( \hat{H}^R(\vec{B}(x_D)) \) at the location of the detector \( x_D \). The \( M \) matrices are defined as

\[
M_1 = P_L S_{R_{0,L}}^{R_D} \cdots P_{L_2} S_{R_{1,L}}^{R_D} P_{L_1} S_{R_{0,1},L}^{R_D} \quad (3.38)
\]

\[
M_\Sigma = S_{R_{FM}}^{R_D\dagger} R_{FM}(0,-\theta_2,0) \Sigma_{FM} R_{FM}(0,\theta_1,0) S_{R_F}^{R_D\dagger} \quad (3.39)
\]

\[
M_2 = P_{L_{N+1}} S_{R_{N,L+1}}^{R_D} \cdots P_{L_{N+1}} S_{R_{N+1}}^{R_D} \quad (3.40)
\]

where \( R_{ini} \) refers to the eigenbasis \(|R_{ini}\rangle\) of \( \hat{H}^R(\vec{B}(0^-)) \) at the initial location of the wavepacket; \( R_i \) refers to the eigenbasis \(|R_i\rangle\) of \( \hat{H}^R(\vec{B}_i^+) \) in region \( i \) of the apparatus, as depicted in Figure 3.2; \( FM \) refers to the \(|IJFM\rangle\) basis where \( \vec{F} \equiv \vec{I} + \vec{J} \) and \( M \) is the projection on the local \( z \) axis; \( \theta_1 \) is the angle between the \( z \) axis and the sample surface normal (see Figure 3.1); \( \theta_2 \) is the angle between the sample surface normal and the \( z' \) axis; \( L_i \) is the length of region \( i \), as depicted in Figure 3.2; \( N \) is the total number of regions between \( x_0 \) and \( x_D \) (see Figure 3.2); \( n \) is the number of regions between the initial position of the wavepacket \( x_0 = 0 \) and the sample position \( x_S \); \( \Sigma_{IJFM} \equiv S_{R_{IJ}}^{FM\dagger} \Sigma S_{R_{IJ}}^{FM\dagger} \) is the scattering transfer matrix written in the \(|IJFM\rangle\) basis; \( R_{IJ} \equiv \text{Im} m_Im_J \); and \( \Sigma \) is the scattering transfer matrix in the \(|Im_Im_J\rangle\) basis. Note that \( \theta_1 + (-\theta_2) \equiv -\theta \), where \( \theta \) is the angle between the branches of the apparatus (see Figure 3.1). Note also that Eqn. (3.39) assumes that the surface normal is in the scattering plane; additional rotation matrices are needed if this is not the case.
By defining a matrix \( \Psi_x \equiv (\vec{\psi}_E^1, \vec{\psi}_E^2, \ldots, \vec{\psi}_E^{N_R}) \), all \( N_R \times N_R \) coefficients \( \beta_{ERD} \) can be simultaneously obtained from

\[
\Psi_x = S_{RD}^\dagger M_2 M_2 M_1 \Psi_x
\]

\[
= S_{RD}^\dagger M_2 M_2 M_1 \mathbb{1}_{N_R},
\]

where \( \Psi_x \equiv \mathbb{1}_{N_R} \) because of the specific definition of the system eigenstates (see Section 3.3.1). Using Eqns. (3.38–3.41), we can obtain \( \beta_{ERD} \), and thus \( P_{\text{detection}} \), as functions of the magnetic field profile, the scattering matrix elements, and the scattering geometry.

### 3.7 Application to ortho-Hydrogen

The theoretical framework described in Sections 3.3–3.6 connects the scattering transfer matrix elements \( \Sigma_{I_mJ_mI'_mJ'_m} \) to the experimentally observed signal, which is proportional to \( P_{\text{detection}} \). By changing the magnetic field profiles in the two arms of the apparatus, one can obtain information about how the scattering affects various hyperfine states. To illustrate our theoretical framework and to demonstrate the impact of the scattering transfer matrix on the experimentally observed signal, we consider a beam of \( oH_2 \) in its ground state and a simplified apparatus that contains only a few regions of constant magnetic field, as depicted in Figure 3.3.

#### 3.7.1 Ortho-Hydrogen Hyperfine Hamiltonian

The Hamiltonian describing the relevant internal degrees of freedom of \( oH_2 \) is \([156]\):

\[
\hat{H}_{\text{ortho}}^R(\vec{B}) = -\alpha \vec{I} \cdot \vec{B} - \beta \vec{J} \cdot \vec{B} - c \vec{I} \cdot \vec{J} + \frac{5d}{(2J-1)(2J+3)} [3(\vec{I} \cdot \vec{J})^2 + \frac{3}{2} \vec{I} \cdot \vec{J} - \vec{I}^2 \vec{J}^2]
\]

where, for simplicity, we have neglected magnetic shielding of the nuclear and rotational magnetic moments by the molecule and diamagnetic interactions of the molecule with the magnetic field; \( \vec{B} \) is the local magnetic field; \( \vec{I} \) is the nuclear spin operator; \( \vec{J} \) is the rotational angular momentum operator; \( \alpha \equiv \mu_I / \hbar \approx 4.258 \text{kHz} \); \( \beta \equiv \mu_J / \hbar \approx 0.6717 \text{kHz} \); \( c \approx 113.8 \text{kHz} \); \( d \approx 57.68 \text{kHz} \); \( I = 1 \) is the total nuclear spin angular momentum in units of \( \hbar \); \( J = 1 \) is the total rotational angular momentum in units of \( \hbar \); \( \mu_I \) is the nuclear magnetic moment.
Figure 3.3: A magnetic field profile that approximates the true magnetic field profile of an experiment using $oH_2$. We combine this approximate field profile with the transfer matrix formalism to calculate the observed signal. $B_i$ refers to the different magnetic field vectors of the control fields. $z'$ and $x'$ refer to the new coordinate system defined to align with the second branch of the apparatus (see Figure 3.1). The sample is located at the cross in the centre of the diagram. The surface normal of the sample is assumed to bisect the angle between the two branches of the apparatus. The propagation direction is $x$ before scattering and $-x'$ after scattering. The angle between $x$ and $-x'$ (i.e. the angle between the two arms of the apparatus) is $\theta = 45^\circ$. $B_1$ is directed along $x$ and $B_2$ is directed along $-x'$, as per the arrows. The fields just after the state selector and just before the detector are directed toward the $z$ and $z'$ directions, respectively and as per the arrows. Additional computational parameters not shown above can be found in Appendix B.3.

of a single nucleus; and $\mu_J$ is the magnetic moment due to molecular rotation. The first two terms describe the interaction of the nuclear and rotational magnetic moments with the external magnetic field, the third term describes the nuclear spin-rotational magnetic interaction [156, 162, 163], and the terms proportional to $d$ describe the magnetic spin-spin interaction of the two nuclei [156, 162, 163].

3.7.2 Experiment and Observables

While there are many possible experimental protocols, we focus on the full interferometer mode used by Godsi et al. [44]. The experiment is performed by initiating a continuous flux of $oH_2$ molecules through the apparatus and measuring the current of the ionization detector while varying the first and second control fields ($B_1$ and $B_2$ in Figure 3.3).

In particular, $B_1$ is set to a specific value while $B_2$ is varied around the point $-B_1$ (i.e. about the spin-echo condition). In principle, $B_2$ could also be set to vary around $+B_1$, where spin echoes have also been observed.
Figure 3.4: Upper Panels: Calculated and experimental signals close to the spin echo condition versus the magnetic field of the second coil \(|B_2|\). Lower Panels: Fourier amplitudes of the upper panels versus the generalized gyromagnetic ratio \(\gamma\). For panels (a), (b), (d), and (e), the field profile is depicted in Figure 3.3; \(B_1 = 440\) gauss; the scattering transfer matrix \(\Sigma = \mathbb{I}_0\) and is constant for all energies; and the signal is sampled at a rate of 300 points per 20 gauss. Panels (a) and (d) only include a single velocity (or, equivalently, a single value of energy) in Eqn. (3.18) while panels (b) and (e) include the full integral. For the experimental data shown in panel (c), \(B_1 = 437\) gauss, the sample was the (111) surface of Cu and the signal was sampled every 0.065 gauss (a sampling rate of \(\sim 308\) points per 20 gauss). Panel (f) shows data for \(oH_2\) scattering off of Cu(111) (blue full circles) and Cu(115) (red open circles). All experimental data was obtained from Godsi et al. [44].

[155], but we choose to vary \(B_2\) about \(-B_1\) to match the relevant experiment by Godsi et al. [44]. This variation of the magnetic fields results in oscillatory curves of the detector current versus \(B_2\), as shown in Figure 3.4 (a–c). These oscillations reflect the interference pattern that occurs when the various wavepackets recombine after passing through the final control field (see Section 3.3). This interference pattern contains information about how the individual hyperfine states of the molecule interact with the sample surface.

The \(x\) directed magnetic fields of a solenoid changes the energies of all of the \(N_R = 9\) hyperfine states and induces all \(\binom{N_R}{2} = 36\) possible transitions. The frequencies of these transitions depend on the magnitude of the magnetic fields. By changing the magnitude of the second magnetic field, we are able to probe the rates of change of these transition frequencies with the magnetic field: the (generalized) gyromagnetic...
ratios $\gamma_{ij}(B) = \left| \frac{\partial \omega_{ij}(B)}{\partial B} \right|$, where $\omega_{ij} \equiv \frac{1}{\hbar} \Delta E_{ij} = \frac{E_i - E_j}{\hbar}$, and $E_i$ is the energy of Ramsey state $i$ [44]. The Fourier transforms of the oscillatory curves that give these gyromagnetic ratios are shown in Figure 3.4 (d–f). To obtain these results, we assumed that the surface normal of the sample bisects the angle defined by the two branches and that the scattering transfer matrix is the identity matrix and is independent of energy, i.e. we assume for the present calculation that the only impact of scattering is the change of propagation direction, as modelled with rotation matrices (Section 3.6.2).

**Figure 3.5:** 2D Fourier amplitude plots formed by the concatenation of spectra plots (such as Figure 3.4 (d-f)) for various values of the magnetic field of the first solenoid $B_1$. Colour indicates the Fourier amplitude. For the theory plots, the field profile is depicted in Figure 3.3; the scattering transfer matrix $\Sigma = 1_9$ and is constant for all energies; $B_2$ was varied from $-(B_1 - 10 \text{ gauss})$ to $-(B_1 + 10 \text{ gauss})$; the signal was sampled at a rate of 300 points per 20 gauss; and all data with a value less than $10^{-3.5}$ has been replaced with $10^{-3.5}$ for clarity. For the experimental plot, the sample was the (111) surface of Cu; all data with a value of less then $10^{-8}$ has been replaced with $10^{-8}$ for clarity; the dashed lines indicate transitions identified by Godsi et al. [44]; and the data was obtained from Godsi et al. [44].

The location of each feature in the spectra is reflective of a gyromagnetic ratio and is independent of the molecule-surface interactions, being only a function of the hyperfine energy level structure of $o\text{H}_2$. The relative height of each feature, however, is dependent on the molecule-surface interactions, as exemplified in the experimental spectrum shown in Figure 3.4 (f). From Figure 3.4, one can see that integrating over the velocity distribution is important to produce the spin-echo effect and to bring the observed signal closer to experiment.

A different spectrum can be obtained for every possible value of $B_1$ and then combined to form a 2D map of the generalized gyromagnetic ratios and their contributing amplitudes as a function of $B_1$, as shown in Figure 3.5. This protocol is equivalent to observing the scattering of molecules with different internal
hyperfine states as different values of the magnetic field in the first branch produce different superpositions of the hyperfine states. One can clearly see both the magnetic field-dependence of the gyromagnetic ratios, the impact of integrating over the velocity distribution, and the stark similarities and differences between the experimental and theory plots.

These differences (also seen in Figure 3.4) occur as we have not taken the details of the molecule-surface interactions into account in the theoretical calculations. That is, the identity scattering transfer matrix that we use to calculate the theoretical plots does not well-represent the molecule-surface interactions present in the experiment. A scattering transfer matrix that well-represents the molecule-surface interactions still needs to be determined; one approach to determining this scattering transfer matrix is discussed in Chapter 4. Furthermore, differences between the theoretical and experimental plots also arise because the field profile used for the theoretical calculations (Figure 3.3) only approximates the real magnetic field profile; we can systematically reduce the differences arising from this approximation by simply increasing the number of field regions in the field profile used, however.

We now examine the sensitivity of the calculated signals to various changes in the scattering transfer matrices. Figures 3.6 and 3.7 demonstrate the impact of random variations of the scattering transfer matrix $\Sigma$ on the oscillatory plots (for $B_1 = 440$ gauss) and their spectra, respectively. For simplicity, we keep the matrix elements of $\Sigma$ independent of energy.

The first row of each figure (labelled RP) reflects the impact of differing phases imparted to each hyperfine state after scattering. Specifically, $\Sigma = \bigotimes_{i=1}^{9} e^{i\theta_i}$ is a diagonal unitary matrix whose nine phases $\theta_i$ are randomly chosen from a uniform distribution of width $2\pi$. Such a form of scattering would result from purely elastic scattering where the different hyperfine states probe the surface for different lengths of time (i.e. each state penetrates to a different depth or encounters a resonance with a different lifetime). Significant differences in the relative peak amplitudes can already be seen at this point, indicating that the calculated signal is sensitive to these phases.

The second row of each figure (labelled RDA) reflects the impact of differing state losses due to scattering. Specifically, $\Sigma = \bigotimes_{i=1}^{9} A_i$ is a diagonal matrix whose diagonal elements are randomly chosen from a uniform distribution on the interval $[0, 1)$. This form models the impact of different losses of each hyperfine state to different scattering directions, reactions with the surface, or adsorption to the surface. Again,
significant changes are observed, indicating sensitivity to these features.

The third and fourth rows (labelled ROM and RUM, respectively) probe the impact of inelastic (projection $m_I m_J$-changing) scattering on the calculated signal. For the third row, $\Sigma$ is an orthogonal matrix randomly drawn according to the Haar measure on $O(9)$, while $\Sigma$ is a unitary matrix randomly drawn according to the Haar measure on $U(9)$ for the fourth row of each figure. Here, randomly drawing according to the Haar measure can be understood as analogous to drawing from the “uniform distribution” over the space of possible matrices [164]. The orthogonal matrices model inelastic scattering where no relative phase changes occur, while the unitary matrices model inelastic scattering where relative phase changes do occur. In both cases, there is no loss of total population during scattering. Clearly, the calculated signals are also sensitive to inelastic scattering events, both with and without relative phase changes.

### 3.8 Comparison with a Semi-Classical Method

The present approach is fully quantum mechanical, while Godsi et al. [44] have described a semi-classical method for calculating $P_{\text{det}}$ that they used to model the propagation of $oH_2$ in their molecular hyperfine interferometer. This semi-classical method treats the internal degrees of freedom of the molecules quantum mechanically and the centre-of-mass motion classically. As a result, the momentum changes induced by the magnetic field are ignored and every internal state is described as propagating at the initial velocity $v_0$ of the molecule. The internal degrees of freedom are treated by applying the time evolution operator for the time period $t_i \equiv \frac{L_i}{v_0}$ spent in each magnetic field of length $L_i$. That is, the propagation is calculated in the molecular reference frame with a time-dependent Hamiltonian. Here, we compare the results of the two approaches.

As per the application of the semi-classical method in Godsi et al. [44], we compare the methods in the “flux detector” mode of the molecular hyperfine interferometer. This flux detector mode has a field profile as shown in Figure[3.3] but with the second arm assumed to be of zero length and $B_2 = 0$. The field $B_1$ is varied. For the sake of comparison, we also set the state selector and detector fields in the transfer matrix method to 100000 gauss so that the basis changes performed by the transfer matrix method out of and into these regions match well the Clebsch-Gordon transformation from $|m_I m_J\rangle$ to $|F m\rangle$ and its inverse, as used by the semi-classical method. Note that the off-diagonal elements of the full discontinuity matrix $K$ (3.30) are still
Figure 3.6: Calculated signals close to the spin echo condition as functions of the magnetic field of the second coil \( |B_2| \). The field profile is depicted in Figure 3.3; \( B_1 = 440 \) gauss; and the signal was sampled at a rate of 300 points per 20 gauss. Each of the plots was created with identical parameters, except for the scattering transfer matrices \( \Sigma \). The scattering transfer matrices are identical for all energies and are randomly chosen for each plot as follows. First row – Random Phases (RP): \( \Sigma = \bigoplus_{i=1}^9 e^{i\theta_i} \) is a diagonal unitary matrix whose nine phases \( \theta_i \) are randomly chosen from a uniform distribution of width \( 2\pi \). Second row – Random Diagonal Amplitudes (RDA): \( \Sigma = \bigoplus_{i=1}^9 A_i \) is a diagonal matrix whose diagonal elements are randomly chosen from a uniform distribution on the interval \([0, 1)\). Third Row – Random Orthogonal Matrices (ROM): \( \Sigma \) is an orthogonal matrix randomly drawn according to the Haar measure on \( O(9) \). Fourth Row – Random Unitary Matrices (RUM): \( \Sigma \) is a unitary matrix randomly drawn according to the Haar measure on \( U(9) \). Here, randomly drawing according to the Haar measure can be understood as analogous to drawing from the “uniform distribution” over the space of possible matrices [164].
Figure 3.7: Fourier transforms of the signals shown in Figure 3.6 as functions of the generalized gyromagnetic ratio \( \gamma \). The panel labels are described in the figure caption to Figure 3.6.

only \( \sim 10^{-5} \) at 100000 gauss, such that their neglect still does not invalidate our fully quantum formalism at these large field strengths. We also retain the rotation from the first branch to the second branch and set the scattering matrix to \( \mathbb{1}_9 \). To maximize sensitivity of the comparison, we use only a single velocity when calculating \( P_{\text{detection}} \) in both methods. All other parameters, including the state selector and detector relative state probabilities \( \eta_{\text{m}_{m_{z}}} \) and \( \kappa_{\text{m}_{m_{z}}} \), are as per Appendix B.3. This allows for a test that includes all incoming and outgoing states and their relative phases at experimentally relevant conditions.
Before discussing the results of the comparison between the two methods, however, I will comment on their relative computational cost. Without performing a detailed analysis, the computational complexities of the two methods appear to be similar as the computational costs of both methods should scale proportionally to the number of field regions included in the magnetic field profile used. As such, it appears that the details of how each method is programmatically implemented will determine whether one method out-performs the other computationally.

I will now return to the results of the comparison between the two methods. We calculate the signals \( P_{\text{detection}}(B_1) \) from \( B_1 \) to \( B_1 + 10 \text{ gauss} \) for various values of \( B_1 \). We include 1500 data points in these ten-gauss intervals. The calculated signals are compared between the two methods by calculating their relative absolute difference at identical conditions. This produces a relative absolute difference at each of the 1500 magnetic field points. We then calculate the maximum, mean, and median relative absolute difference over the ten-gauss interval. Figure 3.8 shows how these maximum, mean, and median values vary as a function of \( B_1 \). At low fields, there is no significant dependence on the magnetic field and the relative absolute difference is below the expected experimental error. This lack of dependence on \( B_1 \) is possibly due to some inherent numerical error present in the implementation of one or both methods, which masks any underlying field-dependence. At approximately 460 gauss, however, the error begins to increase with the magnetic field until it saturates at approximately 46000 gauss at a relative absolute difference of approximately one. This increase in error as a function of magnetic field points to a systematic difference between the two methods.

To illustrate the difference between the two approaches in more detail, we plot the Fourier transforms of the calculated signals at various magnetic field strengths in Figure 3.9. At field strengths below 1000 gauss, little difference is observed. At higher field strengths, the feature locations agree, while the Fourier amplitudes differ. The feature locations are determined by the eigenvalues of the Hamiltonian, identical in both methods, while the amplitudes are a function of the relative phases and amplitudes of the wavefunction components. These amplitudes and phases are expected to differ between the two methods at sufficiently high fields because of the approximations made in the semi-classical method.

In particular, the semi-classical method accounts for most of the relative phase and amplitude changes induced by the controlling magnetic fields. It does this by time-evolving the internal state vector for times that correspond to the time \( t_i \) spent in each magnetic field by a molecule moving at its unchanged initial
Figure 3.8: Maximum, median, and mean relative absolute difference between the calculated signals obtained by the semi-classical method discussed in the Supplementary Material of Godsi et al. [44] and the present method, for various values of the controlling magnetic field. See Section 3.8 for a description of the semi-classical method and the field profiles used. The relative absolute difference between the calculated signals is calculated point-by-point as a function of the magnetic field. The mean, median, and maximum values are then calculated over the magnetic field interval spanned by the calculated signal. The calculated signal is sampled at a rate of 1500 points per 10 gauss; the magnetic field varies from $B_1$ to $B_1 + 10$ gauss for each calculated signal; a single velocity was included in the calculations; the scattering transfer matrix $\Sigma = \mathbb{I}_9$ and is constant for all energies. All other parameters are listed in Appendix B.3.

velocity. However, the semi-classical method ignores the small changes in the molecular velocity caused by the magnetic fields. These changes to the velocity modify the time spent in each magnetic field for each individual component of the internal state vector. Thus, $t_i$ should depend on the internal state $|R\rangle$. It is not immediately clear how to include these state-dependent velocity changes into the semi-classical method, however.

At low fields, these velocity changes and the dependence of $t_i$ on $|R\rangle$ are negligible and the fully quantum calculations agree with the semi-classical results to at least 0.1% for fields below 1000 gauss. However, this agreement can only be expected to occur for surfaces that do not change between the surface-impact events of the spatially-separated wavepacket components (discussed in Section 3.3). The maximum temporal separation between these impact events, caused by the velocity changes, varies from a few to several hundreds of picoseconds. Many surfaces do change on this timescale, as has been measured in several helium-3 spin echo experiments [33–39]. In other words, the semi-classical method cannot be used to probe the dynamics of surfaces, while the method presented in this chapter opens the possibility to account for the surface
dynamics with molecular scattering experiments.

Figure 3.9: Fourier amplitudes of the calculated signals at various magnetic field values computed with the semi-classical method discussed in the Supplementary Material of Godsi et al. [44] (orange) and the present method (blue) as functions of the generalized gyromagnetic ratio $\gamma$. The calculation conditions are identical to those of Figure 3.8.

3.9 Conclusion and Future Directions

In this chapter, we have developed a theoretical framework for simulating a surface-sensitive molecular hyperfine interferometer. The approach treats the interferometer as an effective one-dimensional system, accounting for the real experimental geometry by rotating the quantization axis of the hyperfine states at the scattering point. The time evolution of the molecular states is described fully coherently and accounts for the mixing of the hyperfine states and momentum changes induced by the magnetic fields in the experiment. The present approach is fully quantum mechanical and includes a full description of the internal-state-dependent
spatial superpositions imposed on the molecular wavepackets by the controlling magnetic fields. This opens the door for a description of molecular scattering experiments that aim to probe surface dynamics on the picosecond to hundreds of picosecond time scale. To build the framework, we have derived and implemented a transfer matrix formalism that accounts for the internal (hyperfine) degrees of freedom of molecules and that allows for efficient computation of the experimental signal.

In the present work, the molecule-surface interaction is accounted for by a scattering transfer matrix (a transformed version of the scattering matrix) that is suitable for the description of experiments where the surface changes either much more slowly or much more quickly than the molecule-surface or wavepacket-surface interaction times. The extension to arbitrary surface dynamics (currently under investigation) requires a time-dependent scattering transfer matrix that reflects the underlying time-dependence of the molecule-surface interaction potential. We have demonstrated, using the specific case of oH$_2$, how the different features of the time-independent scattering transfer matrix, such as the phases of the diagonal elements, impact the experimental signal. In addition, we have shown that the experimental signal is sensitive to inelastic scattering events that change the projection quantum numbers of the molecular hyperfine states.

The present approach also sets the stage for solving the inverse scattering problem in molecular hyperfine interferometry by means of machine learning approaches, such as Bayesian optimization [45, 64]. For example, the difference between the experimental observations and the results of the transfer-matrix computations can be efficiently minimized by using the Bayesian optimization algorithm to intelligently vary the scattering matrix elements. We discuss this approach in Chapter 4.

The formalism presented here is general to all closed shell molecules and is flexible to describe various experimental setups. It can be used to explore various experimental protocols and evaluate their effectiveness at determining various molecule-surface interactions and surface properties. Thus, this chapter provides the theoretical framework necessary to interpret a wide range of molecular hyperfine interferometry experiments. These new experiments are poised to provide new information about molecule-surface interactions, surface morphologies, and surface dynamics, such as the impact of molecular orientation on surface reaction rates and the details of molecule-surface potential energy surfaces.
3.10 Summary

This section contains a summary of the key results presented in Chapter 3:

- Proposed using generic closed shell molecules for molecular hyperfine interferometry experiments (Section 3.2)
- Developed a fully quantum formalism to describe molecular hyperfine interferometry with closed shell molecules (Sections 3.3.1-3.6)
- Derived a transfer matrix formalism with internal degrees of freedom (Section 3.6)
- Applied the fully quantum formalism to the description of MHI experiments using $^{2}$H$_{2}$ molecules (Section 3.7)
- Showed that various features of the scattering matrix affect the experimental signal (Section 3.7)
- Compared the quantum formalism with a semi-classical formalism and determined that have the same information at low fields, for static surfaces (Section 3.8)
- Showed how the quantum formalism opens the door for a formalism that can account for dynamic surfaces (Section 3.8)
- Set the stage for the determination of the molecule-surface scattering matrix via machine learning techniques (Sections 3.3.1-3.6)
Chapter 4

Bayesian optimization to determine the molecule-surface scattering matrix

4.1 Introduction

This chapter discusses our first applications of Bayesian optimization to the inverse scattering problem for molecular hyperfine interferometry (MHI) experiments. I begin this chapter by discussing how we approach the inverse scattering problem from an optimization perspective in Section 4.2. Then, I give a non-technical overview of Bayesian optimization by first describing Gaussian processes (Section 4.3.1), then Gaussian process models (Section 4.3.2), and finally the Bayesian optimization algorithm itself (Section 4.3.3). I discuss our initial applications of Bayesian optimization to determine the scattering matrix elements from a given signal in Section 4.4. I then conclude the chapter in Section 4.5 and give a bulleted summary of the chapter results in Section 4.6.

4.2 The Inverse Scattering Problem

In Sections 3.2–3.6 we developed a framework that describes the MHI experiments and that allows us to connect the scattering transfer matrix elements to the detector signal. In general, however, we do not know the scattering matrix elements. We would like to determine these matrix elements from experimental data in order to either gain information about the surface and/or improve surface scattering calculations. This is the inverse scattering problem for MHI.
One simple way to determine the scattering matrix elements is to vary the values of the matrix elements until the calculated signal matches the experimental signal. (Note that, in this chapter, I use the terms “scattering matrix” and “scattering transfer matrix” interchangeably, for if we know the matrix elements of the scattering matrix, we can calculate the elements of the scattering transfer matrix, and vice versa.) If there were only a small number of parameters that define the scattering matrix and if the calculation of the signal \( P_{detection} \) were fast, such a naive approach would work rather well. However, a simple grid search of a three-parameter space requires 1000 evaluations for even a low-resolution \( 10 \times 10 \times 10 \) grid. If each calculation takes several minutes, this begins to be computationally expensive; if we need to explore a parameter space of possibly more than 81 dimensions, such an approach is completely prohibitive.

As such, we need a more sophisticated approach. One such approach is Bayesian optimization, which is a machine learning technique particularly well-suited to situations where every evaluation of the function to be optimized is computationally expensive [45]. This approach has been used for the inverse scattering problem in quantum reaction dynamics [64]. To determine the scattering matrix parameters in MHI experiments via optimization, we minimize the mean square error (MSE) between the calculated and experimental signals. Ideally, we want to find the global minimum of this MSE and all local minima that also give a small MSE. We want these local minima as there may be multiple scattering matrices that can explain the experimental signal and because the global minimum may be over-fit to noise in the experimental signal. This noise in the experimental signal could arise from imperfections in the magnetic fields, errors in the measurement of the magnetic field profile, errors or variation in the current supplied to the solenoids, variation of the molecular velocity distribution between experimental runs, and variation of the detector sensitivity. Ultimately, we may not be able to determine a unique scattering matrix, but we should, at least, be able to determine classes of matrices that match the data and/or put bounds on the values of the matrix elements.

I will now give a non-technical overview of Bayesian optimization and then present the results of our initial applications of Bayesian optimization to the inverse scattering problem for MHI experiments.

### 4.3 Bayesian Optimization

As previously mentioned, Bayesian optimization is most advantageous when each evaluation of the function to be optimized is rather expensive. Bayesian optimization relies on using a Gaussian process model to guide
the algorithm and, in turn, a Gaussian process model utilizes a Gaussian process to give an interpolating model of the data obtained from function evaluations. Additionally, the Gaussian process model provides an estimate of the uncertainty in the model, which is vital for the algorithm to function correctly. In this section, I will give a high-level introduction to Gaussian processes, Gaussian process models, and Bayesian optimization.

4.3.1 Gaussian Processes

Essentially, a stochastic process (or simply process) can be thought of as a continuum limit of a joint probability distribution, such that there is a probability distribution for every point on the real line (there are more general and more rigorous definitions [165], but this suffices for the present purposes). A Gaussian process is then a process where this distribution is Gaussian. To build up an intuition for what a Gaussian process is, let us first look at a one dimensional Gaussian distribution

\[ P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]  

[166]. This distribution is characterized by both a mean \( \mu \) and a standard deviation \( \sigma \).

If we look at a Gaussian distribution over two variables, we have the joint probability distribution

\[ P(x_1, x_2) = (2\pi)^{-1} \left( \sigma_1^2 \sigma_2^2 - \sigma_1^2 \sigma_2^2 \rho^2 \right)^{-\frac{1}{2}} e^{-\frac{1}{2\rho^2} \left( \frac{(x_1-\mu_1)^2}{2\sigma_1^2} + \frac{(x_2-\mu_2)^2}{2\sigma_2^2} + \frac{\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1 \sigma_2} \right)} \]  

[166]. Here, the distribution has five parameters: two means \( \mu_1, \mu_2 \); two standard deviations \( \sigma_1, \sigma_2 \); and the covariance \( \sigma_{12} \equiv \rho \sigma_1 \sigma_2 \), where \( \rho \) is the correlation. We can also define the symmetric covariance matrix \( \Sigma_{\text{cov}} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \) and then write

\[ P(x_1, x_2) = P(\bar{x}) = (2\pi)^{-\frac{d}{2}} \left[ \det(\Sigma_{\text{cov}}) \right]^{-\frac{1}{2}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu})^T \Sigma_{\text{cov}}^{-1} (\bar{x}-\bar{\mu})} \]

where \( \bar{x} \equiv (x_1, x_2) \) and \( \bar{\mu} \equiv (\mu_1, \mu_2) \)

[166].

We can then expand this to the \( d \)-dimensional joint distribution

\[ P(\bar{x}) = (2\pi)^{-\frac{d}{2}} \left[ \det(\Sigma_{\text{cov}}) \right]^{-\frac{1}{2}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu})^T \Sigma_{\text{cov}}^{-1} (\bar{x}-\bar{\mu})} \]

where \( \bar{x} \equiv (x_1, x_2, \ldots, x_d) \) and \( \bar{\mu} \equiv (\mu_1, \mu_2, \ldots, \mu_d) \) are both \( d \)-dimensional and there are \( d \) standard deviations \( \sigma_i \), and \( \binom{d}{2} \) covariance terms \( \sigma_{ij} \). That is \( \Sigma_{\text{cov}} \) is the symmetric \( d \) by \( d \) matrix:

\[
\Sigma_{\text{cov}} = \begin{bmatrix}
\sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1d} \\
\sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2d} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{1d} & \sigma_{2d} & \cdots & \sigma_{dd}
\end{bmatrix}
\]
Figure 4.1: Illustration of a Gaussian process. The colour indicates the probability density $P(x)$ that a given function $\tilde{x}(z)$ drawn from the Gaussian process would pass through the point $(z, x(z))$. Each vertical slice of this distribution has the form of a Gaussian, as shown in the two insets. Each of these insets corresponds to the slice indicated by the neighbouring dashed line. The Gaussian distribution within every vertical slice is parametrized by the mean and standard deviation functions $\mu(z)$ and $\sigma(z)$. The blue line shows $\mu(z)$ for this Gaussian process.

Now, we could already choose the number of dimensions to be infinite, but this would still be a countably infinite number of dimensions. Instead, let us now promote $\vec{x}$ from having components with a discrete index to those with a continuous index. That is, the number of dimensions of the joint distribution is now uncountably infinite.

This is rather like the difference in the matrix representations of operators that act on bound states and on continuum states in quantum mechanics. The hydrogen atom potential has an infinite number of discrete bound states, but it also has a continuum of scattering states [167]. This continuum of scattering states has a Hamiltonian whose “matrix” representation has an uncountably infinite number of dimensions, while the bound state Hamiltonian has a matrix with a countable, yet infinite, number of dimensions. Ultimately, the components $x_i$ of $\vec{x}$ change from being labelled by $i \in \mathbb{Z}$ to $z \in \mathbb{R}$, such that we have $x(z)$. 

81
Thus, we now have a situation where there is a Gaussian distribution over $x(z)$ at every point $z$ on the real line, with correlations in the joint probability distributions between points $z$ and $z'$. This can now be called a Gaussian process. Figure 4.1 gives an illustration of a Gaussian process with the insets showing the Gaussian distributions along a vertical slice of the process, while Figure 4.2 gives an alternative representation of the same Gaussian process. Alternatively, we can consider $x(z)$ as no longer describing a randomly chosen vector, but a randomly chosen function. We may then be tempted to write the following probability distribution over this function: $P(x) \propto e^{-\frac{1}{2} \int dz'dz \left[ (x(z) - \mu(z))^2 \Sigma^{-1}_{\text{cov}}(z,z') (x(z') - \mu(z')) \right]}$, which would be characterized by the parametrizing functions $\mu(z)$, and $\Sigma_{\text{cov}}(z,z')$. Unfortunately, while potentially intuitive, this is ill-defined, even if functional forms of this sort are useful in quantum field theory [92].

At the moment, however, the exact definition is less important than the notion that the Gaussian process represents an ensemble of functions, whose ensemble mean is $\mu(z)$ and whose correlations are governed by $\Sigma_{\text{cov}}(z,z')$. That is, a process can be thought of as a probability distribution over a class of functions; the random variable is a function, rather than a vector or scalar. More precisely and more generally, we can define the Gaussian process to be “a collection of random variables, any finite number of which have a joint Gaussian distribution” [165]. In the situation we are describing, this collection of random variables is uncountably infinite and indexed by a real number. While this formal definition may appear unintuitive, it is made clearer in light of the discussion above. So far, we have described a stochastic process that is a distribution over a singly-valued function; however, processes for multi-valued functions can also be defined.

It happens that the correlation functions $\Sigma_{\text{cov}}(z,z')$ correspond to the so-called “kernels” that appear in the context of the Gaussian process models discussed in the next section [45]; as such, I will often refer to the correlation functions as kernels for the remainder of this thesis. Furthermore, there are many such kernels that are used to define Gaussian processes, each of which bestows particular properties on the Gaussian process. For example, the squared exponential kernel ensures that the stochastic process itself is infinitely mean square differentiable [165]. This mean square derivative can be thought of as the generalization of differentiation to stochastic processes [165]. Figure 4.3 illustrates functions drawn from Gaussian processes defined with various kernel functions.
4.3.2 Gaussian Process Model

One application of Gaussian processes is to interpolation. While the true underlying approach to this type of interpolation is that of Bayesian inference, I will give an intuitive description of the results and refer to Rasmussen and Williams [165] for a more detailed explanation.

Let us assume that we have a set of \( n \) data points \( D = \{x_i, y_i\} \) between which we want to interpolate. As such, we are assuming that there is some functional relationship between \( x \) and \( y \). We will hold few assumptions about this function, but let us assume that it is smooth.

Now, let us take an initial Gaussian process, with mean \( \mu(x) = 0 \) and the square exponential kernel
\[
\Sigma_{\text{cov}}(x, x') = e^{-\frac{|x-x'|^2}{2l^2}}.
\]
This kernel describes correlations in the process that decay with a Gaussian profile (and length scale \( l \)) [165]. See the upper left panel of Figure 4.3 for examples of functions that can come from a Gaussian process defined with this kernel. Presumably, there exists a drawn function that approximates the underlying relationship in the data. However, our goal is not to determine this function. Instead, we will use the mean of the Gaussian process as the prediction and fit the process as a whole to the data. This fitting will narrow the distribution over the functions near the data points, but leave it broad in regions with little data. In this way, we will obtain both a reasonable estimate of the underlying functional relationship.
Figure 4.3: Demonstration of the different functions that can arise when drawn from Gaussian processes that are defined with different kernels. Note that while each Gaussian process has $\mu(z) = 0$ and $\sigma(z) = 1$, the behaviour of each drawn function varies drastically between the different kernels. The mathematical form of the square exponential kernel, the Matérn kernels, and the exponential sine squared kernel can be found in Ref. [165]. Note that the exponential sine squared kernel is also known as the periodic kernel [168].

Before performing this fitting procedure, we initially have some data $D$ and a presumed initial distribution over the functions (our prior). This data and the initial distribution (i.e. the Gaussian process) are shown in the upper left panel of Figure 4.4. As we are generating the data from a known functional relationship between $x$ and $y$, I also plot this relationship for reference. Let us first perform a fit with only two data points.

Now, assuming no noise in the system, we know that the true function must pass through the data points. As such, our fitting procedure forces $\mu(x_3) = y_3$ and $\mu(x_4) = y_4$ and removes any uncertainty in the model at $x_3$ and $x_4$. This results in the upper right panel of Figure 4.4. I refer to Rasmussen and Williams for the
mathematical details behind this fitting procedure [165], as they are not needed for the present discussion. One can see that the model is beginning to approximate the underlying function, albeit extremely poorly. Furthermore, there is little uncertainty in the process near \((x_3, y_3)\) and \((x_4, y_4)\), while the uncertainty grows as one moves away from each of these two points.

If we perform the fitting procedure with more data points, we obtain a much better fit; see the bottom panel of Figure 4.4. While there is still uncertainty between the data points, it is small. Interestingly, the model does include some extrapolation past the region spanned by the data points. However, the uncertainty grows rather quickly outside of this region; the influence of the last data point is decaying as a function of distance. Thus, the model is telling us where we can be more confident in its prediction (there is small uncertainty in the model) and where we should be less confident (there is large uncertainty in the model).

We have now made a Gaussian process model for our collection of data. If we were to add more data points, the model would become increasingly more accurate. As the kernel parameter \(l\) is varied during the fitting procedure to improve the fit, we also receive an automatic measure of a relevant length scale in the data. Furthermore, we are not restricted to modelling functions of a single variable, but can perform this analysis with high dimensional data rather efficiently, even to around 100 dimensions [169, 170]. The number of data points needed for a good fit actually scales linearly in the number of dimensions [171], making the method even more efficient for larger dimensions (for up to around 100 dimensions [169, 170]).

### 4.3.3 Bayesian Optimization

In this section, I now wish to turn to the method of Bayesian optimization. We start with some function \(f\) with \(d\) input dimensions that we wish to optimize. One (simplified) algorithm for Bayesian optimization is as follows [45]:

1. Evaluate \(f\) at \(n\) points \(\mathbf{x}_i\), randomly chosen (see panel (i) of Figure 4.5)
2. Fit a Gaussian process model to the set of data points \(\{\mathbf{x}_i, f(\mathbf{x}_i)\}\) (see panel (ii) of Figure 4.5)
3. Find the parameters \(\mathbf{x}_{\text{new}}\) that minimize the Gaussian process model, taking the uncertainty into account (i.e., \(\mathbf{x}_{\text{new}}\) minimizes the so-called “acquisition function”; see the red cross in panel (ii) of Figure 4.5)
4. Evaluate \(f(\mathbf{x}_{\text{new}})\) and add the result to the set of data points
Figure 4.4: An example of fitting a Gaussian process model. The upper left panel shows the true function (dashed line); the sample points (black circles), whose $x$ values are drawn via Latin hypercube sampling; and the unfit Gaussian process model, whose mean function $\mu(x) = 0$ (blue line) and whose standard deviation function $\sigma(x) = 1$. The shaded blue area encompasses $\mu(x) \pm 2\sigma(x)$. The upper right panel shows the Gaussian process model fit to two sampling points. The bottom panel shows the Gaussian process model fit to all five sampling points. Note that $\mu(x)$ is taken as the prediction of the Gaussian process model, while $\sigma(x)$ is taken as the uncertainty in the model at $x$.

5. Repeat from step 2 until $\vec{x}_{\text{new}}$ converges to a particular point (see panels (iii-xii) of Figure 4.5).

Figure 4.5 illustrates this process. There are a few key things to note about the above algorithm. First, instead of evaluating the computationally expensive function $f$ many thousands of times, as is often required for optimization, the algorithm is evaluating the computationally cheaper Gaussian process model thousands of times. (This occurs when minimizing the acquisition function in step 3.) The trade off is that we must re-perform this minimization every time we update the Gaussian process model.
Figure 4.5: Demonstration of the Bayesian optimization algorithm in one dimension. Panel (i) shows the initial preparatory step of evaluating the function \( f \) at seven initial points \( x_i \), chosen via Latin hypercube sampling. The dashed black line shows the true function, while the black circles are \((x_i, f(x_i))\). The blue line shows the mean of the unfit Gaussian process model, while the blue shaded area shows the uncertainty in the model by encompassing \( \mu \pm 2\sigma \). Panel (ii) shows the prediction and uncertainty of the Gaussian process model that is fit to the initial seven points. Also shown are the acquisition function \( a(x) = \mu(x) - 2.5\sigma(x) \) (dashed orange line) and the minimum \((x_{\text{new}}, a(x_{\text{new}}))\) of this acquisition function in the interval \([-20, 20]\) (red cross). Panel (iii) shows the same features, but with the Gaussian process model now fit to eight sample points. This eighth point is \((x_{\text{new}}, f(x_{\text{new}}))\), where \( x_{\text{new}} \) minimizes the acquisition function of the previous panel. Panels (iv-xii) show the continuation of the Bayesian optimization algorithm, where a new sample point \((x_{\text{new}}, f(x_{\text{new}}))\) is added every panel.
Secondly, taking the uncertainty into account when minimizing the model is a key step. The Gaussian process model gives us not only an interpolation, but also information about the uncertainty of that interpolation. We can incorporate this uncertainty by finding the minimum not of the mean of the Gaussian process model, but of the acquisition function $a(\bar{x}) \equiv \mu(\bar{x}) - \kappa \sigma(\bar{x})$, where $\mu$ is the mean of the Gaussian process model, $\sigma$ its standard deviation, and $\kappa$ a tuning parameter. By including the model uncertainty in the minimization, we are improving the likelihood that we reach the global minimum; the algorithm is forced to explore regions of the parameter space where there is insufficient information to conclude the absence of the global minimum.

Finally, the flip side of incorporating this uncertainty is that we can ignore regions where there is sufficient information to conclude the absence of the global minimum. Thus, this algorithm evaluates the function at points more likely to provide useful information than would algorithms based on simple random or grid sampling. This reduces the number of required function evaluations and improves the efficiency of the algorithm.

To conclude this section, I will state that the algorithm I mentioned above describes only the core ideas behind Bayesian optimization. While following the above algorithm will work, there are various modifications one can make to improve the efficiency of the algorithm. For example, there are many possible forms that the acquisition function can take; what I mention above is known as the lower confidence bound (LCB) acquisition function. Other forms include expected improvement and probability of improvement acquisition functions [172]. One can also improve the fitting of the Gaussian process model to the point where it can be proven that the Bayesian optimization algorithm converges to the global minimum, though this makes the algorithm more complex [173].

### 4.4 Preliminary Results

In this section, I will describe our preliminary application of Bayesian optimization to the problem of determining the scattering matrix from the MHI experimental data. We are particularly considering experiments that utilize ortho-hydrogen as the probe molecule, and thus working with a $9 \times 9$ scattering transfer matrix.

In the ideal case, we would like to obtain all 81 complex matrix elements from the experimental data. This is an ambitious goal, however, and we begin by simply considering diagonal scattering matrices, which
describe elastic scattering. These matrices have up to nine possible parameters, the nine variables $\theta_i$ that define the phases of the nine diagonal elements $e^{i\theta_i}$. The RP matrices used for the first rows of Figures 3.6 and 3.7 are of this form, for example.

However, even determining these nine parameters is not straightforward. As such, we do not work with experimental data immediately, but rather choose a set of target parameters $\theta_i^\ast$. We take these target phases and calculate the expected signal. Then, we attempt to determine the parameters from this signal (i.e., we “pretend” that this signal is experimental data). This approach allows us to tune the algorithm, to diagnose whether the Gaussian process model fit is well-behaved and stable, and to know whether the algorithm is reliably converging to the correct values. Once we are confident that the algorithm is working under these known conditions, we can work with the real experimental data, where there will be additional issues to be dealt with, such as noise in the signal.

Our computational setup is the same as that presented in Section 3.7 and has the field profile shown in Figure 3.3. We choose $B_1 = 400$ gauss. To determine the scattering matrices, we match the calculated signal to the “experimental” signal. That is, we minimize the MSE between the two curves. Note that we work with the MSE instead of the root mean square error (RMSE) as the RMSE has a cusp that occurs at the target values. That is, near $\vec{\theta}^\ast$, the RMSE $\sim |\vec{\theta} - \vec{\theta}^\ast|$, which produces difficulties when fitting the Gaussian process model. Instead, near the target value $\vec{\theta}^\ast$, the MSE $\sim (\vec{\theta} - \vec{\theta}^\ast)^2$, which is differentiable at least once and improves the stability of the fitting procedure. One disadvantage is that the MSE is flatter near the global minimum, which increases the number of algorithmic steps required to converge to a particular precision. Additionally, we calculate $P_{\text{detection}}$ at only ten values of $B_2$ instead of the usual $\sim 300$ values, in order to expedite this refinement stage of the algorithm.

For the Bayesian optimization algorithm itself, the kernel is of the form $\Sigma_{\text{corr}}(\vec{\theta}, \vec{\theta}') = C(\vec{\theta} - \vec{\theta}') \times \text{Matérn}_\nu(\vec{\theta} - \vec{\theta}')$, with the parameter $\nu = 1.5$. This includes the Matérn kernel multiplied by the constant kernel $C(\vec{\theta} - \vec{\theta}')$. The Matérn kernel with $\nu = 1.5$ causes the stochastic process to be once mean square differentiable [165], while the constant kernel allows for scaling of the function range. We choose the Matérn kernel with $\nu = 1.5$ as the MSE is also differentiable once at the global minimum. We tried various other forms of the kernel, but obtain the most reasonable performance of the algorithm with the above kernel, so far; it is likely that the above choice of kernel is still not optimal. We select the $N$ initial random points
\( \bar{\theta}_j \) via Latin hypercube sampling [174]. Latin hypercube sampling shows a greater uniform sampling of the parameter space, in comparison to drawing the points according to a uniform distribution. This greater sampling uniformity reduces the amount of redundant information arising from sample point clustering, which is likely when sampling from a uniform distribution.

In order to proceed in a relatively systematic way, we first applied Bayesian optimization to determine a scattering matrix parametrized by a single phase \( \theta_1 \), with all other phases held constant. As this met with success, we increased the number of parameters to two (\( \theta_1 \) and \( \theta_2 \)); this also met with success. Given these successes, we now increase the number of parameters to three and present the results here.

As the underlying core of the Bayesian optimization algorithm is the Gaussian process model, we visualize this model to ensure that the obtained fits are reasonable. As the mean of the Gaussian process is a function of three variables, we plot the mean on slices of the full parameter space. We choose each slice to pass through the known target phases \((\theta_1^*, \theta_2^*, \theta_3^*)\). In the case where the target phases are unknown, we would choose these slices to pass through the current best estimate of these parameters. In Figure 4.6, we show the mean of a Gaussian process model for an optimization run where the (randomly chosen) target phases are \((\theta_1^*, \theta_2^*, \theta_3^*) \approx (3.468, 2.482, 4.488)\); the remaining six phases are set to zero. To obtain this plot, we choose 300 initial sample points and perform 55 steps of the Bayesian optimization algorithm, which results in a total of 355 evaluations of the MSE. We set the LCB acquisition function tuning parameter \( \kappa = 0.01 \), in line with the values used in Ref. [64]. One can clearly see that the mean of the Gaussian process model is rather smooth and that the minimum of the model is near the known target phases \((\theta_1^*, \theta_2^*, \theta_3^*)\), shown by the orange circles.

In addition to plotting the mean of the Gaussian process model, we can monitor the progress of the algorithm by looking at the RMSE and the determined phases \((\theta_1, \theta_2, \theta_3)\) as a function of the algorithm step. We show these observables in Figure 4.7. One can see that the algorithm converges rather quickly, with a reasonable RMSE of 0.31% between the target and calculated signals occurring after just 20 optimization steps. The determined phases also converge rather quickly, with the phases determined to be approximately \((3.472, 2.494, 4.484)\) after 30 optimization steps. The relative errors between these values and the target values are \((0.11\%, 0.46\%, -0.07\%)\), presumably well within expected experimental errors. These results show that we are able to at least partially determine the scattering matrix from a given signal. Furthermore,
Figure 4.6: Gaussian process model of the mean square error between a calculated and a target $P_{\text{det}}$ curve as a function of the three phases $\theta_1, \theta_2, \theta_3$. These angles are the phases of a $9 \times 9$ diagonal scattering matrix whose diagonal elements $\Sigma_{jj}$ are of the form $e^{i\theta_j}$, with $\theta_4$ to $\theta_9$ set to zero. The goal is to find the phases $\theta_1, \theta_2, \theta_3$ that minimize the mean square error and thereby determine the target scattering matrix. Each subplot of this panel is a 2D slice of the 3D surface, with each slice passing through the known target phases $(\theta_1^*, \theta_2^*, \theta_3^*)$. The orange circles indicate the known target phases. The Gaussian process model is fit with 300 randomly selected points and 55 points selected by the Bayesian optimization algorithm.
Figure 4.7: Left panel: The phases as determined by the Bayesian optimization algorithm as a function of the number of algorithm steps. The dashed lines indicate the known target phases $\theta_1^*, \theta_2^*, \theta_3^*$. Right panel: The root mean square error between ten points of the target $P_{\text{detection}}$ curve and the $P_{\text{detection}}$ curve calculated with the determined phases (as indicated in the left panel), for each step of the Bayesian optimization algorithm.

we are able to determine the target phases to an error of less than 0.5% with just 330 function evaluations. This is significantly more efficient than a simple $10 \times 10 \times 10$ grid search, which could at most determine each phase to a resolution of $2\pi/10 \approx 0.63$.

Interestingly, the algorithm as presented does not always perform this well. Under identical conditions, except for another set of target phases $(3.190, 5.808, 3.002)$, it takes 760 steps of the algorithm (a total of 1060 function evaluations) to reach a RMSE of 1.7% between the calculated and target signals, with the relative errors in the determined phases ($0.65\%, 1.05\%, 0.10\%$). We are still determining the reason for this relatively poor performance (though it is still significantly better than a grid search) and still determining which of these two examples are reflective of the typical behaviour of the algorithm as presented. Certainly, more tuning of the algorithm is required. This may include modifying $\kappa$, selecting a different kernel, or changing the Gaussian process model fitting procedure. Indeed, one can modify the Gaussian process model fitting procedure to ensure that the algorithm always converges to the global minimum [173], which may be an avenue worth exploring.

We have also tried applying the algorithm to problems with larger numbers of unknown phases, but have obtained no successful results so far. We will first understand and improve the behaviour of the algorithm for problems with three unknown phases before again applying the algorithm to larger numbers of phases.
Once we have tuned the algorithm to perform well at determining the known target phases, we will attempt to fit the diagonal scattering matrix to true experimental data.

Interestingly, we may not need to determine nine phases from the experimental signal, as symmetries in the molecule-surface interaction may allow us to work with only three phases. That is, the primary interaction of an $o\text{H}_2$ molecule with the surface is steric. This would modify the rotational degrees of freedom while leaving the nuclear spin degrees of freedom unchanged, such that $\Sigma_{-1m_J} = \Sigma_{0m_J} = \Sigma_{1m_J}$, where the fist index corresponds to $m_J$. Finally, while we do not know if a diagonal matrix is the correct form of the scattering matrix, the results, whether positive or negative, will guide us towards the best next steps.

### 4.5 Conclusion and Future Directions

In this chapter, we have discussed our initial applications of Bayesian optimization to the inverse scattering problem for MHI experiments. The goal of this inverse scattering problem is to determine the scattering matrix elements from the experimental signal. Our approach to this problem is to minimize the difference between the experimental signal and the results of the transfer-matrix computations by varying the elements of the scattering matrix. We performed this minimization efficiently by using the Bayesian optimization algorithm to intelligently vary the scattering matrix elements. In this way, we have successfully applied Bayesian optimization to determine the scattering matrix corresponding to a target calculated signal, when that matrix is diagonal and determined by three or fewer parameters.

Building on this foundation, we will tune the algorithm to improve performance and then apply it to situations where the diagonal scattering matrix is defined by all nine phases. In addition to these diagonal scattering matrices, we will also treat different classes of scattering matrices, such as the **RDA**, **ROM** and **RUM** matrices discussed in Section 3.7.2, as well as fully random matrices. Also, all of the work discussed in this chapter has assumed that the scattering matrices are constant as a function of incident energy. This is known to not be the case and we may expand each scattering matrix element $\Sigma_{ij}(E)$ in a Taylor series; or, a Laurent series, if we expect the presence of some resonances.

Furthermore, there is certain to be some experimental error or noise in the measured signals and it is important to take this into account. As such, we will determine how errors in the experimental signal affect
our determination of the scattering matrix parameters. Indeed, given the complexities involved, we may only be able to determine the parameters to low precision or only determine the classes of matrices that fit the experimental data. Even if we can only determine classes of matrices, however, this would still give us information about what symmetries in the molecule-surface interaction are preserved or broken.

Additionally, we may explore other types of optimization algorithms. If we are able to analytically determine the gradient of the MSE from the transfer matrix formalism discussed in Chapter 3, we could take advantage of gradient-based methods and achieve improved performance [175, 176].

Finally, we may attempt a full Bayesian approach to this problem and begin with a prior distribution over the scattering matrix parameters, which can be updated via Bayesian inference, given the experimental data [177]. This would allow us to place error bars on the scattering matrix parameters.

The work presented in this chapter serves as the first step on the path to determine scattering matrices defined by more than three parameters and to use Bayesian optimization to determine the properties of the scattering matrix elements compatible with a given experimental measurement. These scattering matrix properties can then be used to gain physical insight into molecule-surface interactions and surface properties and used to test approximations used in \textit{ab initio} calculations.
4.6 Summary

This section contains a summary of the key results presented in Chapter 4:

- Applied Bayesian optimization to determine a scattering matrix parametrized by three phases, from data generated from known target parameters (Section 4.2)

- Demonstrated that Bayesian optimization applied to the inverse scattering problem for MHI experiments is significantly more efficient than a grid search with comparable resolution, for a three dimensional parameter space (Section 4.2)

- Demonstrated that tuning of the Bayesian optimization algorithm is required to improve performance for the inverse scattering problem (Section 4.2)

- Laid the foundation to efficiently determine, via Bayesian optimization, matrices defined by more than three parameters
Chapter 5

Monte Carlo enhanced with Bayesian model calibration

5.1 Introduction

In this short chapter, I propose the application of Bayesian model calibration to reduce the negative impact of critical slowing down and/or the sign problem in classical and quantum Monte Carlo calculations. In Section 5.2, I discuss Monte Carlo calculations and some of their limitations. In Section 5.3, I give a non-technical description of Bayesian model calibration. Then, in Section 5.4, I discuss a simplified version of Bayesian model calibration that is less powerful, but easier to implement. In Section 5.5, I provide a proof of principle demonstration that Bayesian model calibration can correct the values of poorly converged Monte Carlo calculations with the information from only a few well-converged calculations. I conclude and discuss future applications of Bayesian model calibration in Section 5.6 and give a chapter summary in Section 5.7.

5.2 Monte Carlo Calculations

Since the landmark paper by Metropolis et al. in 1953 [178], Markov chain Monte Carlo calculations (henceforth referred to as Monte Carlo calculations) have been widely used throughout science and engineering. In terms of complex quantum systems, these Monte Carlo calculations have been used to determine the properties of superfluids [71, 179, 180], calculate the electronic structure of atoms, molecules, and solids [181–183], study the Hubbard model [73, 184], and investigate Fermi-polarons [185, 186], for example.
Despite their massive success, however, Monte Carlo calculations do have some, rather famous, limitations. For example, when calculating observables near a second-order phase transition, basic Monte Carlo methods exhibit what is known as critical slowing down [187]. This occurs as sequential Monte Carlo samples become too strongly correlated near the phase transition. More specifically, for Monte Carlo calculations the variance of an observable often scales as $(N)^{-\frac{1}{2}}$, where $N$ is the number of Monte Carlo samples [188]. However, $N$ properly counts the number of independent Monte Carlo samples. Near a second-order phase transition, the correlation length of the system increases, which also causes an increase in the number of sequential Monte Carlo samples that are correlated (that is, the autocorrelation time increases) [189, 190]. Hence, many more Monte Carlo samples must be generated to obtain reasonable accuracy; this can quickly become prohibitive.

Another major limitation of Monte Carlo methods is known as the sign problem. For example, in diagrammatic Monte Carlo, the algorithm calculates observables by generating and summing a series of Feynman diagrams [191]. Unfortunately, for fermionic systems these diagrams often have an oscillating sign and nearly cancel each other out in the sum [74]. As such, all, or almost all, diagrams of the same order need to be summed to obtain an accurate result. (Roughly, the diagram order is the number of vertices in the diagram.) However, the number of diagrams in a given order $n$ scales as $n!$ and calculations that do not converge at low diagram orders are difficult to complete [74, 192].

Many different approaches have been made to tackle these limitations of Monte Carlo methods. Approaches to the issue of critical slowing down often involve updating the system state more efficiently, such as with cluster updates [189, 190] or loop algorithms [184]. An approach to the sign problem in diagrammatic Monte Carlo involves using the Dyson equation to reformulate the summation of the bare Feynman diagrams into a self-consistent formulation that requires explicit summation over a smaller set of diagrams; this is bold diagrammatic Monte Carlo [74, 185, 186].

In summary, as one nears a second order phase transition or works with a fermionic system, it can become difficult to reduce the uncertainties in the calculated observables because of the large number of Monte Carlo samples or diagram orders required; to then map out these observables as a function of the input parameters is prohibitive. However, these difficulties are only true for well-converged calculations. Calculations with smaller numbers of samples or including only low order diagrams can still be performed.
easily, albeit with an inaccurate final answer. At this point, we ask: *Instead of trying to converge the Monte Carlo calculations by increasing the number of samples, can we just correct the poorly converged calculations in an efficient manner?* Here, we propose to do this correction with machine learning techniques.

In recent years, machine learning techniques have been applied to improve the efficiency of various numerical methods. For example, restricted Boltzmann distributions have been used to learn effective Hamiltonians that are then used to propose more efficient updates in classical Monte Carlo calculations [77, 78]. Also, Cui and Krems have used Bayesian model calibration to improve the efficiency of scattering calculations by combining quantum scattering calculations with classical scattering calculations [60]. That is, classical calculations are used to calculate scattering cross sections at many points of the parameter space as they are computationally low-cost. Then, with just a few computationally expensive quantum calculations, Cui and Krems correct the classical calculations and obtain a reasonably accurate prediction of the scattering cross section as a function of the input parameters.

Here, we propose to use Bayesian model calibration [193] to partially alleviate both the sign problem and critical slowing down in Monte Carlo calculations by using Bayesian model calibration to correct easily-obtained, but poorly-converged, Monte Carlo calculations with the information gleaned from a small number of difficult, but well-converged, calculations. I give a non-technical overview of Bayesian model calibration in the next section.

### 5.3 Bayesian Model Calibration

Bayesian model calibration (BMC) is a machine learning technique based on the Gaussian process model discussed in Section 4.3.2. BMC was proposed in 2001 by Kennedy and O’Hagan [193] and was originally formulated to combine experimental measurements and computer model predictions in order to improve model predictions, even if not all important physical factors were taken into account in the computer model. Note that Kennedy and O’Hagan proposed two similarly-named and closely-related methods in the same paper [193]: Bayesian calibration and Bayesian model calibration. While both of these methods use the same combination of Gaussian process models, the goal of Bayesian calibration is to determine the parameters that maximize the accuracy of a mathematical model (while including error correction) and the goal of BMC is to correct the errors of a given mathematical model.
BMC was originally proposed to handle a scenario like the following \cite{193}: We have measurement data, such as of the NO\textsubscript{x} concentrations in the atmosphere over the course of a year, and we would like to predict these concentrations in future years. Unfortunately, the best mathematical model we have is a poor fit to the measurement data and we do not know of any physically-motivated ways to improve the fit. However, we do know that the model must correlate with the physical situation to some extent, given that the model is physically motivated. We then exploit this knowledge by (1) fitting a Gaussian process model to the correlations between the predictions of the mathematical model and the measurement data and then (2) using the Gaussian process model to correct the predictions of the mathematical model.

Kennedy and O’Hagan \cite{193} showed how to perform such a correction and obtain information about the uncertainty in the final prediction. One begins by training two Gaussian process models simultaneously, one to model the mathematical model and another to model the difference between the mathematical model and the experimental data. We then have the following relationship:

\[ Y(\cdot) = \rho F(\cdot) + G(\cdot), \]  

(5.1)

where \( Y(\cdot) \) is the final prediction, \( F(\cdot) \) is the Gaussian process model that describes our mathematical model, \( G(\cdot) \) is the Gaussian process model that describes the difference between the mathematical model and the experimental prediction, \( \cdot \) represents the input parameters, and \( \rho \) is a tuning parameter.

In principle, we can treat these two Gaussian process models in a fully Bayesian manner, such that we can rigorously determine a final estimate of the uncertainty in the prediction and use the data to determine all hyperparameters, including \( \rho \) \cite{193}. However, because this can be rather complex to implement, we first implement a simplified version of this algorithm. I discuss this simplified algorithm in the next section.

### 5.4 Simplified Bayesian Model Calibration

Ideally, we would like to implement the full Bayesian form of BMC. However, for the present work, we have implemented a simplified version that is both easier and faster to program. As the core idea of BMC is for the algorithm to learn the correlations between a mathematical model and experimental data, we have chosen to implement only this feature and neglect the rest. That is, instead of the full Bayesian treatment where both the mean and standard deviations of the two Gaussian process models are properly taken into
account and the two Gaussian process models are trained simultaneously, we only take the means of the Gaussian process models into account and train the two Gaussian process models individually.

Thus, in our simplified Bayesian model calibration (sBMC) method we set $\rho = 1$ in Eqn. (5.1) and first fit only $F(\cdot)$ to the numerical model. We then fit $G(\cdot)$ to $\chi(\cdot) = y(\cdot) - \mu_F(\cdot)$, where $y(\cdot)$ represents the experimental data and $\mu_F$ is the mean of the Gaussian process $F$. Finally, we define the prediction of sBMC as $Y(\cdot) = \mu_F(\cdot) + \mu_G(\cdot)$, where $\mu_G$ is the mean of the Gaussian process $G$. While this is a rather crude approach and any final results should rely on the full BMC method, this sBMC method has the advantage of simplicity and the ability to still provide proof of principle for our ideas.

Instead of working with a mathematical model and experimental data, as did Kennedy and O’Hagan [193], we will work with the data from poorly converged Monte Carlo calculations and well-converged Monte Carlo calculations, respectively. As both kinds of calculations are physically motivated, there should be a correlation between the inaccurate poorly converged data and the accurate well-converged data. Our goal is to learn these correlations at just a few points in parameter space and then use these correlations to correct the poorly converged calculations in the remainder of the parameter space of interest.

To enact this sBMC correction scheme, we do the following:

1. Perform several poorly-converged calculations that span the parameter space of interest
2. Fit a Gaussian process model $F(\cdot)$ to these data points
3. Perform a sparse set of well-converged calculations in the same parameter space
4. Fit a Gaussian process model $G(\cdot)$ to the difference between $\mu_F(\cdot)$ and the well-converged data points
5. Calculate the final prediction $Y(\cdot) = \mu_F(\cdot) + \mu_G(\cdot)$

In this way, we obtain a Gaussian process model $F(\cdot)$ of the poorly-converged calculations and a Gaussian process model $G(\cdot)$ of the difference between $\mu_F(\cdot)$ and the well-converged data points. Our final prediction consists of $\mu_F(\cdot)$, which models the poorly converged results, corrected by $\mu_G(\cdot)$, which models the correlations between the poorly- and well-converged calculations.
5.5 Proof of Principle Demonstration

We now apply this sBMC correction scheme to the bold diagrammatic Monte Carlo method described by Prokof’ev and Svistunov in Ref. [185]. One application of bold diagrammatic Monte Carlo that they describe in Ref. [185] is to the calculation of the s-wave scattering length of the spherical potential barrier. That is, the potential whose value is $U_0$ for $r < 1$ and zero otherwise, with $U_0 \geq 0$. As this problem has a known analytical solution, it is a useful test bed for the initial proof-of-principle application of sBMC. Also, Prokof’ev has made the relevant Fortran code for this problem publicly available [1]. Using this publicly-available code, we calculate the scattering length as a function of the potential strength $U_0$ and of the number of Monte Carlo steps included in the algorithm. We work with the number of Monte Carlo steps instead of the diagram order as the calculation of the s-wave scattering length of the spherical potential barrier involves only two diagrams in this version of the bold diagrammatic Monte Carlo algorithm.

While the details of the diagrammatic Monte Carlo algorithm are not needed for the present discussion, those interested in learning more about diagrammatic Monte Carlo can consult Prokof’ev’s excellent introduction to Monte Carlo methods [1]. Furthermore, Prokof’ev, Svistunov, and their co-authors have many excellent papers that describe diagrammatic Monte Carlo in significant detail for applications that include the Hubbard model [73], the unitary Fermi gas [75, 194], Holstein polarons [195], Dirac liquids [196], and Fermi-polarons [185] [186]. I find Refs. [73-75] particularly useful for understanding the method. Also, excellent introductions to many-body Feynman diagrams (on which diagrammatic Monte Carlo methods are frequently based) can be found in Chapters 4 and 5 of Condensed Matter Field Theory by A. Altland and B. Simons [92] and Chapters 3 and 7 of Quantum Theory of Many-Particle Systems by A. L. Fetter and J. D. Walecka [197].

We first calculate the s-wave scattering length $a_0$ for the spherical potential barrier with $1 \times 10^9$ Monte Carlo steps for 100 values of $U_0$ chosen via Latin hypercube sampling from the interval $[10^{-3}, 30]$ [174]. We then calculate the scattering length with $2 \times 10^{10}$ Monte Carlo steps for ten values of $U_0$ chosen via Latin hypercube sampling from the interval $[10^{-3}, 20]$. We compare the Monte Carlo calculations with the exact solution for the scattering length of the attractive spherical potential barrier [198]:

$$a_0 = 1 - \frac{\tanh \sqrt{2U_0}}{\sqrt{2U_0}}, \quad (5.2)$$

101
where $U_0 \geq 0$. The relative absolute difference between the Monte Carlo calculations and the exact answer (5.2) are shown in Figure 5.1 for the low Monte Carlo step count results (blue) and the high step count results (orange). It can be seen that the well-converged results show a more consistent low error than the poorly converged results. While there are regions where the poorly-converged calculations appear to be more accurate than the well-converged calculations, the convergence of these diagrammatic Monte Carlo calculations is not always monotonic.

We then run the sBMC algorithm with the many low step count results and the few high step count results. The relative absolute difference between the final prediction $Y(\cdot)$ and the exact result (5.2) is shown as the green curve in Figure 5.1. We see that the error of the sBMC result mostly resembles or improves upon the error of the high step count results, at least within the range spanned by the high step count results. We thus see how the computationally low-cost calculations are improved with comparatively few higher-cost, but more accurate, calculations. Clearly, however, this correction only occurs when interpolating between the sparse, well-converged calculations; in regions outside the span of the well-converged calculations, the error grows rapidly. Admittedly, there are regions where the error is higher than either the low or high step count results, which shows the imperfect nature of these results. Overall, however, we can see an improvement that justifies further refinement and exploration of this approach.

Importantly, we should improve the results by performing the full BMC algorithm, as opposed to the sBMC scheme used here. Particularly, the full BMC algorithm would provide an estimate of the uncertainty in the final prediction. However, in the absence of easily usable and publically-accessible codes that implement the full BMC algorithm (at least in the Python programming language), the sBMC algorithm can be used by researchers wishing to test the principles of the full BMC algorithm for their particular situation before investing the time to implement the full BMC algorithm. Any final results should rely on the full BMC algorithm, however.

As mentioned, the above results only provide a proof-of-principle result and are not yet a demonstrated algorithmic advance. However, our proposal can be extended to situations where there are more dependent variables and to situations where the exact answer is not known. In the absence of the exact result, it is important to determine ways to assess whether the BMC algorithm is indeed improving upon the poorly converged calculations. One simple way is to perform a few more well-converged calculations to compare...
Figure 5.1: The relative absolute difference between the numerical results and the exact scattering length (5.2) of a repulsive spherical barrier versus the height of the barrier $U_0$. The numerical results are obtained from a Monte Carlo calculation with $1 \times 10^9$ Monte Carlo steps (blue), a Monte Carlo calculation with $2 \times 10^{10}$ Monte Carlo steps (orange), and the simplified Bayesian model calibration result (green) discussed in Section 5.5. More details are provided in the text.

5.6 Conclusion

We have proposed applying Bayesian model calibration (BMC) to reduce the negative impact of critical slowing down and the sign problem in Monte Carlo calculations, whether classical or quantum. We have demonstrated an initial proof of principle calculation that we can use BMC to improve the accuracy of poorly converged Monte Carlo calculations with only a few well-converged calculations. In this demonstration, we applied the simplified Bayesian model calibration (sBMC) algorithm to the problem of calculating the scattering length of the repulsive spherical barrier via diagrammatic Monte Carlo. This initial success to the BMC predictions. Finally, as the Gaussian process models can be modified to take into account the uncertainty in the data points (i.e. the models can be fit to the error bars, not just the data points) [45], the BMC algorithm could be used to incorporate the variance of calculated observables. Thus, we need not throw away the statistical information generated by the Monte Carlo algorithm, but can include this information when applying the full BMC algorithm.
prompts us to implement the full BMC algorithm and to apply the algorithm to various other Monte Carlo calculations.

For example, we could apply the BMC algorithm to more sophisticated diagrammatic Monte Carlo schemes, such as those for the Hubbard model [73], angulons [199], or the Fermi-polaron problem [185] [186]. In these cases, we may approach the problem by correcting low Feynman diagram order calculations with a few high diagram order calculations. As another example, we could correct low bead number calculations with a few high bead number calculations in path integral Monte Carlo calculations [71] [72]. And, we could combine poorly and well-converged calculations for classical Monte Carlo calculations near a second order phase transition.

Moreover, it appears that BMC may be useful for many types of calculation, not only Monte Carlo calculations. Particularly, if there is a calculation with a tuning parameter that governs both accuracy and computational cost, the BMC algorithm may be able to improve calculation efficiency by using a few computationally high-cost, “well-tuned” calculations to correct several computationally low-cost, “poorly-tuned” calculations. Examples of tuning parameters include the bond dimension (or number of eigenstates retained) in density matrix renormalization group calculations [67–69], the number of baths in dynamical mean field theory [70], and the step or mesh sizes in the integration of differential equations. Finally, we need not only use data from a single type of calculation, but could combine data from different methods and/or analytical results in order to improve overall accuracy or interpolate between well-known limits. All of these applications are speculative, but warrant further investigation.
5.7 Summary

This section contains a summary of the key points presented in Chapter 5:

- Proposed applying Bayesian model calibration to improve the efficiency of Monte Carlo calculations in regions where the sign problem or critical slowing down are problematic (Sections 5.2-5.4)

- Provided a proof of principle demonstration that Bayesian model calibration can improve the accuracy of many poorly converged Monte Carlo calculations with the information from just a few well-converged calculations (Section 5.5)

- Discussed other types of calculations where Bayesian model calibration may improve calculation efficiency (Section 5.6)
Conclusion

6.1 Summary of Results

In this thesis, I have explored novel manifestations and applications of quantum interference in complex systems and developed new approaches to study complex quantum systems. We have uncovered new physics, modelled a new experimental system and proposed a way to improve the convergence of Monte Carlo methods.

In Chapter 2, we showed that particles whose long-range hopping amplitudes are isotropic can localize in disordered three dimensional lattices. We also showed that cooperative shielding, which mitigates the effects of long-range terms of a Hamiltonian \cite{30,119}, occurs in three-dimensional lattices for these particles. In contrast, if the long-range hopping amplitudes have the same anisotropy as the interaction between two dipoles aligned along a single axis, we see the absence of cooperative shielding and no concrete evidence of localized states. We have shown that the existence of delocalized, non-ergodic extended, or localized states in the presence of disorder depends on both the hopping range exponents $\alpha$ and the anisotropy of the long-range hopping amplitudes.

In Chapter 3, we proposed using molecules instead of $^3$He atoms in surface spin echo experiments. That is, instead of using the nuclear spin states of $^3$He to probe sample surfaces and their adsorbates, we proposed using the molecular hyperfine states of generic closed shell molecules, whose additional degrees of freedom may provide us with more insight into surface properties and dynamics. We also developed a fully quantum model of the experiment that involved re-deriving the transfer matrix formalism to include
the internal degrees of freedom. This formalism is needed to address the inverse scattering problem and to interpret these molecular hyperfine interferometry experiments. We also applied our model to the case of $\alpha\text{H}_2$ and showed that the experimental signal does indeed include information about the molecule-surface scattering matrix elements.

In Chapter 4, we applied Bayesian optimization to the inverse scattering problem for these experiments. Bayesian optimization models the function to be minimized with a Gaussian process model and uses the model, along with its uncertainty, to intelligently sample the function to be minimized. Such a sampling procedure significantly reduces the number of function evaluations required for global minimization. Using Bayesian optimization, we determined a scattering matrix defined by three parameters from data generated from known target parameters. We also demonstrated that the approach is significantly more efficient than grid searches with a comparable resolution.

In Chapter 5, we proposed using Bayesian model calibration to improve the efficiency of Monte Carlo algorithms in regions where the sign problem or critical slowing down inhibit the calculation of observables. Bayesian model calibration learns the correlations between several poorly converged calculations and a few well-converged calculations and uses these correlations to correct the computationally low-cost, but poorly converged results. We provided a proof of principle demonstration by applying a simplified version of Bayesian model calibration to a simple diagrammatic Monte Carlo calculation of the scattering length of a spherical potential barrier.

In conclusion, this thesis has improved our understanding of the quantum localization of particles with long-range hopping, provided the theoretical framework needed to interpret molecular hyperfine interferometry experiments, began to address the inverse scattering problem for these experiments by applying Bayesian optimization, and proposed an application of Bayesian model calibration to Monte Carlo calculations to reduce the impact of the sign problem and critical slowing down.

6.2 Outlook

The results of this thesis motivate many possible future directions, for both experimental and theoretical work. This section contains a list, necessarily incomplete, of these possible directions.
6.2.1 Anderson Localization with Long-range Hopping

- Our analytical results appear to imply that cooperative shielding would also exist in two dimensional systems, but it is important to actually confirm this.

- Our work examined only two possible forms of angular dependence of the hopping amplitudes. It is important to examine how different forms of anisotropy impact cooperative shielding and localization. Different lattice geometries may also impact cooperative shielding and localization. Furthermore, the impact of the lattice geometry may also be a function of the anisotropy and vice versa.

- In a recent experiment, researchers could tune the anisotropy of interactions between atoms dressed by Rydberg excitations, for atoms that are in a two dimensional optical or magnetic lattice [23]. This would be a useful test bed for the ideas presented in Chapter 2.

- The anisotropic form studied in Chapter 2 was that of dipoles aligned along a principal axis of the lattice. This form arises from a model of polar molecules in an optical lattice that are aligned by an external electric field. One may be able to change the form of the anisotropy, or at least its impact on cooperative shielding and localization, by changing the angle of the aligning electric field.

6.2.2 Molecular Hyperfine Interferometry

- Our plan is to expand the fully quantum formalism to also account for the dynamics of surfaces that are on the molecule-surface interaction time scales. That is, we plan to incorporate a time-dependent scattering matrix, which may require the numerical evaluation of the energy integrals and/or coupling between transport channels of differing energy.

- Our formalism can help design experimental protocols that produce a signal which may more clearly elucidate particular features of the scattering matrix or which may be easier to interpret.

- We plan to continue applying Bayesian optimization to the inverse scattering problem of molecular hyperfine interferometry experiments. We will work with different classes of scattering matrices and potentially add energy-dependence to the scattering matrices. Ideally, we will be able to determine all 81+ degrees of freedom of the scattering matrix from the scattering data, but even just finding classes
of matrices compatible with the experimental data will help to elucidate the symmetries at play in the molecule-surface interactions.

- Other types of optimization algorithms may help address this inverse scattering problem, such as gradient-based methods [175,176].

- We may be able to approach the inverse scattering problem from a more Bayesian perspective. That is, we could begin with a prior over the scattering matrix elements and propagate this uncertainty through the fully quantum model we have developed to get an uncertainty in the output signal. Then, we could compare this result to the experimental data and use Bayesian optimization to vary the parameters of the prior until the generated signal and its uncertainty match the experimental data and its uncertainty.

### 6.2.3 Monte Carlo Enhanced with Bayesian Model Calibration

- We plan to implement the full Bayesian model calibration algorithm, which we anticipate will provide better results than those obtained from the simplified Bayesian model calibration algorithm discussed in Chapter 5.

- We plan to apply the full Bayesian model calibration algorithm to many different types of Monte Carlo calculations and thus show the general applicability of the approach.

- We plan to apply Bayesian model calibration to Monte Carlo calculations in order to push into parameter regimes of important models where the convergence is currently too slow to calculate observables with small uncertainties, such as near phase transitions.

- Bayesian model calibration may also improve the efficiency of other types of calculations that have a tuning parameter that governs both accuracy and computational cost, such as the density matrix renormalization group [67–69] and dynamical mean field theory [70].
Bibliography


111


119


References


122

Appendix A

Supplementary material for Anderson localization for particles with long-range hopping

A.1 Outline of the Derivation for Equation 2.26

Here, I give an outline of the derivation of Eqn. 2.26, the eigenvalues of the Hamiltonian (2.24) for an ideal 3D cubic lattice with isotropic hopping of arbitrary range (2.22). We begin with the Hamiltonian (2.24) and evaluate

$$
\hat{H}^I_{\alpha} |k_1 k_2 k_3 \rangle = \frac{1}{\sqrt{V}} \sum_{i,j,l,m,n,\alpha,\beta,\delta = -N/2}^{N/2} e^{i \frac{2\pi}{N} (k_1 \alpha + k_2 \beta + k_3 \delta)} \frac{\zeta}{\alpha} (1 - \delta_{il} \delta_{jm} \delta_{hn}) \hat{c}_{\alpha \beta \delta}^\dagger \hat{c}_{ijh} \hat{c}_{\alpha \beta \delta}^\dagger |0\rangle,
$$

(A.1)

where

$$
|k_1 k_2 k_3 \rangle = \hat{c}_{k_1 k_2 k_3}^\dagger |0\rangle
$$

(A.2)

$$
= \frac{1}{\sqrt{V}} \sum_{\alpha,\beta,\delta = -N/2}^{N/2} e^{i \frac{2\pi}{N} (k_1 \alpha + k_2 \beta + k_3 \delta)} \hat{c}_{\alpha \beta \delta}^\dagger |0\rangle,
$$

(A.3)
\[ V \equiv N^3,\ r \equiv \sqrt{(l-i)^2 + (m-j)^2 + (n-h)^2}, \text{ and } |0\rangle \text{ is the vacuum state. Using the commutation relations} \]

\[ [\hat{c}_{lmn}, \hat{c}^\dagger_{ij\sigma}] = \delta_{li}\delta_{mj}\delta_{\sigma n} \]  

(A.4)

\[ [\hat{c}_{lmn}, \hat{c}_{ij\sigma}] = \begin{bmatrix} \hat{c}^\dagger_{lmn}, \hat{c}^\dagger_{ij\sigma} \end{bmatrix} = 0, \]  

(A.5)

and evaluating the sums over \( \alpha, \beta, \delta \) in Eqn. A.1, we get

\[ \hat{H}_\alpha |k_1k_2k_3\rangle = \frac{\zeta}{\sqrt{V}} \sum_{i,j,h,l,m,n} e^{i\frac{2\pi}{N}(k_1i+k_2j+k_3h)} \frac{1}{r_\alpha} (1 - \delta_{il}\delta_{jm}\delta_{hn}) \hat{c}^\dagger_{lmn} |0\rangle \]  

(A.6)

Then, to evaluate the eigenvalues (2.26), we multiply from the left by \( \langle k_1k_2k_3 | \) and obtain

\[ \langle k_1k_2k_3 | \hat{H}_\alpha | k_1k_2k_3 \rangle = \frac{\zeta}{\sqrt{V}} \sum_{i,j,h,l,m,n} e^{i\frac{2\pi}{N}(k_1(i-\alpha)+k_2(j-\beta)+k_3(h-\delta))} \langle 0 | c_{i\alpha\beta\delta} \hat{c}^\dagger_{lmn} | 0 \rangle \]  

(A.7)

\[ = \frac{\zeta}{\sqrt{V}} \sum_{i,j,h,l,m,n} e^{i\frac{2\pi}{N}(k_1(i-l)+k_2(j-m)+k_3(h-n))} \]  

(A.8)

Note that one can perform this reordering of the sum:

\[ \sum_{i,l=-N-1}^{-1} h(i-l) = \sum_{r_1=-(N-1)}^{N-1} (N - |r_1|) h(r_1), \]  

(A.9)

where \( r_1 = i - l, h \) is some function of \( i - l \), and \( (N - |r_1|) \) counts the number of times a particular value of \( r_1 \) appears in the double sum on the left-hand side.
We perform a similar reordering for Eqn. [A.8] and obtain

\[
\langle k_1 k_2 k_3 | \hat{H}_\alpha^l | k_1 k_2 k_3 \rangle = \frac{\zeta}{V} \sum_{r_1=-(N-1)}^{N-1} (N - |r_1|) \sum_{j,h,m,n=-(N-1)}^{N-1} \frac{1}{r_\alpha} (1 - \delta_{r_10} \delta_{jm} \delta_{hn}) e^{i \frac{2\pi}{N} (k_1 r_1 + k_2 (j-m) + k_3 (h-n))}
\]

(A.10)

\[
= \frac{\zeta}{V} \sum_{j,h,m,n=-(N-1)}^{N-1} e^{i \frac{2\pi}{N} (k_2(j-m) + k_3(h-n))} (N - |r_1|) \frac{1}{r_\alpha} (1 - \delta_{r_10} \delta_{jm} \delta_{hn}) e^{i \frac{2\pi}{N} (k_1 r_1)}
\]

(A.11)

\[
= \frac{\zeta}{V} \sum_{j,h,m,n=-(N-1)}^{N-1} e^{i \frac{2\pi}{N} (k_2(j-m) + k_3(h-n))} f (j-m, h-n),
\]

(A.12)

where \( r_1 = i - l \), now \( r = \sqrt{r_1^2 + (m-j)^2 + (n-h)^2} \), and

\[
f (j-m, h-n) \equiv \sum_{r_1=-(N-1)}^{N-1} (N - |r_1|) \frac{1}{r_\alpha} (1 - \delta_{r_10} \delta_{jm} \delta_{hn}) e^{i \frac{2\pi}{N} (k_1 r_1)}
\]

(A.13)

\[
= N \left( \frac{1 - \delta_{(j-m)0} \delta_{(h-n)0}}{(m-j)^2 + (n-h)^2} \right)^{1/2} + 2 \sum_{r_1=1}^{N-1} \frac{(N - r_1) \cos \left( \frac{2\pi}{N} k_1 r_1 \right)}{r_\alpha},
\]

(A.14)

where the first term of the second line corresponds to \( r_1 = 0 \) and the second term arises as \( \delta_{r_10} = 0 \) \( (r_1 \neq 0 \) in the second term).

Repeating the process for \( r_2 \equiv j - m \), we get

\[
\langle k_1 k_2 k_3 | \hat{H}_\alpha^l | k_1 k_2 k_3 \rangle = \frac{\zeta}{V} \sum_{h,n=-(N-1)}^{N-1} e^{i \frac{2\pi}{N} k_3 (h-n)} g (h-n),
\]

(A.15)

where

\[
g (h-n) \equiv \sum_{r_2=-(N-1)}^{N-1} (N - |r_2|) f (r_2, h-n) e^{i \frac{2\pi}{N} (k_2 r_2)}
\]

(A.16)

\[
= N f (0, h-n) + 2 \sum_{r_2=1}^{N-1} (N - r_2) f (r_2, h-n) \cos \left( \frac{2\pi}{N} k_2 r_2 \right).
\]

(A.17)
Finally, repeating the process again for \( r_3 = h - n \), we get

\[
\langle k_1 k_2 k_3 | \hat{H}_\alpha | k_1 k_2 k_3 \rangle = \frac{\xi}{V} \left[ N g(0) + 2 \sum_{r_3=1}^{N-1} (N-r_3) g(r_3) \cos \left( \frac{2\pi}{N} k_3 r_3 \right) \right].
\] (A.18)

We then have

\[
g(0) = N f(0,0) + 2 \sum_{r_2=1}^{N-1} (N-r_2) f(r_2,0) \cos \left( \frac{2\pi}{N} k_2 r_2 \right)
\] (A.19)

\[
= 2N \sum_{r_1=1}^{N-1} (N-r_1) \frac{\cos \left( \frac{2\pi}{N} k_1 r_1 \right)}{r_1^\alpha} + 2N \sum_{r_2=1}^{N-1} (N-r_2) \frac{\cos \left( \frac{2\pi}{N} k_2 r_2 \right)}{r_2^\alpha}
+ 4 \sum_{r_1=1}^{N-1} \sum_{r_2=1}^{N-1} (N-r_1) (N-r_2) \frac{\cos \left( \frac{2\pi}{N} k_1 r_1 \right) + \cos \left( \frac{2\pi}{N} k_2 r_2 \right)}{(r_1^2 + r_2^2)^{\frac{\alpha}{2}}}
\] (A.20)

as

\[
f(r_2,0) = N \left( 1 - \delta_{r_2,0} \right) + 2 \sum_{r_1=1}^{N-1} (N-r_1) \frac{\cos \left( \frac{2\pi}{N} k_1 r_1 \right)}{r_1^\alpha}
\] and

\[
f(0,0) = 2 \sum_{r_1=1}^{N-1} (N-r_1) \frac{\cos \left( \frac{2\pi}{N} k_1 r_1 \right)}{r_1^\alpha}.
\] (A.21)

\[
g(r_3) = N f(0,r_3) + 2 \sum_{r_2=1}^{N-1} (N-r_2) f(r_2,r_3) \cos \left( \frac{2\pi}{N} k_2 r_2 \right)
\] (A.23)

\[
= \frac{N^2}{r_3^\alpha} + 2N \sum_{r_1=1}^{N-1} (N-r_1) \frac{\cos \left( \frac{2\pi}{N} k_1 r_1 \right)}{(r_1^2 + r_3^2)^{\frac{\alpha}{2}}} + 2N \sum_{r_2=1}^{N-1} (N-r_2) \frac{\cos \left( \frac{2\pi}{N} k_2 r_2 \right)}{(r_2^2 + r_3^2)^{\frac{\alpha}{2}}}
+ 4 \sum_{r_1=1}^{N-1} \sum_{r_2=1}^{N-1} (N-r_1) (N-r_2) \frac{\cos \left( \frac{2\pi}{N} k_1 r_1 \right) \cos \left( \frac{2\pi}{N} k_2 r_2 \right)}{r_3^\alpha},
\] (A.24)
\[ f(r_2, r_3) = \frac{N}{(r_2^2 + r_3^2)^{\frac{\alpha}{2}}} + 2 \sum_{r_1=1}^{N-1} (N - r_1) \frac{\cos \left( \frac{2\pi}{N} k_1 r_1 \right)}{r_1^{\frac{\alpha}{2}}} \quad \text{and} \quad (A.25) \]

\[ f(0, r_3) = \frac{N}{r_3^{\frac{\alpha}{2}}} + 2 \sum_{r_1=1}^{N-1} (N - r_1) \frac{\cos \left( \frac{2\pi}{N} k_1 r_1 \right)}{r_1^{\frac{\alpha}{2}}}, \]  

(A.26)

where the fact that \( r_3 > 0 \) was taken into account (see last term of Eqn. A.18). We then substitute Eqns. A.24 and A.20 into Eqn. A.18. Finally, by expanding and collecting terms, we obtain Eqn. 2.26 after relabelling \( r_1, r_2, \) and \( r_3 \) to \( l, m, \) and \( n, \) respectively.
Appendix B

Supplementary material for molecular hyperfine interferometry

B.1 Schrödinger Equation for Eigenstate Coefficients

Using Eqn. 3.19 the time-independent Schrödinger equation is

\[ \hat{H} |E\tilde{R}\rangle = E |E\tilde{R}\rangle \]  
\[ \hat{\mathcal{K}} |E\tilde{R}\rangle = \left( E - \hat{\mathcal{H}}^R \right) |E\tilde{R}\rangle \]

\[ \sum_R \int dx \Phi^{ER}_R(x) \hat{\mathcal{K}} |xR\rangle = \sum_R \int dx \Phi^{ER}_R(x) \left( E - \hat{\mathcal{H}}^R \right) |xR\rangle \]

\[ \sum_R \int dx \Phi^{ER}_R(x) \langle x_0 R_0 | \hat{\mathcal{K}} |xR\rangle = \sum_R \int dx \Phi^{ER}_R(x) \langle x_0 R_0 | \left( E - \hat{\mathcal{H}}^R \right) |xR\rangle \]  

(B.2)

where \( \hat{H} \) is the total Hamiltonian (3.2) in the current region, \( \hat{\mathcal{K}} \equiv \frac{\mathbf{p}^2}{2m} \), we use \( \hat{\mathcal{H}}^R \) as a shorthand for \( \hat{\mathcal{H}^R}(\tilde{B}_{\text{loc}}) \), and the last line was multiplied by \( \langle x_0 R_0 \rangle \).
The different terms can be evaluated as

\[ \langle x_0 | \hat{K} | xR \rangle = \delta_{R_0R} \langle x_0 | \int \frac{\hbar^2 k^2}{2m} |k\rangle \langle k | x \rangle \]
\[ = \int \frac{dk}{2\pi} \delta_{R_0R} \frac{\hbar^2 k^2}{2m} e^{ik(x_0 - x)}, \tag{B.3} \]

\[ \langle x_0 | E | xR \rangle = \delta_{R_0R} \delta(x - x_0) E, \tag{B.4} \]

\[ \langle x_0 | \hat{H}^R | xR \rangle = \delta(x - x_0) H^R_{R_0}, \tag{B.5} \]

where \( |k\rangle \) is a momentum state with wavenumber \( k \), and \( m \) is the mass of the molecule. The additional factor of \((2\pi)^{-1}\) in Eqn. (B.3) comes from \( \langle x | k \rangle \equiv (2\pi)^{-\frac{1}{2}} e^{ikx} \). After inserting these three equations into Eqn. (B.2) and evaluating most of the sums, we obtain

\[ \int dx \int \frac{dk}{2\pi} \frac{\hbar^2 k^2}{2m} e^{ik(x_0 - x)} \Phi^E_{R_0}(x) = \Phi^E_{R_0}(x_0) E - \sum_R H^R_{R_0R} \Phi^E_{R_0}(x_0) \tag{B.6} \]

Noting that \( k^2 e^{ikx_0} = -\frac{\partial^2}{\partial x_0^2} e^{ikx_0} \) and \( \int \frac{dk}{2\pi} e^{ik(x_0 - x)} = \delta(x_0 - x) \), we obtain Eqn. (3.20) after the relabelling \( x_0 \to x \).

**B.2 Coefficient Relations Across a Discontinuity**

Since \( \Phi^E_R(x) \in C^1(x) \) for a specific value of \( R \) and given Eqn. (3.26), we get the defining equations for the continuity of the wavefunction as

\[ \lim_{x \to 0^-} \Phi^E_R(x) = \lim_{x \to 0^+} \Phi^E_R(x) \]

\[ \lim_{x \to 0^+} \sum_R \Phi^E_R(x) S^R_{R \to R^+} = \lim_{x \to 0^+} \Phi^E_R(x) \]

\[ \lim_{x \to 0^-} \sum_R S^R_{R^+ \to R} \left( A_{R^+} e^{ik_{R^+}x} + B_{R^+} e^{-ik_{R^+}x} \right) = \lim_{x \to 0^+} A_{R^+} e^{ik_{R^+}x} + B_{R^+} e^{-ik_{R^+}x} \tag{Eqn. (3.22)} \]

\[ A_{R^+} + B_{R^+} = \sum_R S^R_{R \to R^+} (A_{R^-} + B_{R^-}) \tag{B.7} \]

where \( S^R_{R^+ \to R} \equiv \langle R^+ | R^- \rangle \), \( k_{R^+} \equiv \sqrt{2m(E - E_{R^+})/\hbar} \), and \( E_{R^+} \equiv \langle R^+ | \hat{H}^R(\vec{B}(0^\pm)) | R^+ \rangle \). There are \( N_R \) such equations, one for each value of \( R^+ \).
Correspondingly, the defining equations for the continuity of the first derivative of the coefficients are

\[
\lim_{x \to 0^-} \frac{\partial}{\partial x} \Phi_{R}^{ER}(x) = \lim_{x \to 0^+} \frac{\partial}{\partial x} \Phi_{R}^{ER}(x)
\]

\[
\lim_{x \to 0^-} \sum_{R} S_{R}^{a}_{R} \frac{\partial}{\partial x} (A_{R} e^{ikR - x} + B_{R} e^{-ikR - x}) = \lim_{x \to 0^+} \frac{\partial}{\partial x} (A_{R} e^{ikR + x} + B_{R} e^{-ikR + x}) \quad \text{(Eqn. (3.22))}
\]

\[
\lim_{x \to 0^-} \sum_{R} S_{R}^{b}_{R} ikR (A_{R} e^{ikR - x} - B_{R} e^{-ikR - x}) = \lim_{x \to 0^+} ikR (A_{R} e^{ikR + x} - B_{R} e^{-ikR + x})
\]

\[
A_{R}^+ - B_{R}^+ = \sum_{R} S_{R}^{a}_{R} \frac{kR}{kR^+} (A_{R} - B_{R})
\]  

(B.8)

Solving Eqns. (B.7) and (B.8) for the coefficients \(A_{R}^+\) and \(B_{R}^+\), we obtain Eqns. (3.27) and (3.28).

### B.3 Computational Parameters Used for the Application to ortho-Hydrogen

We take the mean velocity \(v_0 = 1436.14 \text{ ms}^{-1}\) and the velocity spread to be 4 \%FWHM. When performing the integral of Eqn. (3.18), we take a \(k\)-space grid spacing \(\Delta k = 1 \times 10^4 \text{ cm}^{-1}\) and integrate from \(-7\sigma_k\) to \(+7\sigma_k\), where \(\sigma_k\) is the Gaussian width in momentum space as defined in Section 3.4. For the magnetic field profile and the angles between the two branches of the apparatus, see Figure 3.3. The relative probabilities used for the state selector probabilities \(P_{R_0}\) and the detector coefficients \(c_{RD}\) are given in Table B.1.

Where applicable, the parameters above were chosen to match those in the supplementary information of Godsi et al. [44].

**Table B.1:** Relative probabilities of the state selector \(\eta_{m_im_j}\) and the detector \(\kappa_{m_im_j}\). The state selector probabilities \(P_{R_0}\) are calculated as \(P_{R_0} = P_{m_im_j} \equiv \eta_{m_im_j} / \sum_{m_im_j} \eta_{m_im_j}\) and the detector coefficients \(c_{RD}\) are calculated as \(c_{RD} = c_{m_im_j} \equiv \kappa_{m_im_j} / \sum_{m_im_j} \kappa_{m_im_j}\).

<table>
<thead>
<tr>
<th>(m_I)</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>-1</th>
<th>-1</th>
<th>-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(m_J)</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(\eta_{m_im_j})</td>
<td>1.0000</td>
<td>0.9755</td>
<td>0.7901</td>
<td>0.1465</td>
<td>0.1111</td>
<td>0.0738</td>
<td>0.0343</td>
<td>0.0299</td>
<td>0.0258</td>
</tr>
<tr>
<td>(\kappa_{m_im_j})</td>
<td>1.00</td>
<td>0.96</td>
<td>0.93</td>
<td>0.53</td>
<td>0.42</td>
<td>0.37</td>
<td>0.21</td>
<td>0.19</td>
<td>0.16</td>
</tr>
</tbody>
</table>