

**Applications of machine learning for solving complex
quantum problems**

by

Rodrigo Alejandro Vargas–Hernández

B.Sc. Chemistry, Universidad Nacional Autónoma de México, 2012

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)

December 2018

© Rodrigo Alejandro Vargas–Hernández, 2018

The following individuals certify that they have read, and recommended to the Faculty of Graduate and Postdoctoral Studies for acceptance, a dissertation entitled:

APPLICATIONS OF MACHINE LEARNING FOR SOLVING COMPLEX
QUANTUM PROBLEMS

submitted in partial fulfillment of the requirement
for by Rodrigo A. Vargas-Hernández

the degree of DOCTOR OF PHILOSOPHY

in CHEMISTRY

Examining Committee:

PROFESSOR ROMAN V. KREMS, CHEMISTRY
Supervisor

PROFESSOR EDWARD R. GRANT, CHEMISTRY
Supervisory Committee Member

PROFESSOR MARK THACHUK, CHEMISTRY
Supervisory Committee Member

PROFESSOR JOERG ROTTLE, PHYSICS
University Examiner

PROFESSOR PIERRE KENNEPOHL, CHEMISTRY
University Examiner

Abstract

This thesis illustrates the use of machine learning algorithms and exact numerical methods to study quantum observables for different systems. The first part of this thesis depicts how to construct accurate potential energy surfaces (PESs) using supervised learning algorithms such as Gaussian Process (GP) regression. PESs have a leading part in quantum chemistry since they are used to study chemical reaction dynamics. Constructing the PES from quantum reactive scattering calculations, as the reaction probability, is known as the inverse scattering problem. Here, we illustrate a possible solution to the inverse scattering problem with a two-tiered GP model one GP model interpolates the PES and the second in Bayesian optimization (BO) algorithm. The end result is an accurate PES constructed with a GP with fewer points than with standard methods previously used for PES. BO is an optimization algorithm for black-box functions that use GP regression as an approximation of the interrogative function. We applied BO to find the optimal parameters of hybrid-density functionals. Quantum observables can differ between phases of matter. GP models with kernel combinations can extrapolate quantum observables such as the polaron dispersion energy between different phases and discover phases of matter. The same algorithm can predict quantum observables where standard numerical techniques lack convergence.

In the second half of the dissertation, we studied the evolution of quantum walks in various graphs with Hamiltonians permitting particle number changes. We showed that particle number-changing interactions accelerate quantum walks for any of the graph considered. Quantum simulators to study many-body physics is an active research field. We proposed the use of magnetic atoms trapped in optical lattices to experimentally mimic Bose-Hubbard type models by preparing atoms in different Zeeman states.

Lay Summary

In this dissertation, we used machine learning algorithms to reduce the computational resources needed to predict quantum observables such as the electronic energy of molecules, reaction probabilities, and energy dispersions. Next, we illustrate that machine learning algorithms, trained with data from a single phase, can extrapolate quantum observables and predict new phases of matter. We also introduced an optimization algorithm for black-box functions to solve problems like the inverse scattering problem or the optimization of hybrid-density functionals used in quantum chemistry. Additionally, we illustrate that the spread of quantum particles used in quantum computing, like in quantum search algorithms, can be enhanced if quantum particles are allowed to create/annihilate different particles. Last, we proposed an experimental simulator to study many-body quantum physics using Zeeman excitations of magnetic atoms trapped in an optical lattice.

Preface

Part of the material in Chapter 2 was published in the article: A. Kamath, R. A. Vargas-Hernández, R. V. Krems, T. Carrington Jr., and S. Manzhos, *Neural Networks vs Gaussian Process Regression for Representing Potential Energy Surfaces: a Comparative Study of Fit Quality and Vibrational Spectrum Accuracy*, J. Chem. Phys. **148**, 241702 (2017). The author did the GP models calculations to construct the PESs of the results published in the article.

Part of the material in Chapter 3 was accepted for publication in *New Journal of Physics*, arXiv:1711.06376 (2018): R. A. Vargas-Hernández, Y. Guan, D. H. Zhang, and R. V. Krems, *Bayesian optimization for the inverse scattering problem in quantum reaction dynamics*. This project was done in collaboration with the research group of professor D. H. Zhang at Dalian Institute of Chemical Physics (DICP), in Dalian China. The project was identified and designed by Roman V. Krems and the author. The author developed the Bayesian optimization algorithm and the code that combines BO with the DICP's library, which computes the reaction probabilities and the GP models for the PESs. The author did all the calculations and the data analysis.

Part of the material in Chapter 4 was accepted for publication in *Physical Review Letters*, arXiv:1803.08195 (2018): R. A. Vargas-Hernández, J. Sous, M. Bercui, and R. V. Krems, *Extrapolating quantum observables with machine learning: inferring multiple phase transitions from properties of a single phase*. The project was identified and designed by the author and Roman V. Krems. The polaron dispersions data was provided by J. Sous and M. Bercui. The author developed the code to construct the kernel combinations based on Refs. [58, 60] and carried all the calculations presented in the article.

The material in Chapter 5 and Appendix B was published in the article: R. A. Vargas-Hernández, and R. V. Krems, *Quantum walks assisted by particle number fluctuations*, Phys. Rev. A **98**, 022107 (2018). The project was identified and designed by the author under the guidance of Roman V. Krems. The author developed the code, performed the calculations, the data analysis, and the analytical derivations.

The material in Chapter 6 was published in the article: R. A. Vargas-Hernández, and R. V. Krems, *Engineering extended Hubbard models with Zeeman excitations of ultracold Dy atoms*, J. Phys. B **49**, 235501 (2016). The project was identified and designed by Roman V. Krems and the author. The author developed the code, performed the calculations, and the data analysis.

The material in Appendix A is unpublished work by the author. The project was identified and designed by the author. The code that computes the HOMO energy with density functional theory was provided by R. Ghassemizadeh from the University of Freiburg. The manuscript and further calculations are still in progress.

Results presented in Chapters 2-4 and Appendix A were done using *scikit-learn's* Gaussian Process Regression library [137].

Table of Contents

Abstract	iii
Lay Summary	iv
Preface	v
Table of Contents	vii
List of Tables	xi
List of Figures	xii
Glossary	xxiv
Acknowledgments	xxvi
.xxviii
1 Introduction	1
1.1 Machine learning to study quantum systems	2
1.2 Quantum walks	5
1.3 Quantum simulators of condensed matter	8
1.4 Thesis outline	9
2 Gaussian Processes	11
2.1 Introduction	11
2.2 Gaussian Processes: prediction	12

2.3	Gaussian Processes: training	13
2.4	Kernel functions	16
2.4.1	Constant kernel	17
2.4.2	Square exponential kernel	17
2.4.3	Matern kernel	17
2.4.4	Rational Quadratic kernel	19
2.4.5	Periodic kernel	21
2.4.6	Linear kernel	22
2.5	Potential Energy Surfaces	23
2.5.1	Results	25
2.5.2	Gaussian Processes regression vs Neural Networks	25
2.6	Summary	28
3	Bayesian Optimization	31
3.1	Introduction	31
3.2	Acquisition function	34
3.2.1	Probability of Improvement	34
3.2.2	Expected Improvement	36
3.2.3	Upper Confidence Bound	37
3.3	Results	39
3.3.1	Fitting known quantum dynamics results	41
3.3.2	Fitting without known quantum dynamics results	45
3.3.3	The inverse scattering problem	47
3.4	Summary	50
4	Extrapolation of quantum observables	53
4.1	Introduction	55
4.2	Combination of kernels	56
4.2.1	Kernel combination to predict a single phase transition	58
4.2.2	Kernel combination to predict multiple phase transitions	62
4.3	Model selection	66
4.3.1	Bayesian information criterion	67
4.3.2	Greedy search for model construction	68

4.4	Results	68
4.4.1	Extrapolating quantum observables to unconverged regimes 69	
4.4.2	SSH-BM polaron phase diagram	70
4.5	Discussion	73
5	Accelerating quantum walks	77
5.1	Introduction	77
5.2	Models	79
5.2.1	Ideal 1D lattices	81
5.2.2	Disordered 1D lattices	85
5.2.3	Binary trees	88
5.2.4	Glued binary trees	91
5.3	Conclusion	94
6	Quantum simulators with highly magnetic atoms	97
6.1	Introduction	97
6.2	Lattice Hamiltonian with Zeeman excitations	99
6.3	Engineering lattice models	104
6.3.1	$t - V$ model	104
6.3.2	$t - V$ model with Dy atoms	106
6.3.3	Particle number-non-conserving interactions	110
6.3.4	Anderson localization of Zeeman excitations	112
6.4	Conclusion	113
7	Conclusion	116
7.1	Summary of the thesis	116
7.2	Future work	119
	Bibliography	122
A	Bayesian optimization for hybrid-density functionals	145
B	Schrieffer–Wolff transformation	149

C	Magnetic dipole–dipole interaction	152
	C.0.1 Matrix element	155

List of Tables

Table 2.1	RMSE (test errors computed on 120,000 points) of the PES obtained with the NNs for different training points (N_{pts}). The number in the parenthesis are the number of neurons used in that particular NN. $\langle NN \rangle_{10}$ is the average RMSE for 10 NNs with different sets of N_{pts} . The values are in cm^{-1}	27
Table 2.2	RMSE (test errors computed on 120,000 points) of the PES obtained with GP regression. All GPs use the SE kernel function. N_{pts} is the number of training points. $\langle GP \rangle_{10}$ is the average RMSE for 10 GP models with different sets of N_{pts} . The values are in cm^{-1}	27

List of Figures

Figure 2.1	Interpolation using a GP model trained with 7 points and an exponential squared kernel. The black dashed line is $f(x) = \frac{x}{10} + x^2 + 10\sin(\frac{3}{2}x)$. The solid blue line is the prediction of the GP model, Equation 2.5. The grey shaded area is the standard deviation of the predicted mean of the GP model, Equation 2.4. The blue square symbols are the training data.	14
Figure 2.2	a) The square exponential function with different length scales. b) Prediction of a GP model with the SE kernel using different length scale values.	18
Figure 2.3	a,c) The Matern kernel function for different length scales and $\nu = \frac{3}{2}$ (a) or $\nu = \frac{5}{2}$ (c). b,d) Prediction of a GP model with the MAT kernel using different length scale values and $\nu = \frac{3}{2}$ (b) or $\nu = \frac{5}{2}$ (d).	19
Figure 2.4	a) The rational quadratic function with different ℓ and α values. b) Prediction of a GP model with the RQ kernel using different values for ℓ and α	20
Figure 2.5	a) The exponential sine squared function with different length scales and different periodicities. b) Prediction of a GP model with the PER kernel using different values for ℓ and $p = 8.11$	21
Figure 2.6	Prediction of a GP with the LIN kernel using different σ and polynomial degree.	22

Figure 2.7	H_2 molecule’s energy for different interatomic distances predicted with a GP model trained with 6 points and the SE kernel. The exact PES (dashed black line) is from reference [23]. . . .	24
Figure 2.8	Distribution of energy values of the PES dataset. Figure from reference [102]	26
Figure 2.9	Absolute errors in transition frequencies computed on PESs fitted to 625 points with different methods (NN: top half, GP: bottom half) vs the spectrum computed on the reference PES. Figure from reference [102]	29
Figure 3.1	Example of the BO algorithm over three iterations. Blue markers are GP’s the training data, blue curves are GP’s prediction and black dashed line is the black-box function $f(x) = \frac{x}{10} + x^2 + 10\sin(\frac{3}{2}x)$. The yellow solid curves are the acquisition function, and \blacktriangledown are the maximum of the acquisition function at each iteration, \mathbf{x}_{n+1}	35
Figure 3.2	Illustration of the PI acquisition during BO. We also show the PI function for different values of ϵ	36
Figure 3.3	Illustration of the EI acquisition during BO.	37
Figure 3.4	Illustration of the UCB acquisition during BO. We also show how the UCB function for different values of κ	38
Figure 3.5	a) The reaction probability for the $H_2 + H \rightarrow H + H_2$ reaction as a function of the collision energy. The black solid curve—accurate calculations from [23]. The dashed curves—calculations based on GPs PESs obtained with $22 + 1$ <i>ab initio</i> points, where the 22 initial points (solid blue curve) are fixed and the extra point is randomly sampled in the configuration space. b) The error of the reaction probabilities as a function of the location extra point added to the original 22 points. . . .	42

Figure 3.6	a) The reaction probability for the $\text{H}_2 + \text{H} \rightarrow \text{H} + \text{H}_2$ reaction as a function of the collision energy. The black solid curve – accurate calculations from Ref. [23] based on the surface constructed with 8701 <i>ab initio</i> points. The dashed curves – calculations based on the GP PES obtained with 22 <i>ab initio</i> points (blue); 23 points (orange), 30 points (green) and 37 points (inset). The RMSE of the results with 37 points is 0.009. b) The GP model of the PES for the H_3 reaction system constructed with 30 <i>ab initio</i> points. This surface yields the quantum dynamics results shown by the green curve in the upper panel.	44
Figure 3.7	Convergence of the RMSE of the probabilities for $\text{H}_2 + \text{H} \rightarrow \text{H} + \text{H}_2$ reaction with the number of BO iterations as a function of κ in UCB.	45
Figure 3.8	The reaction probability for the $\text{OH} + \text{H}_2 \rightarrow \text{H} + \text{H}_2\text{O}$ reaction as a function of the collision energy. The black solid curve – accurate calculations from Ref. [42] based on the surface constructed with ~ 17000 <i>ab initio</i> points. The dashed curves – calculations based on the GP PES obtained with 200 <i>ab initio</i> points (blue); 280 points (orange) and 290 points (green). The RMSE of the 290-point result is 0.0076.	46
Figure 3.9	The reaction probabilities for the $\text{H}_2 + \text{H} \rightarrow \text{H} + \text{H}_2$ reaction as functions of the collision energy. The black solid curve – accurate calculations from Ref. [23]. The dashed curves – the results of iterative calculations maximizing the difference between the reaction probabilities in subsequent iterations. The black curve is not used for these calculations. The inset shows the agreement between the reaction probabilities (red symbols) based on the GP approach after 48 iterations (total of 70 <i>ab initio</i> points) and the exact results.	47

Figure 3.10	a) The reaction probabilities for the modified $\text{H}_2 + \text{H} \rightarrow \text{H} + \text{H}_2$ reaction as functions of the collision energy. The black dot-dashed curve is obtained by a modification of the previous results (black solid curve) involving a translation along the energy axis. The ML models are trained to obtain the PES that would describe the new reaction probabilities. The green dashed curve is a results of such training after 30 iterations, which produces a surface constructed with 52 <i>ab initio</i> points. b) Comparison of the original PES (blue) with the new PES (red) found by the BO algorithm. The new PES yields the reaction probabilities described by the green dashed curve in the upper panel. The RMSE of the results shown by the green dashed curve is 0.016 ev.	49
Figure 4.1	Schematic diagram of a quantum system with three phases. . .	55
Figure 4.2	The upper panel is the RMSE of GPs' prediction as a function of λ_{SSH} . The markers illustrate the values of λ_{SSH} used to train the GPs. We consider two sets of training values, (a) $0.4 \leq \lambda_{SSH} \leq 1.3$ and (b) $1.6 \leq \lambda_{SSH} \leq 2.5$. The lower panels illustrate the GPs prediction (solid colour curves) and the accurate energy dispersion (dashed colour curves). We use $k_{MAT} \times k_{LIN}$ as the kernel function for both GP models. $\omega/t = 3$ is held fixed.	60
Figure 4.3	The upper panel is the RMSE of GPs' prediction as a function of λ_{SSH} . The markers illustrate the values of λ_{SSH} used to train the GPs. We consider two sets of training data (a) 100 points distributed from $0.2 \leq \lambda_{SSH} \leq 0.5$ and (b) 125 points distributed in $0.2 \leq \lambda_{SSH} \leq 0.6$. The lower panels illustrates the GPs prediction (solid colour curves) and the accurate energy dispersion (dashed colour curves). We use $k_{MAT} \times k_{LIN}$ as the kernel function for both GP models. For all the energy dispersions display here, we consider $\omega/t = 3$ and $\beta = 0$. . .	61

Figure 4.4	For both panels we illustrate the change in K_{GS} as a function of λ_{SSH} . The blue curves is the K_{GS} from the GP's prediction of the energy dispersion. The black dashed curve is the correct K_{GS} for each λ_{SSH} . The markers are the values of λ_{SSH} that we use to train each GPs. The upper panel shows $0.2 \leq \lambda_{SSH} \leq 0.5$, while the lower panel shows $0.2 \leq \lambda_{SSH} \leq 0.6$. We use $k_{MAT} \times k_{LIN}$ as the kernel function for both GPs. For all the calculations $\omega/t = 3$ and $\beta = 0$	62
Figure 4.5	K_{GS} for the mixed model 4.3. The black dashed curves are the calculations from Ref. [87]. The color map is the prediction of the GP models with the fully optimized kernels. The models are trained by the dispersion of the polarons at the locations represented by the black dots. The different kernels were considered, k_{MAT} , k_{RQ} and k_{RBF}	63
Figure 4.6	K_{GS} for the mixed model 4.3. The black dashed curves are the calculations from Ref. [87]. The color map is the prediction of the GP models with the fully optimized kernels. The models are trained by the dispersion of the polarons at the locations represented by the black dots. The different kernels considered here are all the possible pairwise addition, Equation 4.6, of two simple kernels, k_{MAT} , k_{RQ} and k_{RBF}	65
Figure 4.7	The momentum of the polaron ground state for the mixed model 4.3. The black dashed curves are the calculations from Ref. [87]. The color map is the prediction of the GP models with the fully optimized kernels. The models are trained by the dispersion of the polarons at the locations represented by the black dots. The different kernels considered here are all the possible pairwise multiplication, Equation 4.7, of two simple kernels, k_{MAT} , k_{RQ} and k_{RBF}	66
Figure 4.8	Schematic diagram of the kernel construction method employed to develop a GP model with extrapolation power. At each iteration, the kernel with the highest Bayesian information criterion (Equation 4.10) is selected.	69

- Figure 4.9 Left panel: Holstein polaron dispersion predicted with a GP model with k_{RQ} , black dashed curves, and $k_{RQ} \times k_{RQ} \times k_{RBF} + k_{LIN}$, blue dashes curves. The red solid curves are polaron dispersions computed with the momentum average approximation method [22]. Right panel: The change in the RMSE as a function of depth in the kernel combination search guided by the BIC. We plot the RMSE between GP predicted polaron dispersion and the exact polaron dispersion at different values of ω . The grey area is the range in ω considered for training the GP. For both figures, the value of $g = 1.5$ is fixed. 70
- Figure 4.10 K_{GS} for the mixed model (4.3) as a function of β/α for $\lambda_{SSH} = 2\alpha^2/t\hbar\omega$. The dotted curves are the quantum calculations from Ref. [87]. The color map is the prediction of the GP models. Each panel illustrates the improvement of the predicted phase diagram. The panels correspond to the optimized kernels GPL-0 (left), GPL-1 (right), GPL-2 (centre), where “GPL- X ” denotes the optimal kernel obtained after X depth levels in the algorithm depicted in Figure 4.8. The models are trained by the dispersion of the polarons at the locations represented by the white dots. 72
- Figure 4.11 K_{GS} for the mixed model (4.3) as a function of β/α for $\lambda_{SSH} = 2\alpha^2/t\hbar\omega$. The dotted curves are the quantum calculations from Ref. [87]. The color map is the prediction of the GP models. Each panel illustrates the improvement of the predicted phase diagram. The panels correspond to the optimized kernels GPL-0 (left), GPL-1 (right), GPL-2 (centre), where “GPL- X ” denotes the optimal kernel obtained after X depth levels in the algorithm depicted in Figure 4.8. The models are trained by the dispersion of the polarons at the locations represented by the white dots. 73

Figure 4.12 K_{GS} for the mixed model (4.3) as a function of β/α and $\lambda_{SSH} = 2\alpha^2/t\hbar\omega$. The black dashed curves are the calculations from Ref. [87]. The color map is the prediction of the GP models with the fully optimized kernels. The models are trained by the dispersion of the polarons at two different locations represented by the white dots. Left column panels are the predicted phase diagram with GPL-0 for both training data sets and right column panels are with GPL-2. 74

Figure 5.1 Time dependence of the standard deviation (in units of lattice constant) of the wave packet for a particle initially placed in a single site of a one-dimensional ideal lattice. The solid black curves represent the ballistic expansion governed by the Hamiltonian (5.1) with $v = 0$ and $\hat{V}_{nc} = 0$. Upper panel: The oscillating curves show the size of the wave packets governed by the Hamiltonian (5.1) with $v = 0$, $\gamma = 0$, $\Delta/t = 1$, $t = 1$ and two values of $\Delta\epsilon$: $\Delta\epsilon = 10/t$ (blue) and $\Delta\epsilon = 20/t$ (red). Lower panel: The oscillating curves show the size of the wave packets governed by the Hamiltonian (5.1) with $v = \pm 1$, $\Delta/t = 1$, $t = 1$ and $\Delta\epsilon = 20/t$. The insets show the average number of particles $\langle n \rangle$ as a function of time for the corresponding Hamiltonian parameters. Notice that for $\Delta\epsilon = 20/t$, $\langle n \rangle$ stays below 1.2 at all times. Figure from reference [185]. 82

Figure 5.2 Time dependence of the standard deviation (in units of lattice constant) of the wave packet for a particle initially placed in a single site of a one-dimensional ideal lattice. The solid black curves represent the ballistic expansion governed by the Hamiltonian (5.1) with $v = 0$ and $\hat{V}_{nc} = 0$. The oscillating curves show the size of the wave packets governed by the Hamiltonian (5.1) with $v = 0$, $\gamma = 0$, $\Delta/t = 1$, $t = 1$, $\Delta\epsilon = 20/t$ and a Hilbert space with different number of particles, 1 particle (black), 1 and 3 particles (red) and 1,3 and 5 (blue). The upper inset shows the logarithm of the particle probability distributions in a disordered 1D lattice with 19 sites. The results are averaged over 50 realizations of disorder and are time-independent with $w = 10/t$. And the lower inset depicts the average number of particles $\langle n \rangle$ as a function of time for the three different wave packets. Figure from reference [185]. . . 84

Figure 5.3 Time dependence of the standard deviation (in units of lattice constant) of the wave packet for a particle initially placed in a single site of a one-dimensional ideal lattice. The solid black line represents the ballistic expansion governed by the Hamiltonian (5.1) with $v = 0$, $\Delta = 0$ and $\gamma = 0$. For the dotted red curve $\Delta = 0$ and $\gamma = t$ while for the blue solid curve $\Delta = t$ and $\gamma = 0$. For all of these calculations, $\Delta\epsilon = 20/t$. The inset shows the average number of particles $\langle n \rangle$ as a function of time for $\Delta = 0$ and $\gamma = t$ (dotted red curve) and $\Delta = t$ and $\gamma = 0$ (blue solid curve). Figure from reference [185]. 85

- Figure 5.4 Upper panel: The logarithm of the particle probability distributions in a disordered 1D lattice with 41 sites: diamonds – $\Delta/t = 1$, squares – $\Delta/t = 1/2$, circles – $\Delta/t = 1/10$. The results are averaged over 100 realizations of disorder and are time-independent. The dashed line is an exponential fit to the $\Delta/t = 1/10$ results. Lower panel: the long-time limit of the IPR defined in Eq. (5.3) averaged over 100 instances of disorder as a function of the disorder strength w : solid line – $\Delta = 0$, dashed line – $\Delta/t = 1$. The inset shows the IPR averaged over 100 realizations of disorder for two disorder strengths $w = \{5/t, 10/t\}$ as functions of time: the solid black curves – $\Delta = 0$; the dotted and dot-dashed curves – $\Delta = t$. Figure from reference [185]. 87
- Figure 5.5 Schematic diagram of an ideal binary tree with depth-5 (\mathcal{G}_5). 88
- Figure 5.6 Upper panel: The growth of the wave packet for a single particle placed at the root of the tree (\mathcal{G}_5): black line – $\Delta = 0$; the oscillating curves – $\Delta/t = 1$ with $\Delta\epsilon = 10/t$ (blue) and $\Delta\epsilon = 20/t$ (red). The inset shows the probability of reaching the last node (in layer 5) of the binary tree as a function of time. The curves are color-coded in the same ways as in the main plot. Lower panel: The L_1 norm between the probability distribution of the wave packet and the uniform distribution as a function of time. The uniform distribution is defined by the value $1/\dim(\mathcal{G}_5)$ for each node. The solid black curve is for $\Delta = 0$. The oscillating curves are for $\Delta = t$ with $\Delta\epsilon = 10/t$ (blue) or $\Delta\epsilon = 20/t$ (red). For all curves in both figures $\gamma = 0$. Figure from reference [185]. 89
- Figure 5.7 The squares represent the stationary distribution $\pi_S(v)$ for a quantum walk (QW) on the \mathcal{G}_5 tree with $\Delta = 0$. The circles and diamonds are the stationary distributions $\pi_S(v)$ for quantum walks with $\Delta/t = 1$ and $\Delta\epsilon = 10/t$ and $\Delta\epsilon = 20/t$, respectively. Figure from reference [185]. 92

Figure 5.8	Schematic diagram of an ideal glued binary trees with depth-4 ($\mathcal{GBT} - 4$).	92
Figure 5.9	Average particle probability distribution in a disordered $\mathcal{GBT}4$ graph: upper panel – $\Delta/t = 0$, middle panel – $\Delta\epsilon = 10/t$ and lower panel – $\Delta\epsilon = 20/t$. For all panels we consider 100 realizations of disorder with a strength of $w = 5/t$ and $\gamma = 0$. The wave packet for a particle is initially placed in the head node of a $\mathcal{GBT}4$ graph. Figure from reference [185].	94
Figure 5.10	Average particle probability distribution in a disordered $\mathcal{GBT} - 4$ graph with strength $w = 5/t$. The wave packet for a single particle is initially placed in the head node of a $\mathcal{GBT} - 4$ graph shown in Figure 5.8. The upper panel displays the probability (5.14) summed over all nodes of layer $j = 3$; the lower panel shows $p_j(t)$ for the bottom node $j = 6$. Insets: The particle probability distributions for an ideal $\mathcal{GBT} - 4$ graph with $\Delta\epsilon = 10/t$ (blue dot-dashed) and $\Delta\epsilon = 20/t$ (red dashed). The broken curves show the results obtained with $\Delta/t = 1$ and the full black lines – $\Delta/t = 0$. Figure from reference [185].	96
Figure 6.1	Upper panel: Zeeman levels of a $\text{Dy}(^5I)$ atom in the lowest-energy spin-orbit state characterized by $J = 8$ in a magnetic field $\mathbf{B} = B_0\hat{z}$. Lower panel: the solid curve – difference of the energy gaps $(\epsilon_{M=2} - \epsilon_{M=1}) - (\epsilon_{M=1} - \epsilon_{M=0})$; the dot-dashed curve – difference of the energy gaps $(\epsilon_{M=2} - \epsilon_{M=1}) - (\epsilon_{M=0} - \epsilon_{M=-1})$. The horizontal dashed line shows the magnitude of the matrix element $t_{i,i+1}$ in Eq. (6.21) for Dy atoms with $ g\rangle = J = 8, M = 0\rangle$ and $ e\rangle = J = 8, M = 1\rangle$ in an optical lattice with $a = 266$ nm. Figure from reference [184].	103

Figure 6.2	The magnitudes of the coupling constants t_{ij} (upper panel) and the ratio t_{ij}/v_{ij} (lower panel) with $j = i \pm 1$ for the Zeeman states of Dy corresponding to $ g\rangle \Rightarrow JM\rangle$ and $ e\rangle \Rightarrow JM'\rangle$. The calculations are for the magnetic field $\mathbf{B} = B_0(0.1\hat{x} + \hat{z})$ with $B_0 = 100$ G. The Zeeman states in this magnetic field retain 96% of the eigenstates of \mathbf{J}^2 and \mathbf{J}_z . Figure from reference [184].	107
Figure 6.3	The magnitude of the coupling constant v_{ij} with $j = i \pm 1$ for the Zeeman states of Dy corresponding to $ g\rangle \Rightarrow JM\rangle$ and $ e\rangle \Rightarrow JM'\rangle$. The calculations are for the magnetic field $\mathbf{B} = B_0(0.1\hat{x} + \hat{z})$ with $B_0 = 100$ G. The Zeeman states in this magnetic field retain 96% of the eigenstates of \mathbf{J}^2 and \mathbf{J}_z . Figure from reference [184].	108
Figure 6.4	The ratio $v_{i,i+1}/t_{i,i+1}$ for the Zeeman states of Dy corresponding to $ g\rangle = J = 8, M = -7\rangle$ and $ e\rangle = a J = 8, M = -8\rangle + b J = 8, M'\rangle$: full circles – $a = \sqrt{3/4}, b = \sqrt{1/4}$; open circles – $a = \sqrt{3/5}, b = \sqrt{2/5}$. The calculations are for the magnetic field $\mathbf{B} = B_0(0.1\hat{x} + \hat{z})$ with $B_0 = 100$ G. The Zeeman states in this magnetic field retain 96% of the eigenstates of \mathbf{J}^2 and \mathbf{J}_z . Figure from reference [184].	109
Figure 6.5	The magnetic field dependence of the quantities $t_{i,i+1}$ (squares) and $v_{i,i+1}$ (circles) defined in Eqs. (6.24) and (6.25) for two different pairs of the Zeeman state of Dy($J = 8$) atoms: the full symbols – the results for $ g\rangle = J = 8, M = -8\rangle$ and $ e\rangle = J = 8, M = -7\rangle$; the open symbols – the results for $ g\rangle = J = 8, M = 0\rangle$ and $ e\rangle = J = 8, M = +1\rangle$. The magnetic field is given by $\mathbf{B} = B_0(0.1\hat{x} + \hat{z})$. The Zeeman states in such a magnetic field retain 96% of the eigenstates of \mathbf{J}^2 and \mathbf{J}_z . Figure from reference [184].	110

- Figure 6.6 Anderson localization of the $|J = 8, M = 0\rangle \rightarrow |J = 8, M = +1\rangle$ excitation in a one-dimensional array of Dy atoms on an optical lattice with $a = 266$ nm and 20 % of the lattice sites empty. The upper panel shows the probability distribution for the atoms in the corresponding site to be in the excited state at $t = 2$ seconds formed by a single excitation placed at $t = 0$ in the middle of a lattice with 1000 sites. The lower panel shows the width of the excitation probability distribution as a function of time. Figure from reference [184]. 115
- Figure A.1 The iterations of BO to find the minimizer of \mathcal{L} for the H₂ molecule. The black vertical line is the next proposed point by BO. The blue markers are the training data used to construct the acquisition function. The blue-solid curve is the mean of the GP model and the grey shaded area is $\sigma(\cdot)$ 148

Glossary

This glossary uses the handy `acronym` package to automatically maintain the glossary. It uses the package's `printonlyused` option to include only those acronyms explicitly referenced in the `LATEX` source.

$\alpha(\cdot)$	acquisition function
BO	Bayesian optimization
BIC	Bayesian information criterion
DFT	Density functional theory
EI	Expected Improvement
GP	Gaussian Process
HOMO	highest occupied molecular orbital
$k(\cdot, \cdot)$	kernel function
θ	kernel parameters
KRR	kernel Ridge regression
LHS	Latin hypercube sampling
LR	long-range
\mathcal{L}	loss function
ML	Machine learning

NN	Neural network
PES	potential energy surface
PI	Probability of improvement
QW	quantum walk
RL	Reinforcement learning
SR	short-range
\mathcal{D}	training data
UCB	Upper confidence bound

Acknowledgments

Foremost, I would like to express my sincere gratitude to my supervisor Professor Roman V. Krems for giving me the opportunity to do a Ph.D. in his research group. I would also like to thank him for all his patients, devotion and endless help to make this stage of my life a grateful experience. I would like to thank Marina Litinskaya and Evgeny A Shapiro for all the help they grant me during the first years of my grad studies. I am thankful for the great research group I got to work with, P. Xiang, J. Cui, J. Sous, T. Chattaraj, and J. Cantin; thank you very much for all those great physics and non-physics discussions during these years. I would like to thank R. Chen for all the great machine learning discussions we had.

I am grateful for the incredible experiences I had during my multiple visits to the research group of Professor T. Schaetz in Freiburg Germany, these trips would not had happened without the help of the IRTG-coco program. I would also like to thank my German friends, F. Hakelberg, S. Fuchs and P. Weckesser for all the great memories together. I would like to thank Professor D. H. Zhang and his student Y. Guan who made my visit to the Dalian Institute of Chemical Physics a pleasant experience.

Vancouver became my second home and I was lucky to make new friends. E. Castillo and J. Casanova for all their help during my first years. B. Loosley for the beer/biking moments. K. Lovering and A. Bain for all those cups of coffee. P. L. Esquinas and A. Lee for all those quantum assignments, S. Cuen-Rochin for the basketball moments, M. Vashista, T. Matzat, H. Neurert and R. Garcia-Rojas-Villegas. Thank you all the great memories that we shared during my Ph.D. A special thanks goes to Montse Rueda for helping me be a better human, bike-beers-books.

Like any other immigrant, leaving home behind is one of the hardest challenges because of the people we stop seeing. I am thankful for all the great Mexican friends who supported me and my family during my Ph.D. Romina Arguelles and her family, Mariana Dominguez, Paula Balbi, Izaskun Diaz, Alejandro Evaris, Pablo and Diego Mugierza, Arturo Sanchez, Raymundo Marcial, Emiliano Martinez, Pablo Bulit, Rodrigo Cortés, Salvador Portocarrero and Adrian Vázquez. I would like to thank my best friend Perla Ordaz for all the non-little things she has done for me. I am thankful to Professor Jose Carlos Flores who taught me math in high-school and Professor Carlos Amador-Bedolla for introducing me to the beautiful world of quantum mechanics.

I would like to thank my aunt Julieta Mariaca, my aunts Martha E. Medina Guerrero, Guadalupe Guerrero and their family, and my uncle Walter Louis and his family; thank you all for your help and love to me and my family.

I am grateful I met Mriga Das at the beginning of my Ph.D. Mi amor, thank you very much for all the joy and love you bring to my life. There is no doubt this ride was easier thanks to you.

I was lucky to do my Ph.D. while living with my sister Ana Jimena Vargas-Hernández; thank you for all your help, love, patience and pancakes during this time; every day it felt like home.

Last but not least, I would like to thank my parents, Ana Lilia Hernández-Cuevas and Alejandro Vargas-Cruz for all their love that goes even beyond life and all their support. I would also like to thank them for teaching me never to give up and always try to make this world fairer. Since I was a kid my mother taught me that everything in life is explained with the laws of math and physics. My dad also taught me that chemistry, on its own, is a wonderful science that can make you understand the wonders of nature. I would like to thank my parents for making me passionate about knowledge.

To my mother and her eternal fight for a fairer world.

To my father, for teaching me to always fight for my dreams.

In memory of my father.

¡Hasta la victoria!
— Ernesto Guevara

Chapter 1

Introduction

If I have seen further it is by standing on the shoulders of Giants.
— Sir Isaac Newton (1675)

The field of quantum physics has been theoretically challenging due to dimensionality of the Hilbert space required to accurately describe quantum systems. Major breakthroughs in the field of modern science have been achieved due to growth in computational power in recent years. Nevertheless, there are still many interesting quantum systems that cannot be fully simulated with the current computing capacity and thus different methodologies have to be designed. In 1982 R. Feynman proposed the idea to simulate quantum systems analogously [64]. This new research field is known as *quantum simulators* and its only goal is to use fully controllable quantum systems to improve the human knowledge of quantum physics [30].

As practical as it seems, *quantum simulators* cannot mimic all quantum systems and the hunger for novel numerical tools to study quantum physics is still present [30]. Recently the use of machine learning algorithms has gained momentum in the fields of theoretical chemistry and physics [8, 33, 35, 71]. In this thesis machine learning algorithms are being used to study quantum physics in a wide range of problems; from the description of the electronic energy as a function of the nuclei position to the discovery of new quantum phases of matter. In this chapter, a brief introduction and motivation to each of the research projects that were conducted during my Ph.D. are presented. The last part of the current chapter sum-

marizes each of the chapters dedicated to the research projects that this thesis is comprised of.

1.1 Machine learning to study quantum systems

Physics is devoted to studying observable quantities like position, momentum or energy, to explain how nature behaves. The state of a system, S , determines the measured quantity, o , that is produced by a physical observable \mathcal{F} . This process can be described by,

$$\mathcal{F}(S) = o. \quad (1.1)$$

For example, given the distance and the masses of two bodies, one can compute the gravitational potential energy,

$$U(x_1, x_2, m_1, m_2) = -\frac{G m_1 m_2}{r_{1,2}}, \quad (1.2)$$

where $r_{1,2}$ is the distance between the two particles, and m_i is the mass of each particle. G is the gravitational constant. One can observe that Equation 1.2 is similar to Equation 1.1, where $S = [x_1, x_2, m_1, m_2]$ and $\mathcal{F}(\cdot) \rightarrow U(\cdot)$.

Things are more complicated in the quantum world since quantum observables are described by operators, $\hat{\mathcal{O}}$ [161]. In quantum mechanics, an operator acts on a quantum state $|\psi\rangle$ and produces a scalar, λ , and another quantum state,

$$\hat{\mathcal{O}}|\psi\rangle \rightarrow \lambda|\phi\rangle. \quad (1.3)$$

A specific group of operators is the Hermitian operators whose eigenvalues are real, $\lambda_i \in \mathbb{R}$. All the quantum observables that are studied in this thesis are described by Hermitian operators. When a Hermitian operator acts on its eigenvector, the output vector is the same,

$$\hat{\mathcal{O}}|\psi_i\rangle = \lambda_i|\psi_i\rangle. \quad (1.4)$$

where $|\psi_i\rangle$ is an eigenvector of operator $\hat{\mathcal{O}}$ and λ_i is the eigenvalue associated

with $|\psi_i\rangle$. Furthermore, the dot product between two eigenvectors is equal to a Kronecker delta,

$$\langle \psi_i | \psi_j \rangle = \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j \end{cases}, \quad (1.5)$$

i.e. they are orthogonal. Using Equation 1.4 one can rewrite any Hermitian operator using its spectral decomposition,

$$\hat{\mathcal{O}} = \sum_i \lambda_i |\psi_i\rangle \langle \psi_i|, \quad (1.6)$$

where λ_i are the eigenvalues and $|\psi_i\rangle$ the eigenvectors of the operator $\hat{\mathcal{O}}$. Each $|\psi_i\rangle \langle \psi_i|$ term in Equation 1.6 is a projector or measurement operators in the eigenspace of $\hat{\mathcal{O}}$, $\hat{P}_i = |\psi_i\rangle \langle \psi_i|$. Furthermore, any quantum state $|\psi\rangle$ can be rewritten in terms of the eigenstates of $\hat{\mathcal{O}}$,

$$|\psi\rangle = \sum_i |\psi_i\rangle \langle \psi_i | \psi \rangle, \quad (1.7)$$

where each $\langle \psi_i | \psi \rangle$ term is an inner product between the quantum states $|\psi_i\rangle$ and $|\psi\rangle$. Using Equations 1.6 and 1.7, one can compute the expectation value of a quantum observable as,

$$\begin{aligned} \langle \psi | \hat{\mathcal{O}} | \psi \rangle &= \sum_i \lambda_i \langle \psi | \psi_i \rangle \langle \psi_i | \psi \rangle \\ &= \sum_i \lambda_i \left[\sum_j \langle \psi | \psi_j \rangle \langle \psi_j | \right] |\psi_i\rangle \langle \psi_i| \left[\sum_l \langle \psi_l | \langle \psi | \psi_l \rangle \right] \\ &= \sum_i \lambda_i |\langle \psi | \psi_i \rangle|^2 \end{aligned} \quad (1.8)$$

where $|\langle \psi | \psi_i \rangle|^2 = \langle \psi_i | \psi \rangle \langle \psi | \psi_i \rangle$ is the probability of measuring λ_i . Furthermore, Equation 1.8 is the statistical definition of the expectation value for a discrete random variable X ,

$$\mathbb{E}[X] = \sum_i x_i p_i. \quad (1.9)$$

The ground state of $\hat{\mathcal{O}}$, denoted as $|\psi_{GS}\rangle$, is the eigenvector with the lowest eigenvalue, λ_0 . Using Equation 1.8 and $|\psi\rangle \rightarrow |\psi_{GS}\rangle$ one can measure λ_0 ,

$$\langle \psi_{GS} | \hat{\mathcal{O}} | \psi_{GS} \rangle = \lambda_0. \quad (1.10)$$

Additionally, operators can have internal parameters; for example, the Hamiltonian of a particle in a one-dimensional box depends on the length of the box, L , and the mass of the particle, m . Here we collectively described the parameters of an operator by $\bar{\theta}$. In the case of a particle in a one-dimensional box, $\bar{\theta} = [L, m]$.

One of the central ideas in this dissertation, Equation 1.11, is the use of Machine learning (ML) algorithms to learn the relation between λ_0 and $\bar{\theta}$,

$$\langle \psi_{GS} | \hat{\mathcal{O}}(\bar{\theta}) | \psi_{GS} \rangle = \lambda_0 \rightarrow \mathcal{F}(\bar{\theta}), \quad (1.11)$$

where $\mathcal{F}(\cdot)$ is the ML algorithm that mimics the computation of λ_0 . If $\mathcal{F}(\cdot)$ is capable to learn $\langle \psi_{GS} | \hat{\mathcal{O}}(\bar{\theta}) | \psi_{GS} \rangle$ accurately, one could reduce the computational resources needed to calculate λ_0 for different values of $\bar{\theta}$.

In theoretical chemistry, computing the electronic ground state energy of an ensemble of electrons and nuclei in different fixed positions can be mapped to Equation 1.11. Each of the possible spatial configurations of the nuclei is described by a unique $\bar{\theta} = [x_1, \dots, x_m]$, where x_i is the position of each nucleus. The map between the nuclei positions and the electronic ground state energy, $\mathcal{F}(\cdot)$, is known in the field of quantum chemistry as the potential energy surface (PES) [133]. As it is stated above, inferring $\mathcal{F}(\cdot)$ is crucial to reduce the computational power needed to study quantum systems. This leads to the first question addressed in this thesis: *Can ML be used to interpolate quantum observables accurately to construct PESs?*

For some quantum systems, the computation of λ_0 with standard numerical methods lacked convergence due to the size of the Hilbert space required to describe the quantum system. Furthermore, without the ability to compute λ_0 in the entire space of $\bar{\theta}$, one can not construct the complete phase diagram of a quantum system [160]. This raises the second question: *Can ML be used to extrapolate observables to learn phase diagrams?*

Quantum observables can also depend on other quantum observables, Equation 1.12, making the computations more demanding.

$$\lambda_0 = \langle \phi_{GS} | \hat{\mathcal{O}}_1(\bar{\theta}, \langle \psi_{GS} | \hat{\mathcal{O}}_2(\bar{\theta}') | \psi_{GS} \rangle) | \phi_{GS} \rangle. \quad (1.12)$$

Using Equation 1.11, one can rewrite Equation 1.12 as,

$$\langle \phi_{GS} | \hat{\mathcal{O}}(\bar{\theta}, \mathcal{F}(\bar{\theta}')) | \phi_{GS} \rangle = \langle \phi_{GS} | \hat{\mathcal{O}}(\bar{\theta}, \lambda'_0) | \phi_{GS} \rangle, \quad (1.13)$$

where θ' are the parameters of $\hat{\mathcal{O}}_2$ and λ'_0 is the lowest eigenvalue of $\hat{\mathcal{O}}_2(\theta')$. $\mathcal{F}(\cdot)$ computes the quantum observable of $\hat{\mathcal{O}}_2$ given the values of θ' . The parameters of $\hat{\mathcal{O}}$ are described by $\bar{\theta}$.

Many types of quantum problems can be described by Equation 1.12; for example, to compute any quantum dynamical observable one first needs to fit $\mathcal{F}(\cdot)$ to reduce the complexity of the problem. From Equation 1.13 one can observe the causality relation between λ_0 and $\mathcal{F}(\cdot)$ is $\mathcal{F}(\cdot) \rightarrow \lambda_0$, so one can raise the following question: *Can $\mathcal{F}(\cdot)$ be learned from λ_0 ?* This question amounts in the field of quantum physics to the *inverse scattering problem*. Due to the complexity of the inverse scattering problem, one could rephrase the above question as the third question of this dissertation: *Can the inverse scattering problem be solved using ML?*

1.2 Quantum walks

The goal of ML is designing robust algorithms capable of learning almost any phenomenon given some data. In order to do so, ML algorithms are trained by minimizing a loss function which sometimes is not a simple task. Furthermore, the optimization of ML algorithms is computationally demanding. Some classical algorithms when they are mapped into the quantum framework became more efficient [194]. For example, the spread of quantum walks in some regimes is significantly faster than random walks, making them more efficient for search type algorithms. Quantum walks are the quantum counterpart of random walks, one of the building blocks used in various classical algorithms as statistical models of real-world processes, and they are also used in the field of quantum computing

[46, 89, 142, 187].

There are two types of quantum walks, *discrete* and *continuous* quantum walks. The first type resembles the classical view of a random walk by tossing a quantum coin. The quantum coin describes the transition probabilities from *heads* ($|+\rangle$) to *tails* ($|-\rangle$) or vice-versa. One of the main differences between quantum and random walks is the possibility for a quantum walk to be in a superposition of *heads* and *tails*,

$$|s\rangle = c_1 |+\rangle + c_2 |-\rangle, \quad (1.14)$$

where c_i s are the complex amplitudes for each coin state. The time evolution of a discrete quantum walk is done by sequentially applying the *coin* operator,

$$|s(n)\rangle = \hat{U}^n |s(n=0)\rangle = \left[\prod_{n=1}^N \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \right] |s(n=0)\rangle, \quad (1.15)$$

where \hat{U} , in this case, is the Hadamard operator and $|s(n=0)\rangle$ the initial state. Tossing a quantum coin one time for an initial state in *heads*, $|s(n=0)\rangle = |+\rangle$, leads to a new state that is in an equal superposition of both coin states,

$$|s(n=1)\rangle = \hat{U} |+\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{|+\rangle + |-\rangle}{\sqrt{2}}. \quad (1.16)$$

In order to simulate the movement of the walk in a one-dimensional chain, we introduce the *shift* operator, \hat{S} . If the quantum coins lands *heads*, the quantum walk moves to the right and if it lands *tails*, it moves to the left,

$$\hat{S} = |+\rangle \langle +| \otimes \sum_n |n+1\rangle \langle n| + |-\rangle \langle -| \otimes \sum_n |n-1\rangle \langle n|, \quad (1.17)$$

here the state $|n\rangle$ indicates the location of the quantum walk in a one-dimensional chain. After applying the Hadamard operator on the initial state once, we obtain a

new quantum state with equal probability in the adjacent site

$$\begin{aligned}\hat{S} \otimes \hat{H} [|+\rangle \otimes |n=0\rangle] &= \hat{S} \left[\frac{|+\rangle \otimes |n=0\rangle + |-\rangle \otimes |n=0\rangle}{\sqrt{2}} \right] \\ &= \frac{|+\rangle \otimes |n=1\rangle + |-\rangle \otimes |n=-1\rangle}{\sqrt{2}},\end{aligned}\quad (1.18)$$

for this example the coin operator is $\hat{U} = \hat{S} \otimes \hat{H}$ and the initial state considered is a quantum walk located in the middle of a 1D chain with *heads*. As it is stated above, the quantum walk has moved with equal probability to the adjacent sites.

The time evolution of *continuous* quantum walks is,

$$|\psi(t)\rangle = \hat{U}(t) |\psi(t=0)\rangle \quad (1.19)$$

where $\hat{U}(t)$ is the evolution operator and it is defined as

$$\hat{U}(t) = e^{-i\hbar\hat{H}t}, \quad (1.20)$$

where \hat{H} is the Hamiltonian operator. For discrete quantum walks the evolution operator is the Hadamard operator, Equation 1.15.

One of the most common continuous quantum walks is described by the Hamiltonian of a one-dimensional chain with hopping only between adjacent sites,

$$\hat{H} = -\gamma [|n+1\rangle \langle n| + |n-1\rangle \langle n|], \quad (1.21)$$

where γ is the transition probability. Expanding the evolution operator in a Taylor series we find that a quantum walk, initially located in the centre of a 1D chain, moves to the adjacent sites with equal probability,

$$\begin{aligned}|\psi(t)\rangle &= [1 - i\hbar\hat{H}t + \dots] |n=0\rangle \\ &= |n=0\rangle + i\hbar\gamma t [|n=1\rangle + |n=-1\rangle] + \dots.\end{aligned}\quad (1.22)$$

It must be pointed out that the formulation of continuous quantum walks is independent of coin states, meaning that the movement of the walk does not depend on tossing a quantum coin. In continuous quantum walks, the Hamiltonian itself is

the graph where the quantum walk will move [4, 63], and the transition probabilities, described by the Hamiltonian, are the vertices between different nodes as it is described in Equation 1.21.

One of the key differences between random and quantum walks is the spread of the walk over time. While the spread over time for random walks scales as $\mathcal{O}(\sqrt{t})$, the spread of quantum walks scales linearly, $\mathcal{O}(t)$, also known as *ballistic expansion* [142]. This unique property has made quantum walks extremely valuable in quantum algorithms [45, 47, 171]. In this thesis we also raise the following question: *Can the spread over time be enhanced for quantum walks beyond the ballistic limit?*

1.3 Quantum simulators of condensed matter

R. Feynman proposed a revolutionary idea. He suggested that one could mimic quantum phenomena with a fully controlled simulator [64]. This simple idea had huge experimental consequences, since, in order to achieve quantum simulators, humans have to be able to fully control quantum systems [30]. Over the last decades, various candidates like, ion traps [21, 166], ultracold atoms and molecules trapped in optical lattices [115], have been proposed to mimic the interactions (or physics) of the solid-state/condensed matter. For example, in 1998 D. Jaksch *et al.* [97] demonstrated that the dynamics of an ultracold dilute gas of bosonic atoms in an optical lattice can mimic the Bose-Hubbard model,

$$\hat{H} = -t \sum_{\langle i,j \rangle} \hat{c}_i^\dagger \hat{c}_j + \frac{U}{2} \sum_i \hat{n}_i (\hat{n}_i - 1) - \mu \sum_i \hat{n}_i, \quad (1.23)$$

where \hat{c}_i^\dagger and \hat{c}_i are the creation and annihilation operators, \hat{n}_i is the number of particles operator in the site i , and $\langle i, j \rangle$ restricts the sum only to adjacent sites. The first term of the Bose-Hubbard model represents the hopping of particles between different lattice sites. The second term describes the interaction between two particles at the same lattice site. The last term is known as the chemical potential and is related to the total number of particles in the system. For ultracold atoms trapped in an optical lattice, the parameters of Equation 1.23 depend on the lattice depth that is produced by the laser used to trap the atoms [97].

In 2009 M. Ortner *et al.* proposed the use of polar molecules to simulate the Bose-Hubbard model [135]. This was experimentally realized in 2013 by B. Yan *et al.* [201]. Trapping polar molecules in optical lattice is arguably one of the greatest experimental achievements in modern physics; however, the filling of optical lattices with current experimental setups is not high, currently limited to $\sim 25\%$ [131].

Polar molecules trapped in different sites of the optical lattice interact by dipole-dipole interaction. The constants of the Bose-Hubbard model simulated with polar molecules can be tuned by an external DC electric field [73]. Since the trap depth of the optical lattice has to be large enough to confine polar molecules, the quantum particles that simulate Bose-Hubbard model are Frenkel excitons, which in the case of polar molecules are rotational excitations [73, 84, 86, 110, 121, 128, 138, 198, 201].

On the other hand, ultracold atoms can be trapped in optical lattices with nearly uniform filling [74, 195]. In 2005 Greismaier *et al.* demonstrated that atoms with high magnetic dipole were trapped in optical lattices [75]. This raises the following question, *Can magnetic atoms trap in an optical lattice be used as quantum simulators for Bose-Hubbard type models?*

1.4 Thesis outline

The work presented in this thesis is divided into two parts. In the first part, Chapters 2-4, illustrates the use of ML to study quantum physics. In the second part of the thesis, Chapters 5 and 6, we study quantum walks and quantum simulators of ultracold atoms with large magnetic dipole.

Chapter 2 contains the introduction to one of the most versatile supervised learning algorithms, GP regression. In this chapter, we also explain why GPs are a more accurate interpolation algorithm than neural networks (NNs) for low dimensional problems. We interpolate the electronic energy for different spatial configurations of the formaldehyde molecule using GP regression.

Chapter 3 introduces the novel idea of combining different ML algorithms with the purpose of solving the inverse scattering problem. It also introduces the Bayesian optimization (BO) algorithm to optimize functions without computing

gradients. The combination of BO and GPs can construct accurate PESs for quantum dynamics observables like the reaction probability for the $H + H_2$ and $OH + H_2$ reactions.

Chapter 4 explores the hypothesis that by extrapolating quantum observables one can discover new phases of matter. Illustrating that GPs with a combination of simple kernels can extrapolate quantum observables and predict the existence of new phases of matter. This chapter studies the evolution of the polaron dispersion as a function of Hamiltonian parameters and shows that the change in the ground state momenta leads to a new phase of matter. Additionally, using the same algorithm one can accurately extrapolate quantum observables where traditional numerical methods lack convergence.

Chapter 5 demonstrates the possibility to enhance the spread of quantum walks for various graphs by allowing the number of walks to be a non-conserved quantity. Illustrating that spread of a quantum walk with number-changing interactions is supra-linear at short times, $\mathcal{O}(t^n)$ where $n > 1$. We also show that for disordered graphs the spread of a quantum walk is larger when number-changing interactions are considered in the Hamiltonian.

Chapter 6 is dedicated to an experimental proposal to study extended Bose-Hubbard models using highly magnetic atoms trapped in optical lattices. The proposal presented uses Zeeman excitations to tune the parameters to construct various Bose-Hubbard type models.

Chapter 2

Gaussian Processes

Sometimes people have to create out of pure necessity.

— Ben Shewry

Machine learning (ML) field is divided into three major areas; supervised, unsupervised, and reinforcement learning [20, 132, 181]. Each of these three fields studies a particular task. In the case of supervised learning, the goal is to find the numerical mapping between an input \mathbf{x}_i and an output y_i . The numerical mapping can be viewed as, $y_i = f(\mathbf{x}_i)$. When the output value y_i is discrete, the problem is known as *classification*. On the other hand, when y_i is continuous is known as *interpolation*. This chapter describes one of the most common supervised learning algorithms Gaussian Process (GP) regression [147].

2.1 Introduction

As in any other supervised learning algorithm, the goal of GP regression is to infer an unknown function $f(\cdot)$ by observing some input, \mathbf{X} , and output data, \mathbf{y} . We denote \mathcal{D} as the training dataset that contains both \mathbf{X} and \mathbf{y} . One of the few differences between GP regression and other supervised learning algorithms, like Neural networks (NNs), is that GPs infer a *distribution over functions* given the training data $p(f|\mathbf{X}, \mathbf{y})$.

Gaussian Processes (GPs) are a probabilistic method and assume that each training point is a random variable and they have a joint probability distribution

$p(f(\mathbf{x}_1), \dots, f(\mathbf{x}_N))$. As its name suggests, GPs assume that the joint distribution is Gaussian and has a mean $\mu(\mathbf{x})$ and covariance $K(\mathbf{x}, \mathbf{x}')$. The matrix elements of the covariance are defined as $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ where $k(\cdot, \cdot)$ is a positive defined kernel function. The kernel function plays a key role as it describes the *similarity* relation between two points. A GP is denoted as,

$$f(\mathbf{x}) \sim GP(\mu(\mathbf{x}), K(\mathbf{x}, \mathbf{x}')) \quad (2.1)$$

In the following sections, we describe the training and prediction of GP regression.

2.2 Gaussian Processes: prediction

As it was mentioned above, a GP is the collection of random variables $f(\mathbf{x}_i)$ that follow a joint distribution. The joint Gaussian distribution between \mathcal{D} , \mathbf{x}_* , and $f(\mathbf{x}_*)$ is of the form,

$$\begin{pmatrix} f(X) \\ f(\mathbf{x}_*) \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu \\ \mu_* \end{pmatrix}, \begin{pmatrix} K & K_* \\ K_*^\top & K_{**} \end{pmatrix} \right) \quad (2.2)$$

where \mathbf{K} is the covariance matrix of the training data X , K_* is the covariance matrix between X and \mathbf{x}_* , and K_{**} is the covariance with respect to itself.

The probability distribution over $f(\mathbf{x}_*)$ can be inferred by computing the conditional distribution given the training data, $p(f_* | \mathbf{x}_*, \mathcal{D})$. Conditioning on the training data is the same as selecting the distribution of functions that agree with observed data points \mathbf{y} . The mean and covariance matrix of the conditional distribution are,

$$\mu(\mathbf{x}_*) = K(\mathbf{x}_*, X)^\top K(X, X)^{-1} \mathbf{y} \quad (2.3)$$

$$\sigma_* = K(\mathbf{x}_*, \mathbf{x}_*) - K(\mathbf{x}_*, X)^\top K(X, X)^{-1} K(X, \mathbf{x}_*) \quad (2.4)$$

where σ_* is the predicted variance for \mathbf{x}_* . The mean of the conditional distribution, equation 2.3, can be rewritten as,

$$\mu(\mathbf{x}_*) = \sum_i d(\mathbf{x}_*, \mathbf{x}_i) y_i = \sum_i \alpha_i k(\mathbf{x}_*, \mathbf{x}_i) \quad (2.5)$$

where $\mathbf{d} = K(\mathbf{x}_*, X)^\top K(X, X)^{-1}$ and $\alpha = K(X, X)^{-1} \mathbf{y}$. The mean of the conditional distribution is a linear combination of the training data, \mathbf{y} , or a linear combination of the kernel function between the training points and \mathbf{x}_* . Function $d(\cdot, \cdot)$ can be understood as a *distance* function. It is important to state that the accuracy of the prediction with GPs directly depends on the size of the training data N and the kernel function $k(\cdot, \cdot)$. In Section 2.4 we illustrate the impact that different kernels on GPs' prediction.

In the GPs' framework, the predicted variance or standard deviation of \mathbf{x}_* represents the uncertainty of the model, Equation 2.4. The uncertainty can be used to sample different regions of the function space to search the location of the minimum or maximum of a particular function. This is the idea behind a class of ML algorithms known as Bayesian optimization (BO), which will be discussed in Chapter 3.

An example of GP regression is illustrated in Figure 2.1 where $f(x) = \frac{x}{10} + x^2 + 10 \sin(\frac{3}{2}x)$. We use the *exponential squared* kernel and 7 training points. In the following section, we describe the most common procedure to train GPs.

2.3 Gaussian Processes: training

The goal of any supervised learning algorithm is to infer a function $f(\cdot)$, as accurately as possible, given some example data. In order to quantify the accuracy of a model we define a *loss* function, \mathcal{L} , for example, the difference between the prediction y_i and the real value \hat{y}_i (training points) such as $\mathcal{L} \approx \sum_i^N (y_i - \hat{y}_i)^2$ or $\mathcal{L} \approx \sum_i^N |y_i - \hat{y}_i|$. The parameters of the model \mathbf{w} and \mathcal{L} are interconnected. To illustrate this idea let us assume that $f(\cdot)$ is a simple linear regression model, $f(x) = a + bx$. The loss function for such model is,

$$\mathcal{L} = \sum_i^N (y_i - \hat{y}_i)^2 = \sum_i^N (a + bx_i - \hat{y}_i)^2. \quad (2.6)$$

From the previous equation, we can observe that the value of \mathcal{L} depends on a and b . It can be argued that when \mathcal{L} is large $f(\mathbf{x}_i) \not\approx \hat{y}_i$. On the other hand when $f(\mathbf{x}_i) \approx \hat{y}_i$ the value of \mathcal{L} will tend to zero. Using a loss function to tune the parameters of $f(\cdot)$ is known as the *training* stage in ML. It must be mentioned

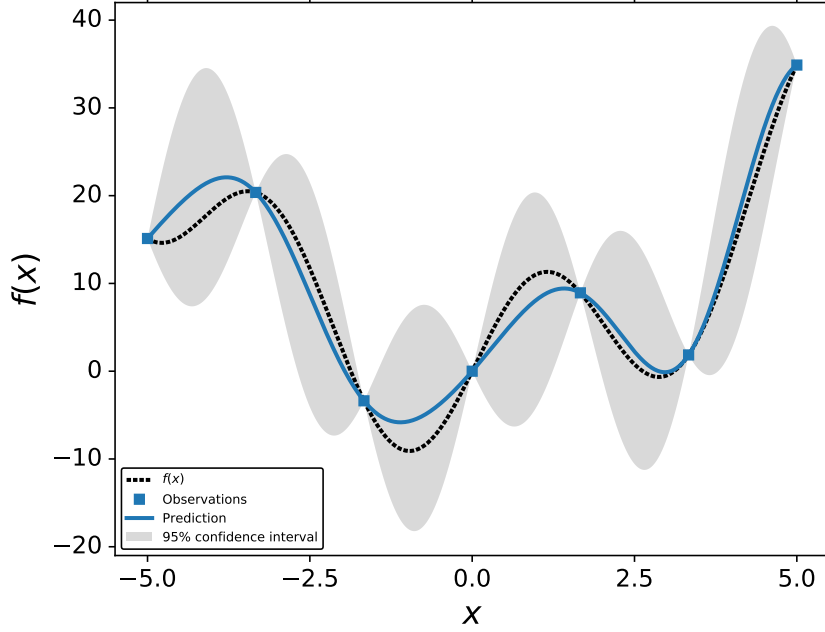


Figure 2.1: Interpolation using a GP model trained with 7 points and an exponential squared kernel. The black dashed line is $f(x) = \frac{x}{10} + x^2 + 10\sin(\frac{3}{2}x)$. The solid blue line is the prediction of the GP model, Equation 2.5. The grey shaded area is the standard deviation of the predicted mean of the GP model, Equation 2.4. The blue square symbols are the training data.

that replicating the training data could also mean that the model “memorized” the training data. This common problem in ML is known as *overfitting*.

GPs models can also be trained using a loss function. GP models are non-parametric models, therefore, the dimensionality of the loss function depends on the number of the parameters of the kernel function. Using a loss function to determine the optimal value for the kernel parameters for non-parametric models is computationally expensive and is prone to overfitting [132, 147]. However, it is possible to train GP methods without a loss function. The most common way to train GPs is by finding the kernel parameters (θ) that maximize the *marginal*

likelihood function,

$$\hat{\theta} = \arg \max_{\theta} p(\mathbf{y}|X, \theta, \mathcal{M}_i) \quad (2.7)$$

where $p(\mathbf{y}|X, \theta, \mathcal{M}_i)$ is the *marginal likelihood* for a given model or kernel function \mathcal{M}_i . Finding the value of θ where $p(\mathbf{y}|X, \theta, \mathcal{M}_i)$ is maximized is known as *type II maximum likelihood* (ML-II) [132, 147]. The *marginal likelihood* or evidence is defined as,

$$p(\mathbf{y}|X, \theta, \mathcal{M}_i) = \int p(\mathbf{y}|X, \mathbf{f}, \mathcal{M}_i) p(\mathbf{f}|\theta, \mathcal{M}_i) d\mathbf{f}. \quad (2.8)$$

In the case of GPs, the *marginal likelihood* has a closed form, Equation 2.9. Finding the kernel parameters that maximize the *marginal likelihood* can be done by maximizing the logarithm of the *marginal likelihood* w.r.t to θ ,

$$\log p(\mathbf{y}|X, \theta, \mathcal{M}_i) = -\frac{1}{2} \mathbf{y}^\top K^{-1} \mathbf{y} - \frac{1}{2} \log |K| - \frac{N}{2} \log 2\pi \quad (2.9)$$

where the first term is known as the data-fit term, the second term reflects the complexity of the model \mathcal{M}_i , and the third term is a constant that depends on the number of training points, N . The value of $\log p(\mathbf{y}|X, \theta, \mathcal{M}_i)$ mainly depends on the data-fit and complexity terms. For instance, for a θ that memorizes the training data the value of $\log |K|$ is large, while $\mathbf{y}^\top K^{-1} \mathbf{y}$ will be small. The tradeoff between the data-fit and the complexity term is key for the optimal value of the kernel parameters.

Standard gradient-based optimization algorithm is the most common method to find the optimal value of the kernel parameters. The logarithm of the *marginal likelihood* has an analytic form, therefore, it is possible to compute the change of $\log p(\mathbf{y}|X, \theta, \mathcal{M}_i)$ as a function of the kernel parameters, $\frac{\partial \log p(\mathbf{y}|X, \theta, \mathcal{M}_i)}{\partial \theta_i}$. For most cases, the function $\log p(\mathbf{y}|X, \theta, \mathcal{M}_i)$ is not a convex function; thus there is a possibility that the optimization algorithm gets trapped in one of the local maxima [147]. For GP regression, the value of θ does depend on the training data \mathbf{y} , and having different training data sets could lead to different values for the kernel parameters.

In the following section, we explain the use of GPs as an interpolation tool to fit

multidimensional functions. In Chapter 3, we illustrate that two-tiered GPs models are capable of solving the so-called *inverse scattering* problem. In Chapter 4, we also illustrate the use of GP regression to predict beyond the training data to discover phases of matter.

2.4 Kernel functions

In the previous sections, we explained how a GP model is trained and also how it can make predictions. We also assumed that GPs need a kernel function in order to construct the covariance matrices \mathbf{K} and \mathbf{K}_* . In this section, we introduce various kernels that are used for training GPs and illustrate how prediction with GPs can drastically change depending on which kernel is used. As mentioned previously, the kernel function should describe the similarity between two points. In kernel regression, two points that are *similar*, under some metric, should have a similar output value y_i , predicted by Equation 2.5.

The kernel function $k(\cdot, \cdot)$ must be a positive definite function; this restriction comes from the requirement to invert the covariance matrix during training and prediction see, Equations 2.5 and 2.9. The covariance matrix must be a symmetric matrix which forces the kernel function to be symmetric too, $k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}', \mathbf{x})$. The Cholesky factorization is the most common algorithm to invert matrices with $\mathcal{O}(N^3)$ complexity where N is the number of training points [147].

All of the kernels that are used in this thesis are *stationary* kernels except for the *linear* kernel. Any stationary kernel can be rewritten in terms of the Mahalanobis distance,

$$r^2(\mathbf{x}, \mathbf{x}') = (\mathbf{x} - \mathbf{x}')^\top \Lambda (\mathbf{x} - \mathbf{x}') = \sum_i^d \ell_i (x_i - x'_i)^2 \quad (2.10)$$

where the Λ is a diagonal matrix with ℓ_i being a particular length-scale for each dimension. The Mahalanobis distance reduces to the square of the Euclidian distance when all $\ell_i = 1$, $r^2(\mathbf{x}, \mathbf{x}') = (\mathbf{x} - \mathbf{x}')^\top (\mathbf{x} - \mathbf{x}')$. Furthermore, when all ℓ_i have the same value, the kernel function is an *isotropic* kernel. In the following sections, we explain some of the most common kernels that are used in GP models.

2.4.1 Constant kernel

The constant kernel, arguably the most simple kernel, assumes that the similarity relation between two points is constant,

$$k_C(r) = \ell \quad (2.11)$$

The kernel parameter ℓ can be optimized by maximizing the logarithm of the marginal likelihood. $k_C(\cdot)$ is a one-time differentiable function.

2.4.2 Square exponential kernel

The square exponential (SE) kernel, also known as the radial basis function (RBF) kernel, is probably the most used kernel in GP regression. The SE kernel is a stationary kernel since it depends on the difference between two points,

$$k_{SE}(r) = \exp(-r^2). \quad (2.12)$$

The kernel parameters of the SE kernel are the matrix elements of the Λ matrix. Each ℓ_i defines the *characteristic length-scale* for each dimension. Depending on the user, the SE kernel can be isotropic, all ℓ_i have the same value, or anisotropic, each ℓ_i has a different value. As it was mentioned in the previous section, the values of Λ_i are optimized by maximizing the logarithm of the marginal likelihood. The training of an isotropic SE kernel is faster since the total number of parameters in the kernel is one, while anisotropic SE kernels have d parameters, d is the dimension of \mathbf{x} . The SE kernel is infinitely differentiable. Figure 2.2 illustrates the square exponential kernel and GPs' prediction with SE kernel as a function of different ℓ s.

2.4.3 Matern kernel

The Matern (MAT) kernels are probably the second most used kernel for GPs,

$$k_{MAT}(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}r}{\ell} \right)^\nu K_\nu \left(\frac{\sqrt{2\nu}r}{\ell} \right) \quad (2.13)$$

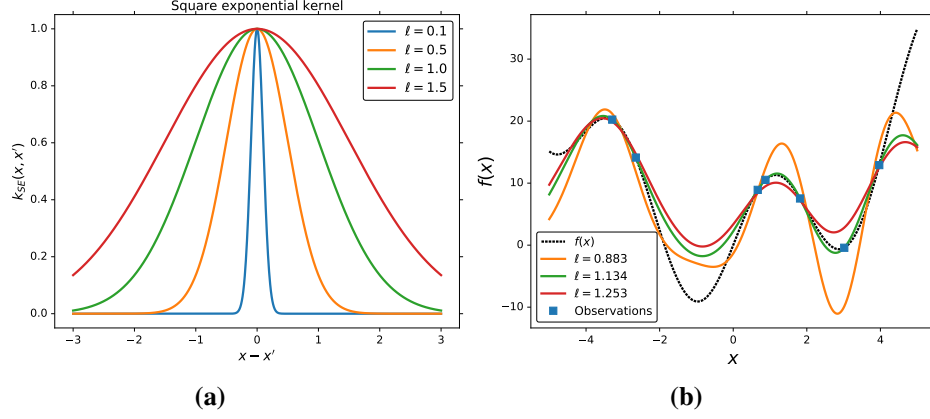


Figure 2.2: a) The square exponential function with different length scales.
b) Prediction of a GP model with the SE kernel using different length scale values.

where ν and ℓ are both positive constants, $\Gamma(\cdot)$ is the gamma function, and K_ν is a modified Bessel function. In the limit of $\nu \rightarrow \infty$ the Matern kernel reduces to the SE kernel. The two most common Matern kernels are for $\nu = \frac{3}{2}$ and $\nu = \frac{5}{2}$,

$$k_{\nu=\frac{3}{2}}(r) = (1 + \sqrt{3}r) \exp(-\sqrt{3}r) \quad (2.14)$$

$$k_{\nu=\frac{5}{2}}(r) = \left(1 + \sqrt{5}r + \frac{5}{3}r^2\right) \exp(-\sqrt{5}r) \quad (2.15)$$

when $\nu = \frac{3}{2}$ the Matern kernel is one-time differentiable, while for $\nu = \frac{5}{2}$ the Matern kernel is twice differentiable. Both kernels, Equations 2.14 and 2.15, are *anisotropic* kernels if r is computed with the Mahalanobis distance. Figure 2.3 illustrates the Matern function and GPs' prediction with Matern kernel for different ℓ s and $\nu = \frac{3}{2}$ or $\nu = \frac{5}{2}$. Figure 2.3 illustrates the MAT kernel and GPs' prediction with these kernels for different values of the kernel parameters.

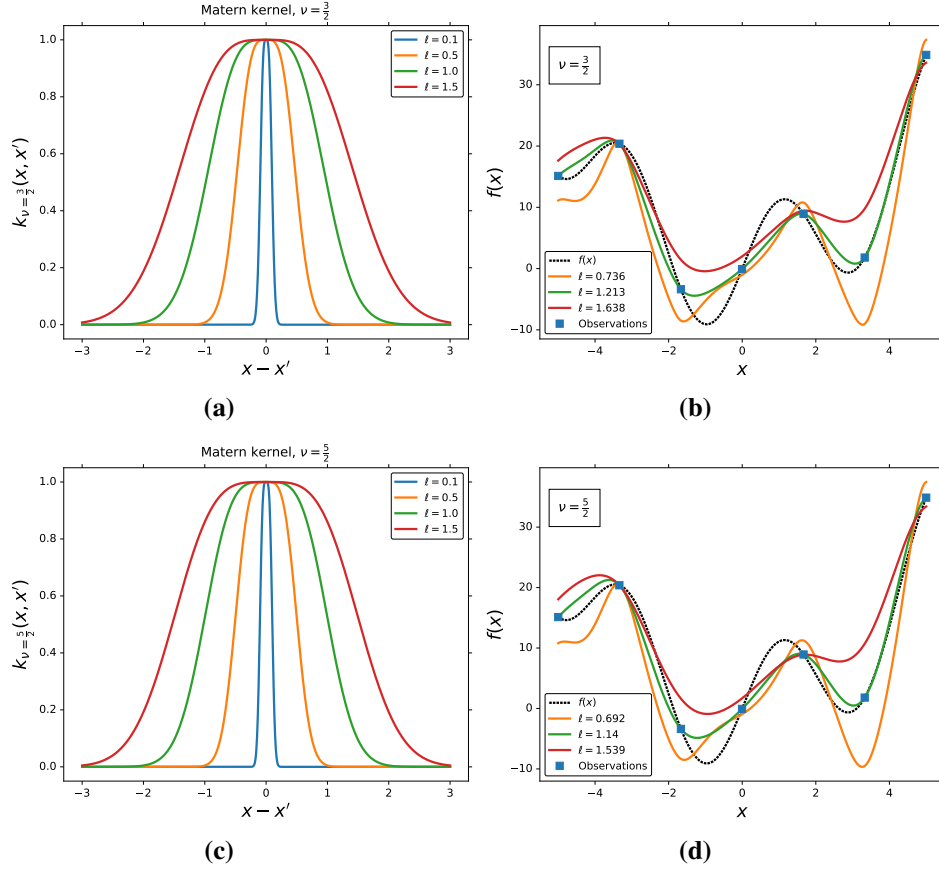


Figure 2.3: a,c) The Matern kernel function for different length scales and $\nu = \frac{3}{2}$ (a) or $\nu = \frac{5}{2}$ (c). b,d) Prediction of a GP model with the MAT kernel using different length scale values and $\nu = \frac{3}{2}$ (b) or $\nu = \frac{5}{2}$ (d).

2.4.4 Rational Quadratic kernel

The Rational quadratic kernel (RQ) is also a stationary kernel,

$$k_{RQ}(r) = \left(1 + \frac{r^2}{2\alpha\ell^2}\right)^{-\alpha} \quad (2.16)$$

where α and ℓ are the kernel parameters. In the case where $\alpha \rightarrow \infty$ the RQ kernel is identical to the SE kernel. r is computed with the Euclidian distance. Figure 2.4 illustrates the RQ kernel and different GPs' prediction for various α and ℓ values.

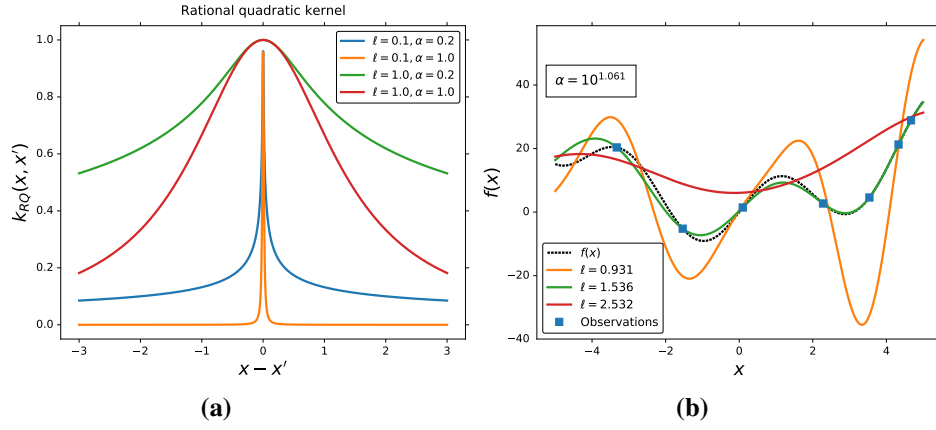


Figure 2.4: a) The rational quadratic function with different ℓ and α values.
 b) Prediction of a GP model with the RQ kernel using different values for ℓ and α .

2.4.5 Periodic kernel

None of the previously mentioned kernels are capable of describing periodic functions. Periodicity can be described by trigonometric function like $\cos(x)$, $\sin(x)$, $\cos^2(x)$ or $\sin^2(x)$. Since any kernel function must be a semi-positive definite function, $\cos^2(x)$ and $\sin^2(x)$ are the only capable trigonometric functions that can be used as kernel functions. The periodic (PER) kernel has the form,

$$k_{PER}(r) = \exp\left(-\frac{2 \sin^2\left(\frac{\pi r}{p}\right)}{\ell^2}\right) \quad (2.17)$$

where p and ℓ are the kernel parameters. p describes the intrinsic periodicity in the data and ℓ is the length-scale parameter. Figure 2.5 shows the exponential sine squared function for different ℓ and p values, and also how GPs' prediction changes for different ℓ and p .

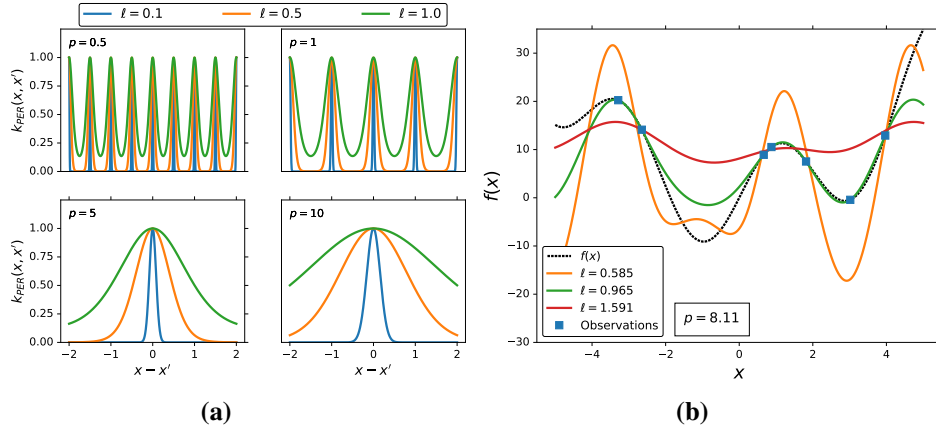


Figure 2.5: a) The exponential sine squared function with different length scales and different periodicities. b) Prediction of a GP model with the PER kernel using different values for ℓ and $p = 8.11$.

2.4.6 Linear kernel

The linear kernel (LIN) is a *non-stationary* kernel and is also known as the dot product kernel,

$$k_{LIN}(\mathbf{x}, \mathbf{x}') = \mathbf{x}^\top \mathbf{x}' + \sigma = \sum_i^d x_i x'_i + \sigma \quad (2.18)$$

where σ is the offset value of the kernel. If $\sigma = 0$ the linear kernel is considered to be *homogenous*. A more general form of the linear kernel is, $k_{LIN}(\mathbf{x}, \mathbf{x}') = \mathbf{x}^\top \Lambda \mathbf{x}'$, where the matrix Λ is a diagonal matrix with unique length-scale parameters for each dimension in \mathbf{x} . The linear kernel is the base of the polynomial kernel,

$$k_{POL}(\mathbf{x}, \mathbf{x}') = (k_{LIN}(\mathbf{x}, \mathbf{x}'))^p = (\mathbf{x}^\top \mathbf{x}' + \sigma)^p = \sigma + \left(\sum_i^d x_i x'_i \right)^p \quad (2.19)$$

where p is the polynomial degree. Figure 2.6 shows the prediction of GP regression with the LIN kernel with different polynomial degrees.

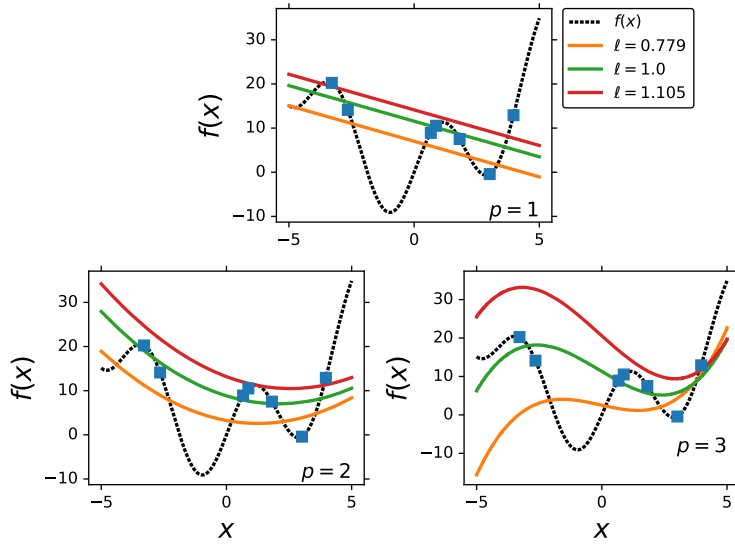


Figure 2.6: Prediction of a GP with the LIN kernel using different σ and polynomial degree.

In the following section, we illustrate the use of GP models to interpolate multidimensional functions. We compare the accuracy of GP regression with NNs, the most well-known ML algorithm for supervised learning.

2.5 Potential Energy Surfaces

One of the greatest challenges in computational chemistry is describing the change of the electronic ground state energy, E_{elec} , as a function of the position of the atoms. For each E_{elec} we need to solve the Schrödinger equation where the Hamiltonian (\mathcal{H}_{elec}) describes a system of an ensemble of N electrons and M nuclei with fix positions. This Hamiltonian is known as the *electronic* Hamiltonian,

$$\mathcal{H}_{elec} = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}}, \quad (2.20)$$

where the first term is the kinetic energy of electrons, the second term is the Coulomb interaction between electrons and nuclei, and the last term is the electron-electron interaction. Z_A is the atomic number of nucleus A. The eigenvalues of \mathcal{H}_{elec} *parametrically* depend on the positions of the nuclei because of the electron-nucleus interaction, the r_{iA} terms in Equation 2.20.

Finding the solutions of \mathcal{H}_{elec} is still one of the most challenging problems in quantum chemistry. In 1998, the Nobel committee for chemistry laureated W. Kohn and J. A. Pople for their contributions in the field of quantum chemistry. W. Kohn is the father of Density functional theory (DFT), which is one of the most known methodologies for solving Equation 2.20, and J. A. Pople is the pioneer in the development of various computational methods in quantum chemistry. Both scientists dedicated their research to find the solutions of the electronic Hamiltonian. It is out of the scope of this thesis to explain or propose new methodologies dedicated to computing the eigenvalues and eigenvectors of the electronic Hamiltonian. Instead, we propose the use of machine learning to reduce the overall computational resources needed to study problems in quantum chemistry.

As it is stated above, the electronic energy depends on the positions of the nuclei. For instance, two hydrogen atoms have different energies for different interatomic distances; quantum chemists call this function a potential energy sur-

face (PES). Figure 2.7 illustrates the prediction of the hydrogen molecule's energy for different interatomic distances using GP regression.

A more general definition of a PES is the function that describes the electronic energy of a system for different spatial configuration of the atoms,

$$E_{elect} = f(\mathbf{R}_0, \dots, \mathbf{R}_M) \quad (2.21)$$

where \mathbf{R}_i is the position of the nucleus i . The function $f(\cdot)$ can be any quantum chemistry method used to evaluate the electronic energy for a given configuration of atoms, $E_{elec} \sim \langle H_{elec} \rangle$. Trying to infer $f(\cdot)$ from some available data is the definition of a supervised learning problem.

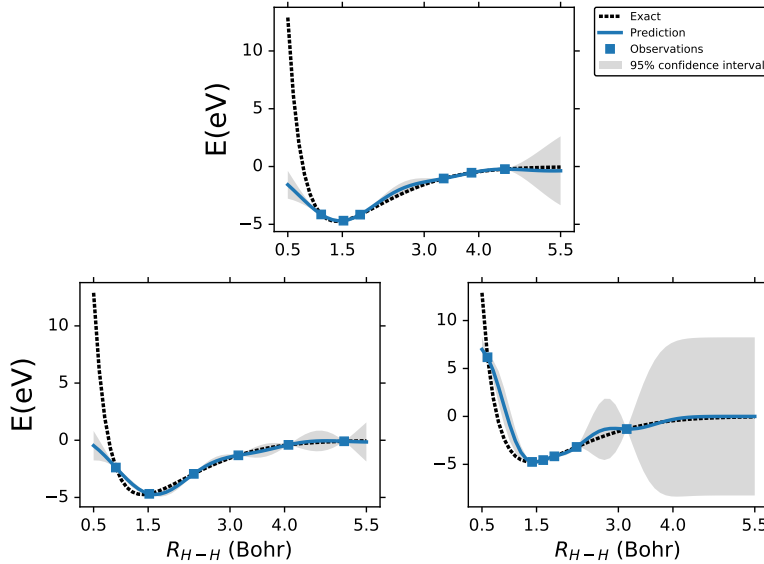


Figure 2.7: H_2 molecule's energy for different interatomic distances predicted with a GP model trained with 6 points and the SE kernel. The exact PES (dashed black line) is from reference [23].

Using ML or any other regression model to infer $f(\cdot)$ is not a novel idea [133, 165, 183]. In 2006 Manzhos *et al.* used NNs to fit PESs [122, 124]. The use of GP regression to construct PESs was introduced by A. P. Bartók *et al.* in Ref. [14] to study bulk crystals where the PESs are used in the calculation of var-

ious properties at high temperatures. Cui *et al.* interpolated the PES of the N_4 molecule [52] using GP regression. However, ML has rapidly evolved in recent years with the creation of new algorithms that could reduce the number of training points needed to make more accurate fits for various PESs [13, 15, 78, 169].

The prediction of bond-breaking energies, unimolecular reactions, vibrational spectra, and reaction rates are a few of the observables that depend upon the accuracy of PES fits [133, 165, 183]. Being able to interpolate the energy of a system for different molecular geometries can also give synthetic chemists information such as reaction mechanisms or transition states [186]. Generally speaking, PESs play a crucial role in chemistry.

In the following section, we present the results obtained using NNs and GP regression to predict the PES for the H_2CO molecule. We also discuss the impact that different kernels have on GP regression and the number of neurons in NNs. Additionally, we summarize the impact the training points have on the test error for both algorithms.

2.5.1 Results

2.5.2 Gaussian Processes regression vs Neural Networks

NNs and GP regression are two of the most known supervised learning algorithms [132, 159]. Over the course of this section, we summarize the results obtained when NNs and GP regression both are used to fit the PES of the H_2CO molecule. We also discuss the importance of the training data size for both algorithms and explain under which circumstances which algorithm is more accurate for interpolating PESs. The results presented in this section are published in reference [102].

The PES of Formaldehyde is a 6D function where the six coordinates are,

- $CO \rightarrow$ distance between the Carbon and Oxygen atoms.
- $CH_1 \rightarrow$ distance between the Carbon atom and one of the Hydrogen atoms.
- $CH_2 \rightarrow$ distance between the Carbon atom and the other Hydrogen atoms.
- $\angle OCH_1 \rightarrow$ angle between the Carbon, Hydrogen and Oxygen atoms.

- $\angle OCH_2 \rightarrow$ angle between the Carbon, second Hydrogen and Oxygen atoms.
- $\angle H_2OCH_1 \rightarrow$ dihedral angle between the OCH planes.

We consider a data set of 120,000 energy points for the H_2CO molecule obtained from reference [36]. Figure 2.8 illustrates the distribution over the energy values of the complete data set. Each training data set used to train every NN or GP model is sampled using the Latin hypercube sampling (LHS) to ensure the points are efficiently spread [126].

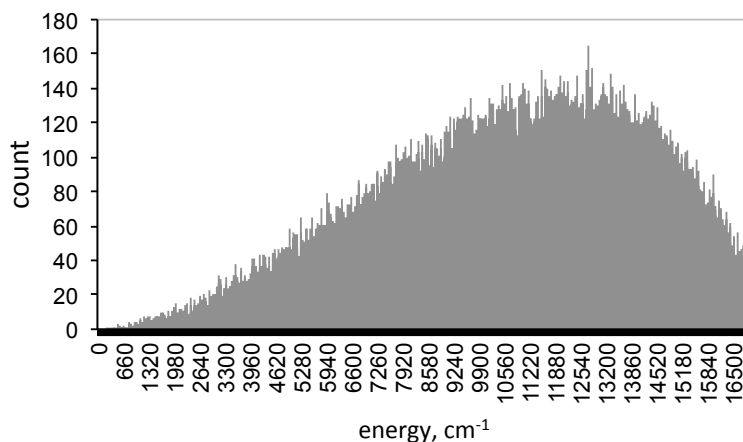


Figure 2.8: Distribution of energy values of the PES dataset. Figure from reference [102]

NNs are a powerful and complex supervised learning algorithm [132]. We consider one of the simplest NNs architectures, single layer NN with sigmoid activation functions. Even in the limit of a single layer of neurons, NNs can be used as an interpolation algorithm. We study the effect that the number of neurons has in the accuracy of single layer NNs. Each NN is trained using 200 epochs with the Levenberg-Marquardt algorithm. All the results that are reported here were obtained using *MATLAB*'s Neural Network ToolboxTM. The data used to train NNs is scaled to $[0, 1]$. The results obtained using NNs are summarized in Table 2.1.

GPs are a robust supervised learning algorithm; we test their accuracy by considering different kernel functions and various number of training points. The optimization of the kernel parameters is carried out by maximizing the logarithm of

Table 2.1: RMSE (test errors computed on 120,000 points) of the PES obtained with the NNs for different training points (N_{pts}). The number in the parenthesis are the number of neurons used in that particular NN. $\langle NN \rangle_{10}$ is the average RMSE for 10 NNs with different sets of N_{pts} . The values are in cm^{-1} .

N_{pts}	NN			$\langle NN \rangle_{10}$		
313	198.00(20)	103.93(30)	87.77(40)	119.11(20)	53.97(30)	43.90(40)
625	21.12(50)	12.91(75)	12.03(100)	13.36(50)	7.52(75)	6.53(100)
1250	9.29(70)	5.74(100)	4.38(150)	5.74(70)	3.36(100)	2.54(150)
2500	4.59(100)	2.43(150)	1.12(250)	2.27(100)	1.23(150)	0.86(250)

Table 2.2: RMSE (test errors computed on 120,000 points) of the PES obtained with GP regression. All GPs use the SE kernel function. N_{pts} is the number of training points. $\langle GP \rangle_{10}$ is the average RMSE for 10 GP models with different sets of N_{pts} . The values are in cm^{-1} .

N_{pts}	GP	$\langle GP \rangle_{10}$
313	29.09	17.18
625	5.98	3.87
1250	2.17	1.13
2500	1.08	0.62

the *marginal likelihood*, Section 2.3. All GP models are constructed with *scikit-learn*'s Gaussian Process Regression library [137]. Interpolation results using GPs are summarized in Table 2.2.

As discussed, using different kernel functions leads to different accuracies in the predictions of GPs. For example, the RMSE of a GP model with the $MAT(\nu = \frac{5}{2})$ kernel and 2000 training points is 43.63 cm^{-1} , while with the same number of training points but using the RQ kernel the RMSE is 49.87 cm^{-1} . The RMSE for both GPs is computed over 30,000 points. The kernel function is not the only variable that GPs' accuracy is sensitive to, different representations of the data also make GP regression more or less accurate. For instance, if we train a GP model with inverse of bond-lengths, like CO , CH_1 and CH_2 , the RMSE decreases to 29.95 cm^{-1} using the $MAT(\nu = \frac{5}{2})$ kernel and 2000 training points.

NNs are a parametric model, meaning that the number of parameters is fixed, unlike GP regression. We consider a single layer NN, which only single hyper-

parameter is the number of neurons. By increasing the number of neurons we are able to decrease the RMSE, Table 2.1. NNs become a more robust interpolation algorithm when the number of neurons increases which means more parameters. However, we also notice that a large number of neurons does not necessarily reduce the RMSE. For example, a NN trained with 625 points but 75 and 100 neurons have similar RMSEs. When the number of parameters for NN is large more training points are needed for better training a NN. The total number of parameters of a single layer NN is $(d + 1) * n_H$, where d is the dimensionality of the problem and n_H is the number of neurons. A well-known problem in NNs is that for a fixed number parameters, more training data could lead to a better optimization which means a lower RMSE. We must remember that the loss function to train NNs is an m -dimensional function where m is the number of parameters, whereas the dimensionality of the cost function for GPs is the number of kernel parameters. Consequently, GP models are often easier to train than NNs.

One of many applications of fitting PESs is the ability to predict the vibrational spectrum of molecules. We compute the error of the predicted vibrational spectrum using the NNs and GPs trained to interpolate the PES of the H_2CO molecule. The vibrational spectrum is determined using space-fixed Cartesian kinetic energy operator and Gaussian basis functions (SFGB) [123]. Figure 2.9 shows that the GPs fit clearly outperforms the NNs fit. The RMSE between the accurate vibrational frequencies and the ones computed using a GPs fit is 0.05 cm^{-1} , while NNs' RMSE is 0.30 cm^{-1} , when both models are trained with 625 points. GP models are not only more accurate than NNs in the interpolation of PESs but also in predicting vibrational spectra. GP regression also offers the advantage of requiring fewer training points than NNs.

2.6 Summary

We have shown that it is possible to fit PESs with supervised learning. Furthermore, we compared two of the most common, novel regression tools, NNs and GP models, to interpolate the energy for a spatial arrangement of atoms. We demonstrated that GP models are a more accurate interpolation tool over NNs for low dimensional systems. Additionally, we explored the impact that the number of training

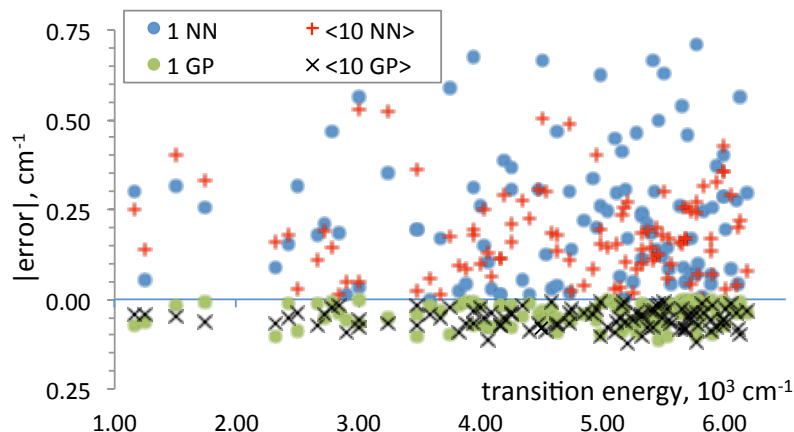


Figure 2.9: Absolute errors in transition frequencies computed on PESs fitted to 625 points with different methods (NN: top half, GP: bottom half) vs the spectrum computed on the reference PES. Figure from reference [102]

points has on the prediction’s accuracy for both methods, NNs and GP regression. Vibrational spectra computed using both regression methods were compared as a second test to determine which method is more accurate with GPs still outperforming NNs.

The computational time required to evaluate a single point (\mathbf{x}_*) depends on the architecture of the models; for example, a single layer NNs’s prediction time scales as $\mathcal{O}(n_H d + n_H)$ where d is the dimensionality of the problem and n_H are the number of neurons. For GPs, because of the K^{-1} term in Equation 2.5 the prediction scales as $\mathcal{O}(N^3)$ where N is the number of training points. However, inversion of the covariance matrix K^{-1} just needs to be done once since it does not depend on \mathbf{x}_* . In the limit of few training points, GPs’ prediction is faster than NNs only if the number of neurons needed for accurate prediction is large.

One of the unanswered questions in quantum chemistry is, which regression model needs the least number of energy points to make accurate predictions? For some systems, each energy point used to train a regression model may require high computational resources and the resources needed to have a large number of points is a problem. We argue that for low dimensional systems GPs are accurate

interpolation algorithms that require less training points, Tables 2.1 and 2.2.

The prediction accuracy of single layer NNs depends on the number of neurons. In Table 2.1 we showed that for a fixed number of training points the RMSE of a single layer NN decreases as n_H increase. NNs with a large number of neurons requires more training points in order to become accurate; for example the NN’s RMSE with $n_N = 100$ and train with 625 and 1250 points changes from 12.03 cm^{-1} to 5.74 cm^{-1} . We must remember that the dimensionality of the loss function used to train single layer NNs directly depends on the number of neurons, $\dim(\mathcal{L}_{NN}) = n_H d + n_H$. The minimizer of \mathcal{L}_{NN} represents the best parameters of, for the NN. The dimensionality of the negative log-marginal likelihood, GPs’ loss function, depends only on the number of parameters in the kernel. If we consider an isotropic kernel the $\dim(\mathcal{L}_{GP}) = 1$, whereas if we consider an isotropic kernel $\dim(\mathcal{L}_{GP}) = d$. Given these points, GP models can be trained more accurately partly because the low dimensionality of their loss function which makes the minimization more efficiently.

The accuracy of fitted PESs can also be evaluated by computing some physical observables that depend on the PES, for instance, the vibrational spectrum of a molecule. We compute the absolute difference between the accurate vibrational spectrum of Formaldehyde and the one predicted using a PES trained with NNs or GPs. Figure 2.9 illustrates that the vibrational spectrum predicted using a GPs trained with only 625 points and the SE kernel is more accurate than with a NN with 100 neurons.

In recent years, ML algorithms have shown the potential to tackle high-dimension complex problems like playing the Go game, self-driving cars, among others, making it a popular/novel tool in many scientific areas like physics and chemistry. In this chapter, we discussed how using 625 energy points and two of the most simple ML algorithms, it is possible to reduce the computational time to study many-body physics like the prediction of the vibrational spectrum of a molecule. It should be noted that the architecture of the NNs used in our research is not the NN that defeated Lee Sedol and future research should be done on how deep-NNs [72], more than a single layer of neurons, can help chemists to construct the PESs for molecules or proteins where GPs cannot be used.

Chapter 3

Bayesian Optimization

To those who do not know mathematics it is difficult to get across a real feeling as to the beauty, the deepest beauty, of nature. If you want to learn about nature, to appreciate nature, it is necessary to understand the language that she speaks in.

— Richard P. Feynman

The optimization of machine learning algorithms is one of the most important research areas since the parameters of each supervised learning algorithm needs to be trained [177]. However, the same tools can be used to optimize physical-chemistry problems. Over the course of this chapter, we introduce one of the most novel ML algorithms to optimize black-box functions known as BO. Using BO and GPs, we present an innovative algorithm that allows us to fit PESs with accurate quantum dynamics observables. For the $\text{H} + \text{H}_2 \rightarrow \text{H}_2 + \text{H}$ system we show that a GP model trained with 30 points is enough to predict accurately the reaction probability for this system. We also illustrate that the same algorithm works for higher dimensional systems like $\text{OH} + \text{H}_2 \rightarrow \text{H}_2\text{O} + \text{H}$, where a GP model trained 290 points fits a PES that again predicts accurate reaction probabilities.

3.1 Introduction

In Chapter 2, we introduced two of the most common supervised learning algorithms, NNs and GP models. We compared the accuracy of both methods by interpolating the PES of the H_2CO molecule when different number of training points

were used. We conclude that GP models need fewer training points in order to accurately interpolate the PES of the H₂CO molecule. Also, we discussed that for low dimensional problems, GP models are more accurate than NNs. Furthermore, in this chapter, we exemplify how GP models can also be used to solve optimization problems.

Here we present again the mean (μ_*) and standard deviation (σ_*) of GPs used in the prediction,

$$\mu(\mathbf{x}_*) = \sum_i^n d(\mathbf{x}_*, \mathbf{x}_i) y_i = \sum_i^n \alpha_i k(\mathbf{x}_*, \mathbf{x}_i) \quad (3.1)$$

$$\sigma(\mathbf{x}_*) = K(\mathbf{x}_*, \mathbf{x}_*) - K(\mathbf{x}_*, X)^\top K(X, X)^{-1} K(X, \mathbf{x}_*) \quad (3.2)$$

where $K(\cdot, \cdot)$ is the covariance matrix with matrix elements $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ and $k(\cdot, \cdot)$ is the kernel function. The values of the kernel parameters are optimized by maximizing the log-marginal likelihood function, Section 2.3.

The goal of an *optimization problem* is to find the best solution among all potential solutions. In the field of ML, the optimization problem is associated with the search for the values of the parameters of a model that better described the problem, e.g. minimizing a loss function [132, 177]. Synthetic chemistry also has optimization problems, for example varying the reaction conditions to increase percent yield. The optimization of a function is also a supervised learning problem, but instead of finding the best global representation of function $f(\cdot)$, the goal is to find the \mathbf{x} where $f(\cdot)$ is minimum,

$$\mathbf{x}_* = \arg \min_{\mathbf{x}} f(\mathbf{x}). \quad (3.3)$$

The most common optimization algorithm for continuous functions is *gradient descent* (GD) [132]. GD algorithm is designed to minimize a function iteratively by displacing the current point in the direction of the *negative* gradient of the function,

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \eta \nabla f(\mathbf{x}_n) \quad (3.4)$$

where the parameter η is known as the *learning rate*. η is also related in the trade-off between exploitation and exploration and plays a key role in the convergence

of the algorithm [132]. For example, when η is small GD is exploiting $\mathbf{x}_{n+1} \approx \mathbf{x}_n$; where as for $\eta \gg 0$ is related to exploration. GD is one of the first optimization algorithms used to train NNs [158].

GD has been a widely successful optimization algorithm. However, not every function can be optimized using GD. For example, there is no analytic function that describes the relation between the percent yield given some experimental conditions for a chemical reaction, therefore one can not use GD to increase the percent yield. There are many other problems that are described by non-analytic functions or *black-box* functions, where evaluations are point-wise. BO is designed to tackle the optimization of black-box functions where gradients are not available. For obvious reasons, trying to find the minimum of $f(\cdot)$ by randomly sampling is not the smartest strategy, since it may take a large number of evaluations from $f(\cdot)$ before finding the minimum. BO tries to infer the location of the minimum of a black-box function by proposing a smarter iterative sampling scheme. In the case of GD we assume that the gradient gives us the information of where to sample the next point in order to get closer to the minimum. Considering that black-box functions do not have a gradient, it is necessary to propose a metric that quantifies the informational gain as a function of the space. The core of BO relies in two components,

1. $\mathcal{F}(\cdot) \rightarrow$ model that mimics the black-box function.
2. $\alpha(\cdot) \rightarrow$ acquisition function that quantifies the information gain for a given point.

To mimic the unknown function $f(\cdot)$ we can use any supervised learning algorithm, like NNs. However, if $\mathcal{F}(\cdot)$ is not capable to learn at every iteration, we may waste some of the evaluations because of the lack robustness of the model. GP models are a great candidate for $\mathcal{F}(\cdot)$ due to the accuracy and robustness to interpolate any continuous function. Additionally, the ability of GP models to quantify the prediction’s uncertainty $\sigma(\mathbf{x})$ without the need of extra data is what makes them the strongest candidate for BO.

Figure 3.1 illustrates how BO works to find the minimum of $f(\cdot)$ without using gradients. The maximum of the acquisition function is the query point where the black-box function is evaluated next, $f(\mathbf{x}_{n+1})$, and at each iteration we add the

new point \mathbf{x}_{n+1} to the training data and retrain the GP model. Algorithm 1 is the pseudocode of BO [25, 170]. In Section 3.2 we explain different acquisition functions that are used in BO.

Algorithm 1 Bayesian optimization

Input: Acquisition function $\alpha(\cdot)$, black-box function $f(\cdot)$, data set \mathcal{D} .

- 1: **for** $n = 1, 2, \dots$, **do**
- 2: Optimize the acquisition function,

$$\mathbf{x}_{n+1} = \arg \max_{\mathbf{x}} \alpha(\mathbf{x}, \mathcal{D})$$

- 3: Evaluate $f(\mathbf{x}_{n+1})$.
 - 4: Augment data $\mathcal{D}_{n+1} = \{\mathcal{D}_n, (\mathbf{x}_{n+1}, f(\mathbf{x}_{n+1}))\}$.
 - 5: Update model.
-

3.2 Acquisition function

BO is an optimization algorithm designed for problems where gradients are not available. As it was mention above, the acquisition function is designed to represent which point in the space has the most information. By iteratively evaluating the black-box function where the acquisition function is maximum we learn a more certain representation of $f(\cdot)$ where the minimum could be. There are many different acquisition functions, here we cover the three most used,

1. Probability of improvement (PI)
2. Expected Improvement (EI)
3. Upper confidence bound (UCB)

3.2.1 Probability of Improvement

In 1964 H. Kushner proposed as an acquisition function to maximize the probability of improvement, the probability when $f(x) > \tau$ [112]. H. Kushner showed that if $f(x)$ is Gaussian distributed, $P(f(x) > \tau)$ can be written as,

$$\alpha_{PI}(\mathbf{x}; \tau) := P(f(x) > \tau) = \Phi\left(\frac{\mu(\mathbf{x}) - \tau}{\sigma(\mathbf{x})}\right) \quad (3.5)$$

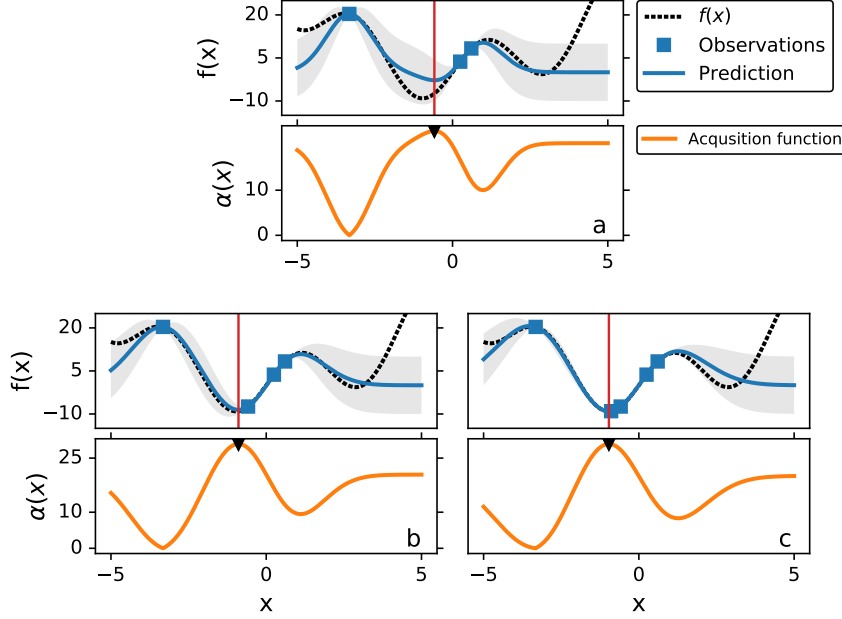


Figure 3.1: Example of the BO algorithm over three iterations. Blue markers are GP's the training data, blue curves are GP's prediction and black dashed line is the black-box function $f(x) = \frac{x}{10} + x^2 + 10\sin(\frac{3}{2}x)$. The yellow solid curves are the acquisition function, and \blacktriangledown are the maximum of the acquisition function at each iteration, \mathbf{x}_{n+1} .

where $\Phi(\cdot)$ is the normal cumulative distribution function, $\mu(\cdot)$ and $\sigma(\cdot)$ are the predicted mean and standard deviation of a GP model trained with the data set \mathcal{D}_n , and τ is the target value. Since the goal of BO is to find τ we can approximate it with the best known value in the set \mathcal{D}_n , for example $\tau = \max_{i=1:N} y_i$. If y_{n+1} is greater than the current value of τ , we update τ . PI is known as a greedy acquisition function, however if we relax the value of τ by adding a constant, ϵ , we can make exploratory moves. For example, in Figure 3.2 we illustrate how the maximum of $\alpha_{PI}(\cdot)$ changes for different values of ϵ .

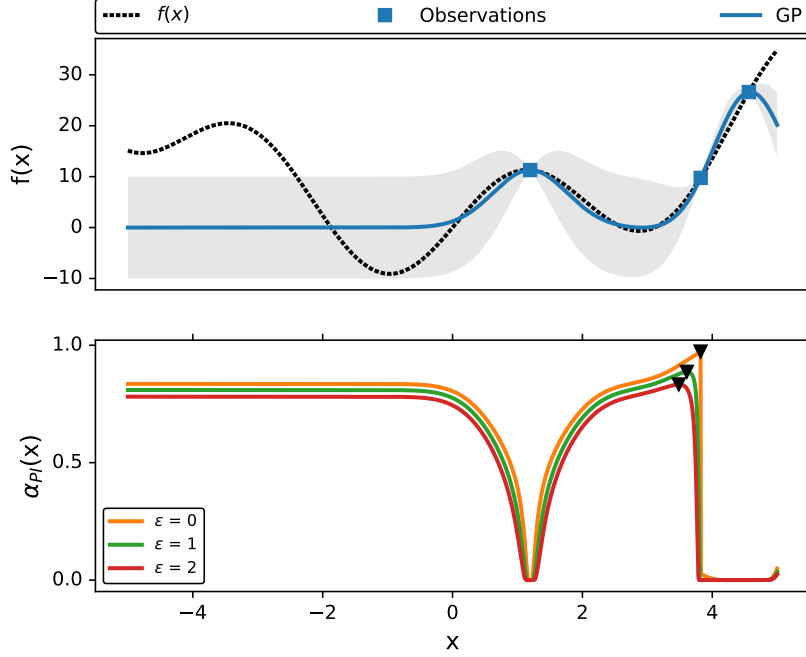


Figure 3.2: Illustration of the PI acquisition during BO. We also show the PI function for different values of ϵ .

3.2.2 Expected Improvement

The expected improvement is one of the most known acquisition functions. The improvement is defined as the difference between the predicted point and the best known point (τ),

$$I(\mathbf{x}; \tau) = \max\{0, \mu(\mathbf{x}) - \tau\} \quad (3.6)$$

and the expected improvement is defined as,

$$\mathbb{E}[I(\mathbf{x}; \tau)] = \int I(\mathbf{x}) p(I(\mathbf{x}) | \tau, \mu(x), \sigma(\mathbf{x})) dx \quad (3.7)$$

where $p(I(\mathbf{x}) | \tau, \mu(x), \sigma(\mathbf{x}))$ is the probability distribution over the improvement. If we consider that $p(I(\mathbf{x}) | \tau, \mu(x), \sigma(\mathbf{x}))$ is Gaussian distributed with a mean $\mu(x) -$

τ , the expectation integral has a closed-form expression [25, 100],

$$\alpha_{EI}(\mathbf{x}; \tau) = \begin{cases} 0 & \text{if } \sigma(\mathbf{x}) = 0 \\ (\mu(\mathbf{x}) - \tau)\Phi(z(\mathbf{x}; \tau)) + \sigma(\mathbf{x})\phi(z(\mathbf{x}; \tau)) & \text{if } \sigma(\mathbf{x}) > 0 \end{cases} \quad (3.8)$$

where $z(\mathbf{x}; \tau) = \frac{\mu(\mathbf{x}) - \tau}{\sigma(\mathbf{x})}$. $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ are the mean and standard deviation of a GP. $\Phi(\cdot)$ is the normal cumulative distribution and $\phi(\cdot)$ is the normal probability distribution. Figure 3.3 illustrates the EI function.

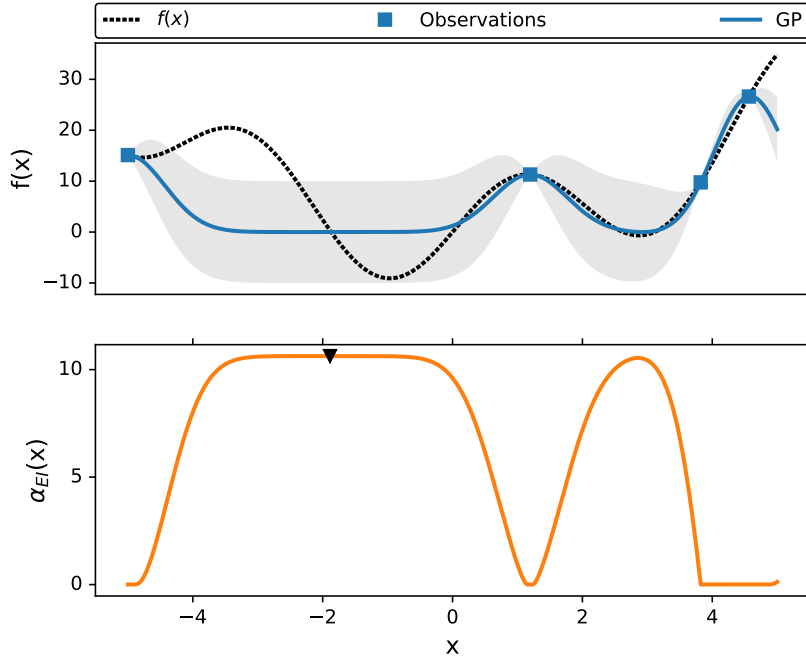


Figure 3.3: Illustration of the EI acquisition during BO.

3.2.3 Upper Confidence Bound

The last acquisition function that we present in this dissertation is the upper confidence bound (UCB) function,

$$\alpha_{UCB}(\mathbf{x}) = \mu(\mathbf{x}) + \kappa\sigma(\mathbf{x}) \quad (3.9)$$

where $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ are the predicted mean and standard deviation of a GP model. The constant κ plays a key role since it describes the trade-off between exploration and exploitation. When κ is small the acquisition function relies more on the mean of the GP which is associated with exploitation, opposite to when $\kappa \gg 0$ which makes $\alpha_{EI}(\cdot) \approx \sigma(\cdot)$ and we explore the input space. Figure 3.4 illustrates how the κ parameter changes the sampling schedule of a black-box function. There have been many proposals on how to change the value of κ as a function of the number of iterations; for example, set κ to a large value at the beginning for more exploration and reduce its value at the end for exploitation [25, 170, 178]. The UCB function can also be used with other ML algorithms that are capable of computing the prediction's uncertainty $\sigma(\cdot)$, like bayesian-NN [134, 172].

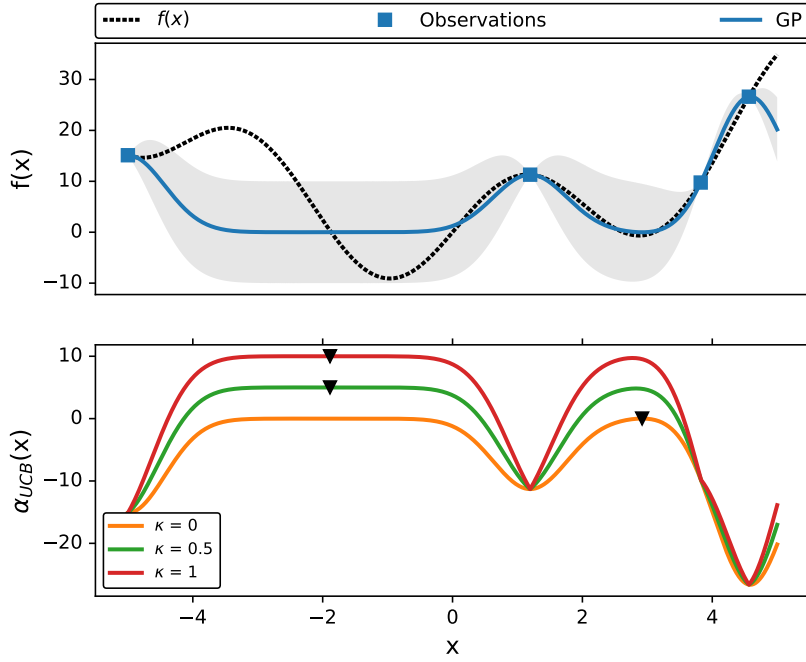


Figure 3.4: Illustration of the UCB acquisition during BO. We also show how the UCB function for different values of κ

In the following section we present the results for fitting PES through the error of quantum dynamic calculations. We also show that by combining GP regression

and BO one can solve the *inverse scattering* problem.

3.3 Results

The main idea of Chapter 2 is to illustrate that PESs can be represented by a supervised learning algorithm like NNs or GPs. We conclude that GPs are more accurate regression models when the training data is few and the dimensionality of the system is low; for instance, the RMSE of a GP and NN trained with 1250 points is 1.13 cm^{-1} and 2.54 cm^{-1} restively, Tables 2.1 and 2.2. We also compare the vibrational spectra of a PES fit with GP regression and NN with respect to the real spectrum, and our results demonstrate that when a PES is fitted using a GP model the vibrational spectra is also more accurate [102].

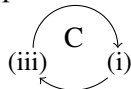
PESs have a more profound meaning in the field of quantum molecular dynamics. PESs are used to reduce the computational complexity of quantum dynamics calculations. PESs are also used to understand reaction mechanisms or to study transition states. To obtain the PES yielding the most accurate description of the experimental dynamical results, one could start with a rough PES based on a small number of *ab initio* calculations and systematically improve the surface by one of the following two iterative procedures:



where (i) is computing the potential energy for a wide range of relative atomic coordinates by an *ab initio* quantum chemistry method; (ii) fitting an analytical representation of the PES; (iii) integrating the Schrödinger equation for the motion of the atomic nuclei on this PES.

In the first approach (cycle A), one would compute and add more *ab initio* points to the PES at each iteration, thus placing an emphasis on the parts of the configuration space most relevant for the dynamics. In the second approach (cycle B), one could attempt to solve the inverse scattering problem by first computing the global PES and then modifying the analytical fit of the PES through an iterative procedure [61, 92, 127]. However, there are two problems that make these iterative approaches unfeasible in the application to quantum reaction dynamics. Firstly,

step (ii) above, i.e. finding the best analytical representation of the PES for multi-atom systems is extremely laborious, almost always requiring manual work [24, 51, 79, 91, 133]. Secondly, step (iii) takes minutes to hours of computation time, which severely limits the number of loops in any of the optimization cycles above. Here we propose a more efficient optimization cycle,



where step (ii) can be easily eliminated by fitting PESs using GP models. Cycle C can be implemented for low dimensional reaction systems by means of a two-tiered GP regression. Where the first GP is used to fit PESs, and the second GP is used to optimize the location and magnitude of the energy points to produce the PES which gives the best description of the observable.

We consider two chemical reactions:



and for all the results display in the following sections. We compute the reaction probabilities using the time-dependent wave packet dynamics approach described in Ref. [41, 55, 146, 190], explicitly accounting for all degrees of freedom. The basis sets of the reaction dynamics calculations are chosen to ensure full convergence of the dynamical results.

Both of these chemical reactions have been studied before [42, 180]. For reaction (1), Su *et al* [180] computed the reaction probabilities with the 3D PES from Ref. [23], constructed using an analytical fit to 8701 *ab initio* energy points. For reaction (2), Chen *et al* [42] computed the reaction probabilities with the 6D PES constructed using NN fits to $\sim 17\,000$ *ab initio* calculations.

In the following sections we highlight how it is possible to optimize cycle C when the quantum dynamics results are and are not known, and how to overcome the inaccuracy of the quantum chemistry calculations to get a better fit of any PES with experimental data.

3.3.1 Fitting known quantum dynamics results

In this section, we explain the algorithm used to optimize cycle C when the reaction probabilities are known. First we randomly select a small number of energy points ($n = 20$ for the 3D surface and $n = 280$ for the 6D surface) from the original PESs [23, 42] and construct an approximate PES with a GP, Equation 3.1. We denote this GP model of the surface by $\mathcal{G}(n)$. When n is small any regression model is likely to be highly inaccurate and the quantum dynamics calculation with this surface is also expected to produce highly inaccurate results, Figure 3.5.

Given $\mathcal{G}(n)$, we then ask the following question: if one *ab initio* point is added to the original sample of few points, where in the configuration space should it be added to result in the maximum improvement of the quantum dynamics results? In principle, this question could be answered without ML by a series of quantum dynamics calculations based on $\mathcal{G}(n+1)$ with the added point moved around the configuration space. However, such an approach would be completely unfeasible as it would require about 10^d dynamical calculations for each added *ab initio* point, where d is the dimensionality of the configuration space. We also need to quantify the improvement gained by adding a single point, which can be done with an utility function $f(\cdot)$. Care must be taken since there are many ways to quantify the improvement e.g. the RMSE function between a $\mathcal{G}(n)$ ’s reaction probability and the exact value from either from a calculation with the full surface or from an experiment.

We seek the utility function’s minimum, $\arg \min f(\cdot)$, since it describes which point shapes the PES in the most accurate manner. Finding the minimum of $f(\cdot)$ could not have been done without using current ML algorithms like BO. Figure 3.5a illustrates the change in the reaction probability as a function of a single point added to a fix set of points. Depending on which point is considered in $\mathcal{G}(n+1)$ the improvement in the reaction probability changes, Figure 3.5b.

As we state above, BO requires a model that mimics the utility function $f(\cdot)$. We denoted as \mathcal{F} , the GP model that learns the utility function and is also used to construct the acquisition function that is needed in BO. This GP model is trained by approximately $15 \times d$ quantum dynamics calculations. We use as utility function

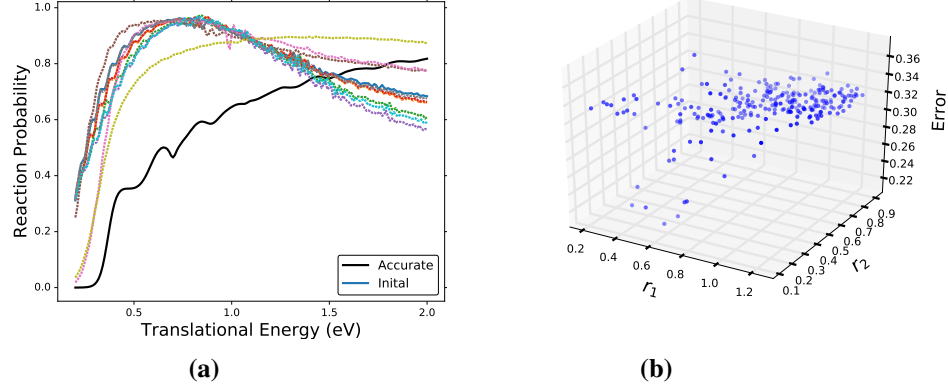


Figure 3.5: a) The reaction probability for the $\text{H}_2 + \text{H} \rightarrow \text{H} + \text{H}_2$ reaction as a function of the collision energy. The black solid curve—accurate calculations from [23]. The dashed curves—calculations based on GPs PESs obtained with 22 + 1 *ab initio* points, where the 22 initial points (solid blue curve) are fixed and the extra point is randomly sampled in the configuration space. b) The error of the reaction probabilities as a function of the location extra point added to the original 22 points.

the root mean square error (RMSE),

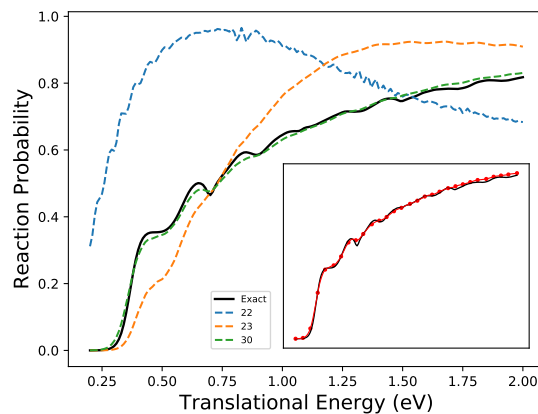
$$\varepsilon(\mathbf{x}_{n+1}) = \sqrt{\sum_i (z(e_i) - \hat{z}(e_i))^2} \quad (3.12)$$

where $z(e_i)$ is the value of the reaction probability at different collision energie e_i with a surface by $\mathcal{G}(n+1)$. \mathbf{x}_{n+1} is the position of the added point and $\hat{z}(\cdot)$ is the accurate reaction probability. For reaction (1) the values of $\hat{z}(\cdot)$ are reported in Ref. [180] and for reaction (2) in Ref. [42]. Figure 3.5b displays the change in the RMSE for different locations of the added point to $\mathcal{G}(n+1)$.

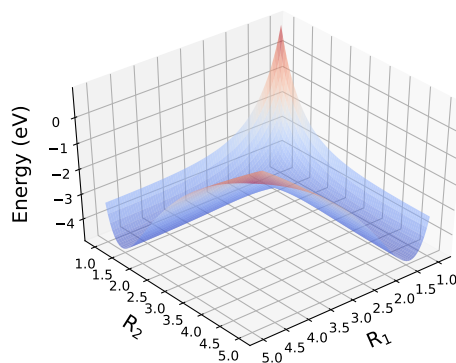
For both systems, reactions (1) and (2), we use the UCB acquisition function, Equation 3.9, with $\kappa = 0.005$. After a fixed number of iterations in the BO algorithm, 35 quantum dynamics calculations for H_3 and 60 calculations for OH_3 , we choose the point where the utility function is minimum, and added to the n set. With this procedure, the minimization of $\mathcal{F}(\cdot)$ for each value of n requires about 35 quantum dynamics calculations for reaction (1) and 60 calculations for

reaction (2). We carried iteratively the same algorithm until we converge to a PES that has an accurate reaction probability. Sequentially adding the points that makes the highest improvement corresponds to a class of ‘greedy’ reinforcement learning strategies in ML [181].

Figure 3.6a illustrates the performance of this algorithm in search of the best PES for reaction (1). As can be seen, the starting model $\mathcal{G}(\cdot)$ of the PES based on 22 *ab initio* points produces highly inaccurate results, but the BO scheme converges to the correct PES after only 8 iterations. Accurate results for the reaction probabilities (green dashed line) can be achieved with a GP model trained with only 30 *ab initio* points. Figure 3.6b shows the model $\mathcal{G}(\cdot)$ of the PES obtained with $n = 30$ *ab initio* points, illustrating that Equation 3.1 produce a physical surface. Figure 3.7 illustrates the dependence of BO as a function of the value of κ in $\alpha_{UCB}(\cdot)$.



(a)



(b)

Figure 3.6: a) The reaction probability for the $\text{H}_2 + \text{H} \rightarrow \text{H} + \text{H}_2$ reaction as a function of the collision energy. The black solid curve – accurate calculations from Ref. [23] based on the surface constructed with 8701 *ab initio* points. The dashed curves – calculations based on the GP PES obtained with 22 *ab initio* points (blue); 23 points (orange), 30 points (green) and 37 points (inset). The RMSE of the results with 37 points is 0.009. b) The GP model of the PES for the H_3 reaction system constructed with 30 *ab initio* points. This surface yields the quantum dynamics results shown by the green curve in the upper panel.

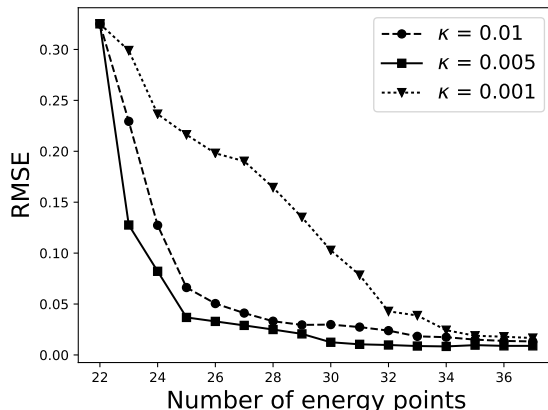


Figure 3.7: Convergence of the RMSE of the probabilities for $\text{H}_2 + \text{H} \rightarrow \text{H} + \text{H}_2$ reaction with the number of BO iterations as a function of κ in UCB.

Figure 3.8 illustrates the performance of this algorithm for reaction (2). As the dimensionality of the configuration space increases, so does the number of points required to represent accurately the PES. Nevertheless, accurate results for the reaction probabilities (green dashed line) are obtained with 290 *ab initio* points, much smaller than the set of $\sim 17,000$ points used in previous work [42] to construct the PES with a NN fit. The RMSE of the reaction probabilities thus obtained is 0.0076. Note that, as any supervised learning technique, this algorithm is guaranteed to become more accurate when trained with more *ab initio* points.

3.3.2 Fitting without known quantum dynamics results

In the previous section we assume that Equation 3.12 accurately describes the improvement when accurate dynamics results are known. However, for all systems, the quantum dynamics results may not be known before hand. Therefore, we propose to use an utility function that quantifies the maximum improvement between the observables computed with $\mathcal{G}(n+1)$ and $\mathcal{G}(n)$ at each iteration. This is justified by the observation that $\mathcal{G}(n \rightarrow \infty)$ must produce the best surface so the maximum improvement of the surface at each iteration is achieved when \mathbf{x}_{n+1} corresponds to the maximum of Equation 3.12, where $\hat{z}(\cdot)$ are the results from $\mathcal{G}(n)$. The min-

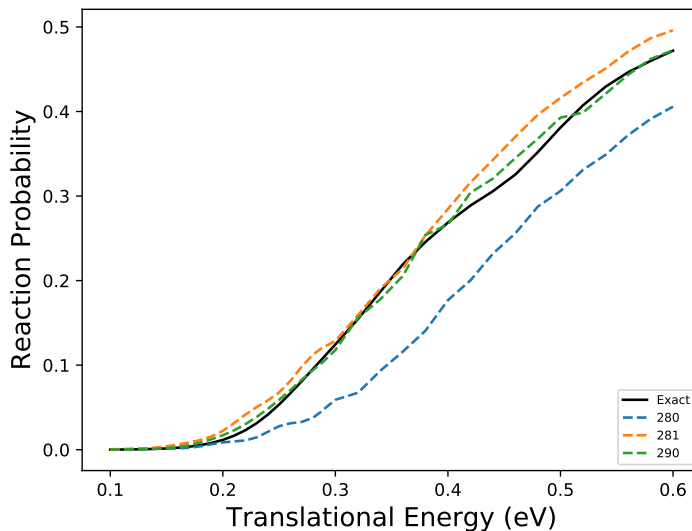


Figure 3.8: The reaction probability for the $\text{OH} + \text{H}_2 \rightarrow \text{H} + \text{H}_2\text{O}$ reaction as a function of the collision energy. The black solid curve – accurate calculations from Ref. [42] based on the surface constructed with ~ 17000 *ab initio* points. The dashed curves – calculations based on the GP PES obtained with 200 *ab initio* points (blue); 280 points (orange) and 290 points (green). The RMSE of the 290-point result is 0.0076.

imum of Equation 3.12 indicates that both $\mathcal{G}(n+1)$ and $\mathcal{G}(n)$ produce the same reaction probabilities.

To illustrate the validity of this assumption, we show in Figure 3.9 a series of computations as functions of n , showing the convergence of the iterative calculations to the accurate results (black solid curve). Figure 3.9 shows that the optimization loop converges to the accurate PES after 48 iterations. We emphasize that the accurate results (black curve) were not used in any way in this calculation. Care must be taken since many of the $\mathcal{G}(n+1)$ could lead to unphysical reactive probabilities. To avoid this issue we penalized the utility function by decreasing its value, preventing the BO algorithm from sample points from that space’s region.

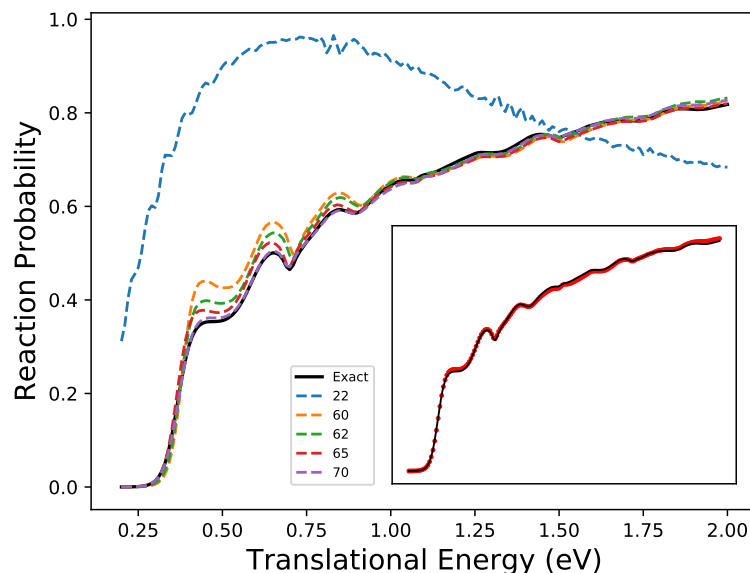


Figure 3.9: The reaction probabilities for the $\text{H}_2 + \text{H} \rightarrow \text{H} + \text{H}_2$ reaction as functions of the collision energy. The black solid curve – accurate calculations from Ref. [23]. The dashed curves – the results of iterative calculations maximizing the difference between the reaction probabilities in subsequent iterations. The black curve is **not used** for these calculations. The inset shows the agreement between the reaction probabilities (red symbols) based on the GP approach after 48 iterations (total of 70 *ab initio* points) and the exact results.

3.3.3 The inverse scattering problem

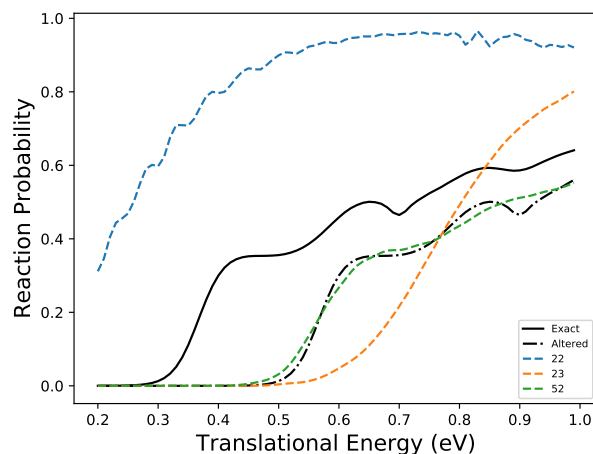
Integrating of Schrödinger’s equation for the motion of the atomic nuclei on any PES is computationally demanding, step (iii); thus the computational resources needed in quantum dynamic calculations increases as a function of the system’s size. To overcome this problem, there is a proposal to infer the shape of a PES using experimental quantum dynamical observables. This scheme is known as the *inverse scattering* problem; however, due to its complexity, this problem has not been fully solved for large systems. Unfortunately, it is impossible to compute the potential energy in step (i) without errors and any theoretical predictions of observ-

ables are subject to uncertainties stemming from the errors of quantum chemistry calculations. These errors become more significant for heavier atoms and are often unknown. Therefore, it would be ideal to develop an approach that either bypasses quantum chemistry calculations or corrects the errors of the *ab initio* calculations. This could be achieved by deriving the empirical PES from the experimental data [50, 61, 92, 127].

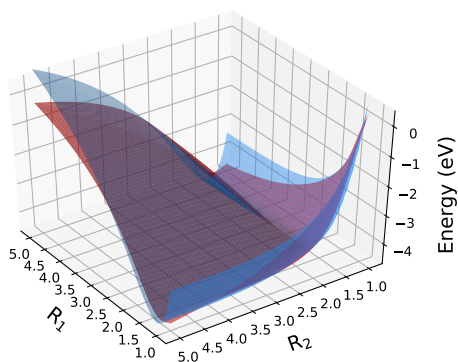
Here, we extend the previous sections to construct a PES that, when used in quantum scattering calculations, reproduces an arbitrary set of observables. We first modify the exact scattering results of Figure 3.6 by shifting along the energy axis and randomly modulating the black curve. This produces an arbitrary energy dependence of the reaction probabilities shown by the dot-dashed curve in Figure 3.10a. The goal is to construct a PES that reproduces these arbitrarily chosen reaction probabilities. Note that the dot-dashed curve extends the interval of energies, where the reaction probability is zero, which means that the PES for this reaction must have a higher reaction barrier and cannot be reproduced with the original PES for H_3 .

We assume that a small ensemble of energy points is known from some (not necessarily accurate) quantum chemistry calculation, denoted by E_i° . As before, this ensemble serves as a starting point to fit a PES with a $\mathcal{G}(n)$. However, in order to allow for the improvement of the PES and to overcome the inaccuracy of the quantum chemistry calculations we vary E_i° by the amount of $\varepsilon(\mathbf{x})$, $E_i = E_i^\circ + \varepsilon(\mathbf{x})$. Intuitively the variable ε depends on the atomic coordinates because depending on the space’s region we may need to increase or decrease the energy to change the shape of the PES.

In the previous sections $\mathcal{F}(\cdot)$ only depended on $\mathbf{x}_{\mathbf{n}+1}$, because we assume that $E_i^\circ = E_i$. A more robust approximation is to learn the dependence of the utility function for both variables, ε and \mathbf{x} . Since ε increases the dimensionality of the utility function by one, we need more data so that BO’s $\mathcal{F}(\cdot)$ accurately describes the minimum of the utility function for $\mathbf{x}_{\mathbf{n}+1}$ and the value of ε . Figure 3.10a shows that this algorithm converges to the arbitrarily modified reaction probabilities after 32 iterations, producing a PES depicted in Figure 3.10b.



(a)



(b)

Figure 3.10: a) The reaction probabilities for the modified $\text{H}_2 + \text{H} \rightarrow \text{H} + \text{H}_2$ reaction as functions of the collision energy. The black dot-dashed curve is obtained by a modification of the previous results (black solid curve) involving a translation along the energy axis. The ML models are trained to obtain the PES that would describe the new reaction probabilities. The green dashed curve is a results of such training after 30 iterations, which produces a surface constructed with 52 *ab initio* points. b) Comparison of the original PES (blue) with the new PES (red) found by the BO algorithm. The new PES yields the reaction probabilities described by the green dashed curve in the upper panel. The RMSE of the results shown by the green dashed curve is 0.016 eV.

3.4 Summary

The results we present in this chapter can be summarized as,

- GPs are a robust regression algorithm that can also be used to minimize/-maximize black-box functions.
- BO is a powerful algorithm that without relying on gradients is capable of minimize/maximize functions like Equation 3.12.
- Accurate quantum reactive scattering results can be obtained with PES based on a very small number of *ab initio* points,
 - 30 for reaction: $\text{H} + \text{H}_2 \rightarrow \text{H}_2 + \text{H}$
 - 290 for reaction: $\text{OH} + \text{H}_2 \rightarrow \text{H} + \text{H}_2\text{O}$
- Accurate quantum reactive scattering results can be obtained with PES based on 70 points for reaction $\text{H} + \text{H}_2 \rightarrow \text{H}_2 + \text{H}$ when the quantum dynamics results are not known.
- Our approach is robust enough that it can bypass the inaccuracy of the quantum chemistry calculations by also learning to correct E_i .
- Most of the energy points that are used to fit PES are unnecessary for accurate quantum dynamics calculations like the reaction probabilities.
- We introduce a two-tiered GP model, which gives the reaction probabilities as explicit functions of the position of an added *ab initio* point.

As it was mentioned, evaluating the energy with accurate *ab initio* quantum chemistry methods and quantum dynamical observables are both computationally demanding. The results we present indicate that ML reduces the total computation of similar problems by using interpolation methods that required less training points (Chapter 2), and better search algorithms like BO. Incorporating ML in the quantum dynamics calculations is a research problem that should be considered, since it could reduce the computational resources needed to evaluate quantum dynamics observables for large systems where is currently impossible.

GPs are not well suited regression algorithms for high dimensional problems, therefore the two-tiered GPs regression method we proposed has some flaws to fit the PES for systems with large number of atoms. However the same strategy can be used with other Bayesian optimization algorithms like the ones proposed by R. P. Adams *et al.* suggested in Ref. [172], *PHOENICS* in Ref. [83] or using random forest (RF) [94].

The optimization of cycle C using GP models and BO show that the total number of points needed to fit a PES with accurate quantum dynamics observables can be reduced. However, the total number of points used by BO to minimized the utility function could be reduced by studying different acquisition functions or change the value of κ as a function of the iteration number [25, 170, 178].

The optimization algorithm for cycle C is greedy since at each iteration we only consider points to train $\mathcal{G}(n)$ with data that represent the minimum of the utility function. However, we did not explore the impact of a non-greedy policy in both the accuracy and the number of points required for a $\mathcal{G}(n)$ with accurate reaction probabilities. A non-greedy policy, know as ϵ -greedy [181], allow us to make new moves that could lead to a better over all strategy that could reduce the number points or achieve even more accurate quantum dynamics observables.

One of the key elements in the scheme we proposed is the utility function. $f(\cdot)$ describes the quality of the points, e.g. Equation 3.12 computes the difference in the complete range of the collision energy. However we could benefit from the intrinsic correlations that multiple utility functions could have, e.g partition the collision energy to have multiple error functions as utility functions. This new strategy can be tackled using multi-task BO [182] and could cut down the number of quantum dynamics calculations. Also, multiple quantum dynamics observables could be used to fit a single PES instead of a single one which is the strategy we present.

We must emphasize that the most important concept of this chapter is the ability to optimize black-box functions with the current ML tools without using gradients. There are many open problems in chemistry and physics that can be formulated in terms of the optimization of black-box functions, e.g. improve the percentage yield in chemical reactions [152], or in cold atoms physics tuning the experimental parameters to enhance the lifetime of quantum particles. It must be remembered,

however, that while using BO the utility function should capture the problem in the most robust manner so that ML can help us solve new problems.

Chapter 4

Extrapolation of quantum observables

The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed.

— Paul A. M. Dirac

In the field of many-body physics quantum observables like ground state energies, particle correlations, particle densities, to mention few, are key to understand the underline physics of a phenomenon. The computation of quantum observables can be understood as the prediction of a scalar black box function where the goal is to infer such function given some training data,

$$\langle \hat{O} \rangle \sim f(\mathbf{x}) \quad (4.1)$$

where $f(\cdot)$ represents the numerical or experimental method used to evaluate the quantum observable of operator \hat{O} . In the case of electronic ground state energies, the operator \hat{O} is the electronic Hamiltonian, Equation 2.20, and \mathbf{x} is all the distances between the nuclei in the molecule. For a condensed matter system, \mathbf{x} is the value of the parameters of the Hamiltonian; for example, for the Hubbard

model \mathbf{x} can be the values of the onsite energy, hopping amplitude, to mention few, Chapter 6.

Using ML to infer quantum observables has been proved to be a novel approach to study many-body physics. For example, J. Carrasquilla *et al.* used NNs to characterize many-body quantum states from different phases of matter for various spin Hamiltonians [35]. The quantum observable used to train the NNs is a discrete variable that labels the phase of matter, i.e. ferromagnetic or antiferromagnetic. As L.-F. Arsenault *et al.* stated in Ref.[8], the use of ML algorithms should be focused to solve true quantum many-body problems. Here the authors predict the self-energy $\Sigma(\omega)$ or the local Green function $G(\omega)$ for the Hubbard model with second-neighbour hopping for a 3D cubic lattice using kernel Ridge regression (KRR) method [132]. The phase diagram of this many-body system has one transition, from the metallic to the Mott insulator phase. For each phase, a different KRR method was used to predict the quantum observable. The phase transition as a function of the Hamiltonian parameters is also learned by a classification algorithm, decision forest [8].

Predicting discrete or continuous quantum observables is a daily task for computational physicists; naturally, the use of ML algorithms has reduced the computational effort needed to study many-body physics. However, when the goal is to discover new physics like unknown phases of matter or the value of quantum observables where the current experimental/numerical tools cannot be assessed, more novel and powerful ML algorithms have to be developed. Over the course of this chapter, we present the idea of applying ML algorithms like GP models to extrapolate quantum observables to discover new phases of matter.

This chapter exemplifies the power of GP models with non-single kernels by extrapolating the energy dispersion, denoted as $E(\cdot)$, of a single particle dressed by bosons in an infinite 1D lattice. We also illustrate a non-bias manner to construct more robust kernels that can interpolate between different phases of matter and also extrapolate where training data is not used.

4.1 Introduction

It is common in quantum physics to encounter a problem described by a Hamiltonian like,

$$\hat{H} = \hat{H}_0 + \alpha \hat{H}_1 + \beta \hat{H}_2 \quad (4.2)$$

whose eigenspectrum can be computed/measured in certain limits of α and β e.g. $\alpha = 0$ or $\alpha \gg \beta$, but not at arbitrary values of α and β . For such problems, it is necessary to interpolate the quantum properties of a system between the known limits, if there are more than one like we did in Chapters 2 and 3. If only one limit is accessible, one must extrapolate from this limit, which we will exemplify below, Section 4.2. Both the interpolation and extrapolation become exceedingly complex if the system properties undergo sharp transitions at some values of α and/or β . Sharp transitions separate the phases of the Hamiltonian (4.2), as shown schematically in Figure 4.1. Since the properties of the quantum system vary drastically in the different phases [160], an extrapolation of quantum properties across phase transition lines is generally considered unfeasible [160]. We consider the phase diagrams of polaron Hamiltonians, some of which have three phases as depicted in Figure 4.1, and show that the sharp transitions in these diagrams can be identified by machine learning models trained with data from only one of the phases.

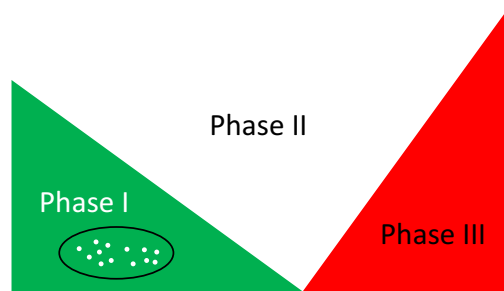


Figure 4.1: Schematic diagram of a quantum system with three phases.

To illustrate the possibility to extrapolate quantum observable using ML, we consider a generalized polaron model describing an electron in a one-dimensional

lattice with $N \rightarrow \infty$ sites coupled to a phonon field:

$$\hat{\mathcal{H}} = \sum_k \epsilon_k c_k^\dagger c_k + \sum_q \omega_q b_q^\dagger b_q + V_{e-ph}, \quad (4.3)$$

where c_k and b_q are the annihilation operators for the electron with momentum k and phonons with momentum q , $\epsilon_k = 2t \cos(k)$ and $\omega_q = \omega = \text{const}$ are the electron and phonon dispersions, and V_{e-ph} is the electron - phonon coupling. We choose V_{e-ph} to be a combination of two qualitatively different terms $V_{e-ph} = \alpha H_1 + \beta H_2$, where

$$H_1 = \sum_{k,q} \frac{2i}{\sqrt{N}} [\sin(k+q) - \sin(k)] c_{k+q}^\dagger c_k (b_{-q}^\dagger + b_q) \quad (4.4)$$

describes the Su-Schrieffer-Haeger (SSH) [125] electron - phonon coupling, and

$$H_2 = \sum_{k,q} \frac{2i}{\sqrt{N}} \sin(q) c_{k+q}^\dagger c_k (b_{-q}^\dagger + b_q) \quad (4.5)$$

is the breathing-mode model [113]. The eigenstates of the model (4.3) are polarons that are known to exhibit two sharp transitions as the ratio α/β increases from zero to a large value [87]. At $\alpha = 0$, the model (4.3) describes breathing-mode polaron, which have no sharp transitions [69]. At $\beta = 0$, the model (4.3) describes SSH polarons, which exhibit one sharp transition in the polaron phase diagram [125]. At these transitions, the ground state momentum of the polaron changes abruptly, as shown in Figure 4.3 and in the black curve in Figure 4.4 for the SSH polaron.

In the following section, we first illustrate how a GP model with a combination of simple kernels became a more robust supervised learning model. Furthermore, we explain how with GP regression it is possible to extrapolate quantum observables of any quantum system like the polaron Hamiltonian that can lead to the discovery of new phases of matter.

4.2 Combination of kernels

The efficiency of GP models depend on the kernel function and the size of the training data set [147]. In the limit of a large number of training points, any GP

model with a stationary kernel produces accurate results. However, there is no restriction to use more than one kernel to construct the covariance matrices, thus we could ask *what is there to gain by combining kernels?*. In this section, we investigate the premise of using GPs with more than one kernel.

For many problems, there might be a suitable kernel form that describes the system more accurately, for example, the periodic kernel for recurrent data. However, to custom made a kernel function for every single problem is not trivial. One of the restrictions is that the kernel function used in GP models must be a positive-defined function [147]. It is possible to construct more robust kernels that satisfied the positive-defined restriction using two simple operations [58–60],

$$k_{\alpha} + k_{\beta} = k_{\alpha}(\cdot, \cdot) + k_{\beta}(\cdot, \cdot) \quad (4.6)$$

$$k_{\alpha} \times k_{\beta} = k_{\alpha}(\cdot, \cdot) \times k_{\beta}(\cdot, \cdot) \quad (4.7)$$

where $k_i(\cdot, \cdot)$ is any of the simple kernels introduced in Chapter 2 .

To illustrate the power of combining kernels, we go back to fitting the PES for H_2CO [102], Chapter 2. For a GP model with k_{MAT} the RMSE is 29.95 cm^{-1} , while using $k_{MAT} \times k_{RQ}$ the RMSE is 10.97 cm^{-1} ; both GP models were trained with the same 2,000 points, and tested on a set of 30,000 points. By including a new kernel we reduced the RMSE by a factor of two, however, we must consider that a GP model with $k_{MAT} \times k_{RQ}$ has more parameters than a GP model with k_{MAT} , making it a more robust supervised model.

The core of GP models is the kernel function which must capture the similarity relation between two points. So far we have shown that GPs with single and multiple kernel functions are accurate interpolation models, but can GPs work for extrapolation?. Extrapolation is defined as the ability to predict beyond the training data range. In principle, if we could design or propose a kernel function that captures the correlation of the data in a robust manner GPs would be capable of extrapolating observables. For example, a GP model with a periodic kernel is capable of extrapolating if there is some intrinsic periodicity in the data.

As we already state in Chapter 2, there are two types of kernels, *stationary* and *non-stationary*. In the case of the stationary kernels, it indisputable that they are not suited for extrapolation since the kernel function for two distant points should

be zero. Consequently, if the distance between the training data and the new point, \mathbf{x}^* , is large, the prediction with GPs is zero, as can be seen in Figure 2.2a. On the contrary, a non-stationary kernel, like the linear kernel Equation 2.18, allows the mean of the predicted distribution to be non-zero, as is illustrated in Figure 2.6.

In the following sections, we illustrate the possibility of using multiple kernels to extrapolate quantum observables to predict phase transitions.

4.2.1 Kernel combination to predict a single phase transition

D. Duvenaud *et al.* proposed that the multiplication of any stationary kernel with the linear kernel, k_{LIN} , leads to a GP model with a non-zero mean that can be used for extrapolation [58, 60]. Here we illustrate that this simple kernel i.e. $k_{MAT} \times k_{LIN}$, is robust enough to make extrapolation predictions and to envision changes in the shape of the energy dispersion of the SSH polaron.

The first case we study is the energy dispersion of the polaron, when β/α is held fixed to $\beta/\alpha = 1.0$. Using Equation 4.1 we defined the lowest polaron eigenenergy or polaron dispersion as,

$$E(K, \alpha, \beta, \omega) = \langle \hat{\mathcal{H}}(K, \alpha, \beta, \omega) \rangle \quad (4.8)$$

where K is the total momentum, α and β are the polaron constants and ω is the phonon frequency. $E(\cdot)$ depends on the parameters of the Hamiltonian 5.1, K, α, β and ω . We denote value of the K where the polaron dispersion is minimum as,

$$K_{GS} = \arg \min_K E(K, \alpha, \beta, \omega). \quad (4.9)$$

The first case we consider is to extrapolate the polaron energy for $\beta/\alpha \geq 1.0$, where $K_{GS} = 0$ [87, 113]. We used GP regression with $k_{MAT} \times k_{LIN}$ to illustrate that it is possible to predict $E(\cdot)$ for $\lambda_{SSH} > 1.3$ while only being trained from $0 < \lambda_{SSH} < 1.3$, Figure 4.2. We defined $\lambda_{SSH} = 2\alpha^2/t\hbar\omega$ as the effective SSH coupling strength.

Predicting energy dispersion with roughly the same shape is still challenging; however, the energy dispersions of the SSH model, for $\beta = 0$, suffer a change of

shape as a function of α . Due to the change in the polaron dispersion, the ground state momentum gets shifted from $K_{GS} = 0$ to $K_{GS} > 0$. Figure 4.3 illustrates that in the case for $\omega/t = 3$ the change in K_{GS} happens right at $\lambda_{SSH} = 0.6$, denoted as λ_{SSH}^* . Indicating that the SSH polaron undergoes a phase transition [125].

We use GP models to extrapolate $E(\cdot)$ for $\lambda_{SSH} > \lambda_{SSH}^*$ using the same kernel, $k_{MAT} \times k_{LIN}$, to exhibit that simple ML algorithm can predict phase transitions without knowing there exists such phases of matter. Here, λ_{SSH}^* denotes the value where the shift in K_{GS} happens. The shape of the energy dispersion before and after λ_{SSH}^* changes, Figure 4.3; nevertheless, GP model trained with only values of λ_{SSH} before the transition ($\lambda_{SSH} < \lambda_{SSH}^*$) are capable to learn the intrinsic trend in $E(\cdot)$ that shifts the ground state momenta, Figure 4.2. The training data and the kernel are equally important in GP regression. When GP models have more information like an extra λ_{SSH} point, they became more accurate as it is displayed in Figure 4.4 where two different GP models were considered. The first GP model was trained from $0.2 \leq \lambda_{SSH} \leq 0.5$ and the second from $0.2 \leq \lambda_{SSH} \leq 0.6$. All the GP models are trained with only 250 total points. To quantify the accuracy of the GPs prediction we computed the RMSE between the predicted energy dispersion and the accurate energy dispersion [16, 87, 174, 175]. In Figure 4.2 we can observe that GP models can make accurate extrapolation for at least $|\lambda_{SSH} - \hat{\lambda}_{SSH}| \approx 1$, where $\hat{\lambda}_{SSH}$ is the greatest/smallest training value.

In the following section, we consider alternative kernel combinations to predict the complete phase diagram of the polaron model.

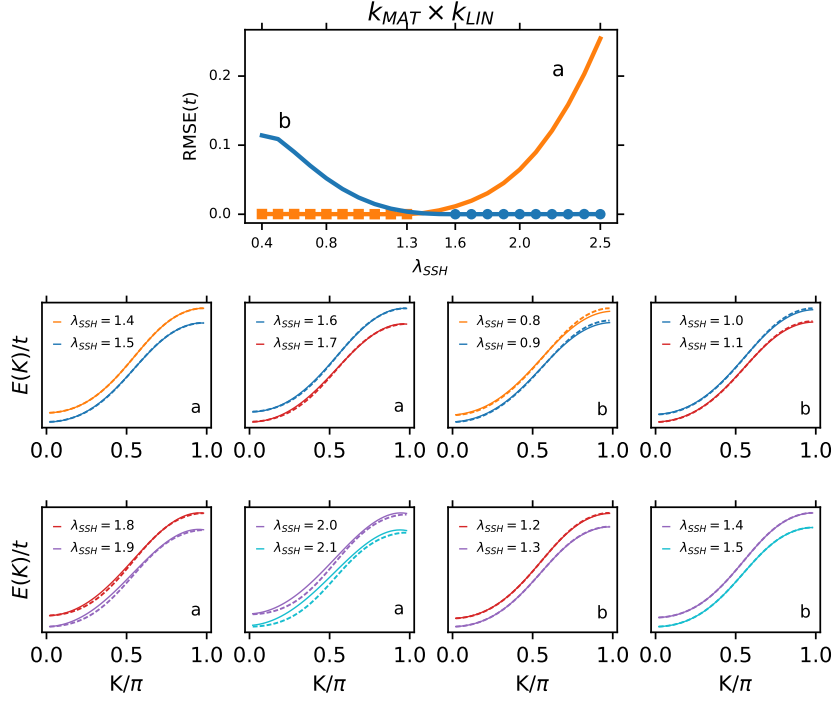


Figure 4.2: The upper panel is the RMSE of GPs' prediction as a function of λ_{SSH} . The markers illustrate the values of λ_{SSH} used to train the GPs. We consider two sets of training values, (a) $0.4 \leq \lambda_{SSH} \leq 1.3$ and (b) $1.6 \leq \lambda_{SSH} \leq 2.5$. The lower panels illustrate the GPs prediction (solid colour curves) and the accurate energy dispersion (dashed colour curves). We use $k_{MAT} \times k_{LIN}$ as the kernel function for both GP models. $\omega/t = 3$ is held fixed.

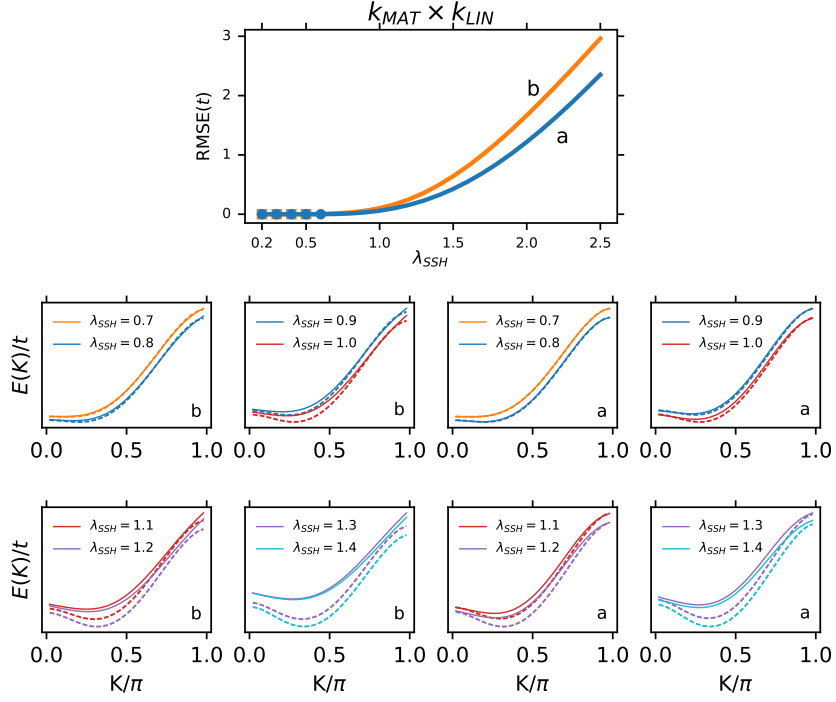


Figure 4.3: The upper panel is the RMSE of GPs' prediction as a function of λ_{SSH} . The markers illustrate the values of λ_{SSH} used to train the GPs. We consider two sets of training data (a) 100 points distributed from $0.2 \leq \lambda_{SSH} \leq 0.5$ and (b) 125 points distributed in $0.2 \leq \lambda_{SSH} \leq 0.6$. The lower panels illustrates the GPs prediction (solid colour curves) and the accurate energy dispersion (dashed colour curves). We use $k_{MAT} \times k_{LIN}$ as the kernel function for both GP models. For all the energy dispersions display here, we consider $\omega/t = 3$ and $\beta = 0$.

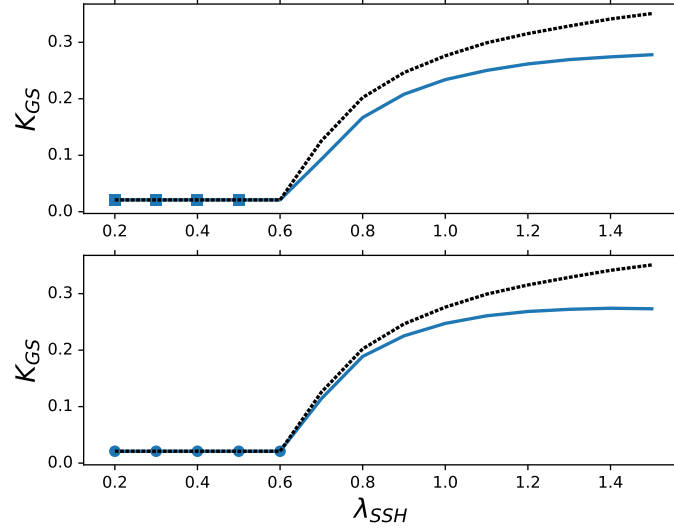


Figure 4.4: For both panels we illustrate the change in K_{GS} as a function of λ_{SSH} . The blue curves is the K_{GS} from the GP's prediction of the energy dispersion. The black dashed curve is the correct K_{GS} for each λ_{SSH} . The markers are the values of λ_{SSH} that we use to train each GPs. The upper panel shows $0.2 \leq \lambda_{SSH} \leq 0.5$, while the lower panel shows $0.2 \leq \lambda_{SSH} \leq 0.6$. We use $k_{MAT} \times k_{LIN}$ as the kernel function for both GPs. For all the calculations $\omega/t = 3$ and $\beta = 0$.

4.2.2 Kernel combination to predict multiple phase transitions

The complete phase diagram of the polaron model has three different phases [87], as is sketched in Figure 4.1. Each phase represents the value of K_{GS} as function of the Hamiltonian parameters,

- Phase I $\rightarrow K_{GS} = \pi$
- Phase II $\rightarrow 0 < K_{GS} < \pi$
- Phase III $\rightarrow K_{GS} = 0$

In this section, we exploit the power of ML to learn the complete phase diagram of the polaron model when few values of α and β are used as training data.

The first problem we consider is to determine the phase transitions by training a GP model with a few data points from phase I and III. Figure 4.5 illustrates the K_{GS} of the predicted $E(\cdot)$ with a GP model trained with different simple kernels. As it can be observed, the only kernel that is capable to predict the existence of phase II is the *Matern* kernel, however the transition lines are only predicted quantitatively close the the training data regime, Figure 4.5. The interpolation of the energy dispersions inside phase II is challenging due to the difference in the physical properties between the other two known phases.

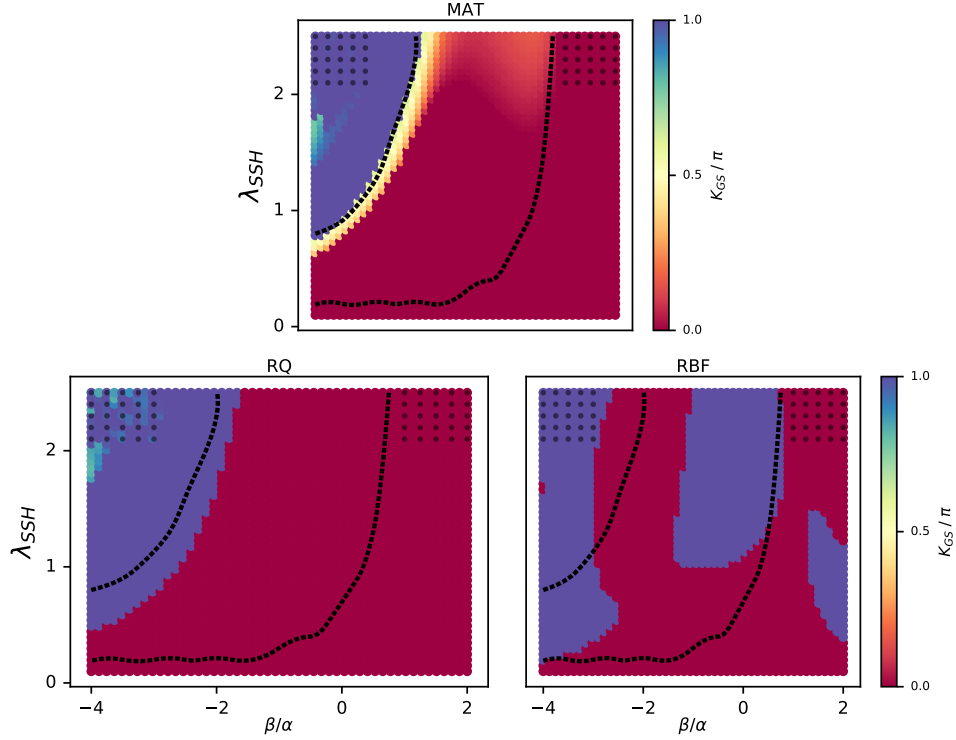


Figure 4.5: K_{GS} for the mixed model 4.3. The black dashed curves are the calculations from Ref. [87]. The color map is the prediction of the GP models with the fully optimized kernels. The models are trained by the dispersion of the polarons at the locations represented by the black dots. The different kernels were considered, k_{MAT} , k_{RQ} and k_{RBF} .

As we showed in the previous section, Section 4.2.1, to increase the learning capacity of a GP model we can use any possible combination of simple kernels

using Equations 4.6 and 4.7. The kernel parameters of any complex kernel were optimized by maximizing the log marginal likelihood. Here we evaluated all the possible combinations of two simple kernels. Each kernel combination proposed was trained with the same data points, 1250 points distributed from $2.0 < \lambda_{SSH} \leq 2.5$ for phases I and III, and $-4 \leq \beta/\alpha \leq -3$ (phase I) and $1.0 \leq \beta/\alpha \leq 2.0$ (phase III). More over, for each value of α and β we used 25 points from $0 < K < \pi$. The phase diagram predicted with each of all the possible combination of two kernels is shown in Figures 4.6 and 4.7. Figure 4.6 displays the predicted phase diagram when two kernels are combined by the addition operation, Equation 4.6. The predicted phase diagrams in Figure 4.7 were computed by multiplying two kernels, Equation 4.7.

As discussed in Chapter 2, ML models with a large number of parameters are not always more capable of learning. This can be observed in Figures 4.6 and 4.7, where increasing the number of kernels does not make the GPs regression prediction of $E(\cdot)$ more accurate. For example, a GP with $k_{RQ} + k_{RQ}$ is still not capable of predicting phase II where $0 < K_{GS} < \pi$. On the other hand, if the right combination of two kernels is chosen, e.g. $k_{MAT} + k_{RBF}$, $k_{RBF} + k_{RBF}$, $k_{MAT} \times k_{RBF}$ or $k_{MAT} \times k_{LIN}$, GP predicted phase diagram improves significantly,

With current computational power training GPs models with a more complex combination of kernels is doable; however, there are two major issues. The first problem is the number of kernel combinations that are required to fully describe a system, and the second obstacle is the need for test data to evaluate the accuracy of each kernel combination. In the following section, we address the lack of test data to construct more a robust kernel by reformulating the kernel combination problem into a Reinforcement learning (RL) problem where each taken action is towards constructing the ‘best’ kernel.

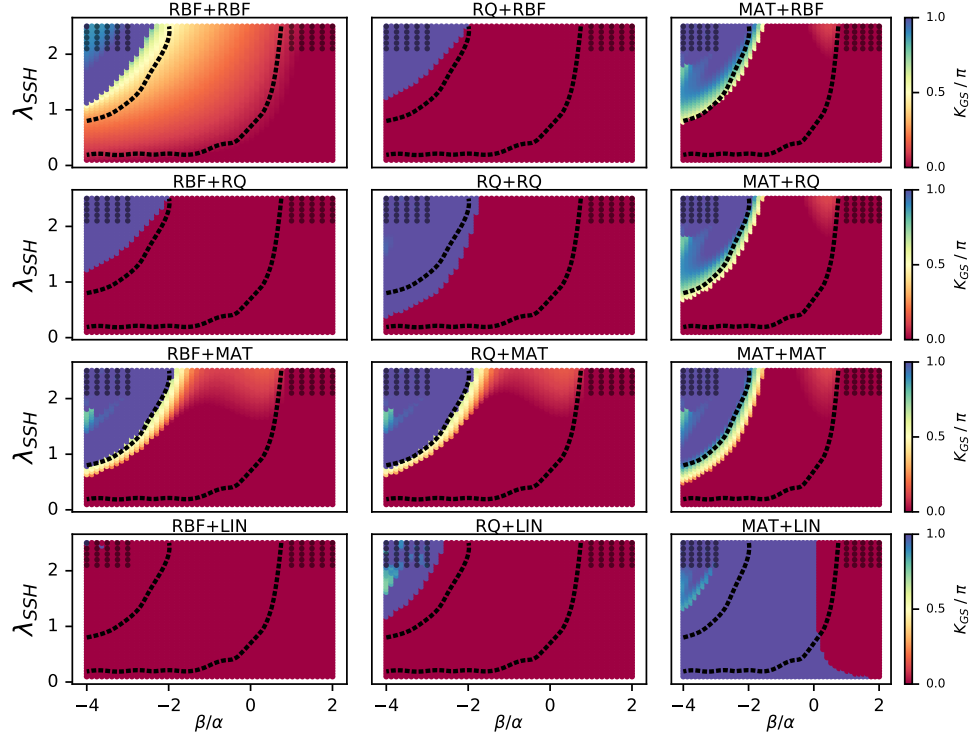


Figure 4.6: K_{GS} for the mixed model 4.3. The black dashed curves are the calculations from Ref. [87]. The color map is the prediction of the GP models with the fully optimized kernels. The models are trained by the dispersion of the polarons at the locations represented by the black dots. The different kernels considered here are all the possible pairwise addition, Equation 4.6, of two simple kernels, k_{MAT} , k_{RQ} and k_{RBF} .

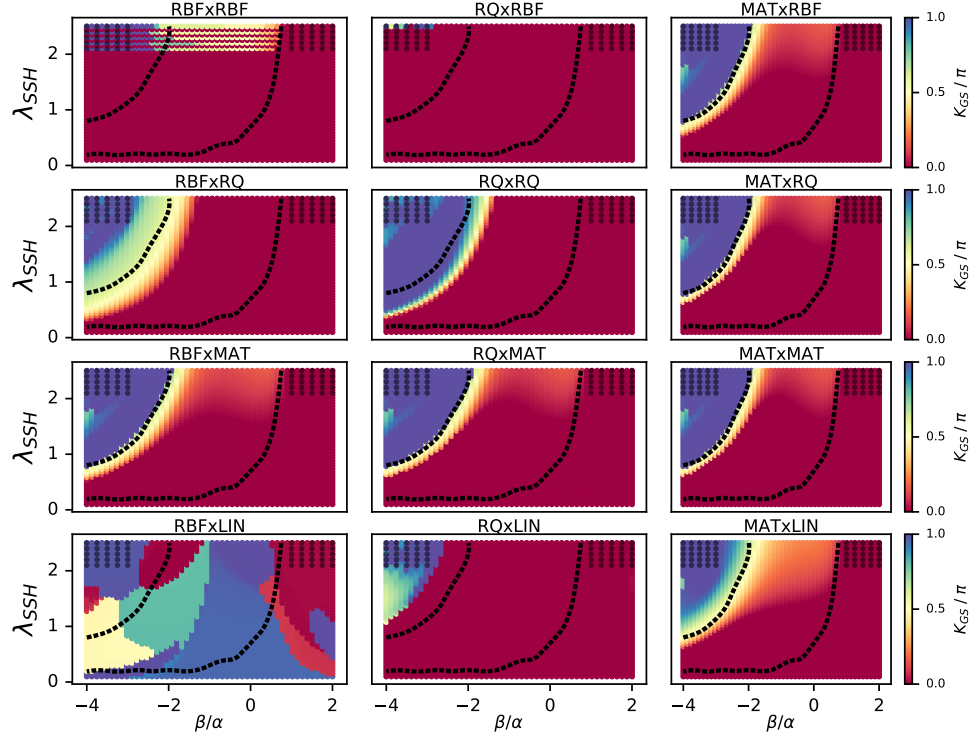


Figure 4.7: The momentum of the polaron ground state for the mixed model 4.3. The black dashed curves are the calculations from Ref. [87]. The color map is the prediction of the GP models with the fully optimized kernels. The models are trained by the dispersion of the polarons at the locations represented by the black dots. The different kernels considered here are all the possible pairwise amultiplication, Equation 4.7, of two simple kernels, k_{MAT} , k_{RQ} and k_{RBF} .

4.3 Model selection

As we stated in the previous section, we could define a GP with a complex kernel structure by combining simple kernels. However, given two GP models with totally different kernels, how do we select the ‘best’ kernel form?. Under the supervised learning framework, once can simply evaluate the test error for each GP model and choose the one with the lowest test error. In order to the compute test error, extra set of data is required, raising the question: *How can the most accurate combination*

of kernels be elected with the lack of test points?. This section presents that taking the advantage that GPs are probabilistic methods, it is possible to select the most accurate model without test points.

Under the framework of Bayesian statistics, it is possible to compute how probable is the data given a model \mathcal{M}_i , $p(\mathcal{D}|\mathcal{M}_i)$. In the case of GPs, $p(\mathcal{D}|\mathcal{M}_i)$ is used to optimize the kernel parameters as we explained in Section 2.3 and can be evaluated analytically. For illustrative purposes let us defined a first GP with $k_{MAT} + k_{RQ}$, and a second GP with $k_{MAT} + k_{MAT}$. For each model, we compute the *marginal likelihood* $p(\mathbf{y}|X, \theta, \mathcal{M}_i)$, where \mathcal{M}_i is one of the two proposed kernels and θ are the kernel parameters. Once both GP models are trained we can select the GP model with the highest marginal likelihood, since it resembles the relationship between the data and the model. However, it may not be fair to compare marginal likelihoods for kernels with different number of parameters. We must remember that a model with a larger number of parameters tends to memorize the data easily which could lead to a higher marginal likelihood [20, 132]. It is possible to penalize the marginal likelihood to compare models with different number of parameters more fairly using the Bayesian information criterion (BIC) [58, 132, 168].

4.3.1 Bayesian information criterion

The training stage of a GP with a fixed kernel form is also known in the Bayesian statistics framework as model selection, where each different kernels parameters' value is a unique model. The most appropriate model is the one with the highest marginal likelihood. In the case of GPs, the parameters of the kernel are continuous variables and the optimization can be carried out numerically as shown in Section 2.3. Unfortunately, we can not use the exact same procedure to search for the 'best' kernel combination since the combination of kernels is not a continuous variable that can be numerically optimized. To overcome this problem, we select different kernel combinations using the Bayesian information criterion [58, 132, 168],

$$\text{BIC}(\mathcal{M}_i) = \log p(\mathbf{y}|X, \hat{\theta}, \mathcal{M}_i) - \frac{1}{2} |\mathcal{M}_i| \log N \quad (4.10)$$

where $|\mathcal{M}_i|$ is the number of parameters of kernel \mathcal{M}_i , and N is the number of training points. The term $-\frac{1}{2}|\mathcal{M}_i|\log N$ is a penalizing term for kernels with a larger number of parameters. The $\log p(\mathbf{y}|X, \theta^*, \mathcal{M}_i)$ is the logarithm of the marginal likelihood for an optimized kernel. Using the BIC we can compare different kernels without the need of test data, however, there can be more than one good combination of simple kernels that could accurately represent the structure of the training data. With the computational power available today, it may seem possible to create and train GP models with as many combinations of kernels as can be imagined. Unfortunately, the combination of kernels is a non-tractable combinatorial problem and this makes the problem almost impossible.

4.3.2 Greedy search for model construction

To overcome the combinatorial problem of possible kernel combinations, we use a *greedy* algorithm to narrow the search space for the most suitable combination of kernels [58, 60]. As we mentioned in Chapter 3, *greedy* strategies are well known in the RL literature [181]. We use BIC as our selection criteria to find the ‘best’ combination of kernels.

The first step in searching for a more robust kernel is to train different GP models with different simple kernels. We denoted simple kernels as all the kernel listed in Section 2.4. We select the kernel with the highest BIC, and denoted as $k_0(\cdot, \cdot)$. In the following step, we use $k_0(\cdot, \cdot)$ as a base kernel and we combine it with all $k_i(\cdot, \cdot)$. We combine the two kernels by multiplying them, $k_0(\cdot, \cdot) \times k_i(\cdot, \cdot)$, or adding them, $k_0(\cdot, \cdot) + k_i(\cdot, \cdot)$. For all the possible combinations we compute the BIC and select the kernel with the highest BIC which is denoted as $k_1(\cdot, \cdot)$. We carried the same procedure iteratively where the base kernel is $k_\ell(\cdot, \cdot)$, and $k_{\ell+1}(\cdot, \cdot)$ is the kernel combination that has the highest BIC. Figure 4.8 illustrates the greedy search algorithm explained above.

4.4 Results

In this section, we illustrate the use of GP models with complex kernels that are constructed by an iterative maximization of the BIC explained in the previous section. We illustrate the possibility to extrapolate quantum observables to regimes

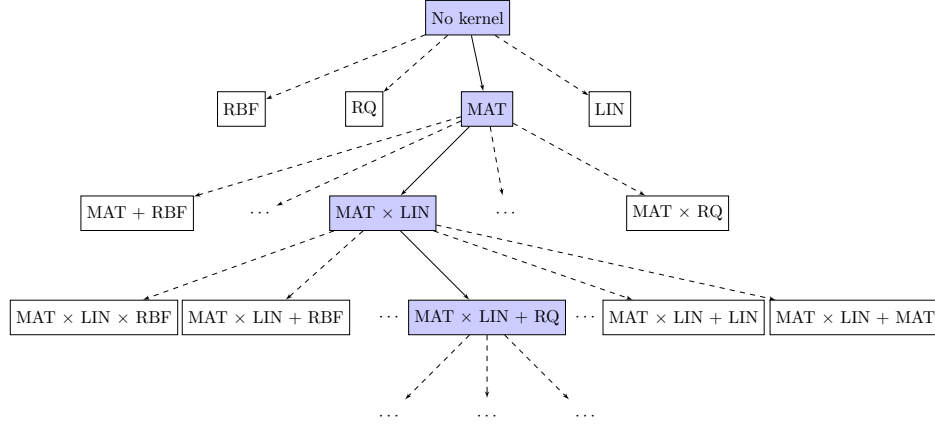


Figure 4.8: Schematic diagram of the kernel construction method employed to develop a GP model with extrapolation power. At each iteration, the kernel with the highest Bayesian information criterion (Equation 4.10) is selected.

in the spaces where traditional numerical algorithms have a lack of convergence. Furthermore, we also trained a GP model with data only from phases I and III, as explained in Section 4.2.2 and predicted the complete phase diagram of the polaron model.

4.4.1 Extrapolating quantum observables to unconverged regimes

This section illustrates how kernel combinations can predict quantum observables where standard numerical methods fail to converge due to the Hilbert space dimensionality. We consider the Holstein polaron model, where the V_{e-ph} term in Equation 4.3 is,

$$H_2 = \frac{g}{\sqrt{N}} \sum_{k,q} c_{k+q}^\dagger c_k (b_{-q}^\dagger + b_q). \quad (4.11)$$

Many methods have been developed to study dressed particles like the Holstein polaron, e.g. quantum Monte Carlo based methods, variational exact diagonalization [22], momentum average approximation [16], to mention few. All these methods demand higher computational resources when the value of ω decreases

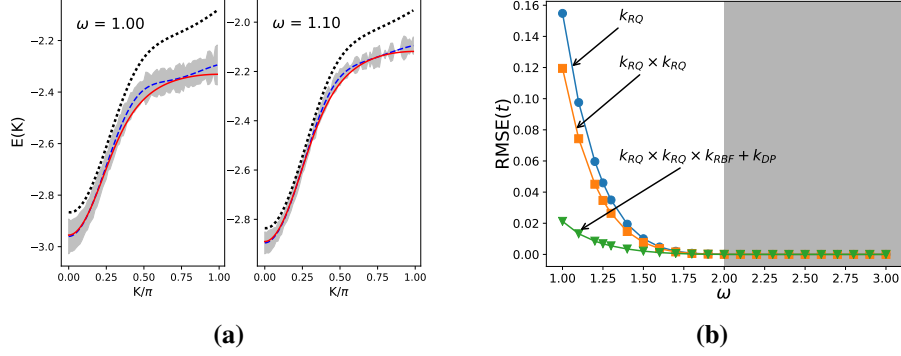


Figure 4.9: Left panel: Holstein polaron dispersion predicted with a GP model with k_{RQ} , black dashed curves, and $k_{RQ} \times k_{RQ} \times k_{RBF} + k_{LIN}$, blue dashes curves. The red solid curves are polaron dispersions computed with the momentum average approximation method [22]. Right panel: The change in the RMSE as a function of depth in the kernel combination search guided by the BIC. We plot the RMSE between GP predicted polaron dispersion and the exact polaron dispersion at different values of ω . The grey area is the range in ω considered for training the GP. For both figures, the value of $g = 1.5$ is fixed.

since the increase of the Hilbert space dimension. As a proof-of-principle, we present that GP models with kernel combination can also extrapolate $E(\cdot)$ in the regime where ω is small. The phonon frequency range considered for training was, $2.0 < \omega < 3.0$. We used the BIC to search for the most appropriate kernel combination to extrapolate $E(\cdot)$ in the low phonon frequency regime, Figure 4.8. Figure 4.9b illustrates that at each iteration of the BIC search, the learning of a GP model has a better representation of $E(\cdot)$ for $\omega < 2.0$. We must point out that the single kernel with the highest BIC, k_{RQ} , is capable to extrapolate $E(\cdot)$ qualitatively. On the other hand, a GP model with $k_{RQ} \times k_{RQ} \times k_{RBF} + k_{LIN}$ yields to accurate results even when $\omega \approx 1.0$, Figure 4.9a.

4.4.2 SSH-BM polaron phase diagram

In this section, we present predicted phase diagram of the SSH-BM polaron with a GP model where the kernel is constructed by the BIC, as it is depicted in Figure 4.8,

For clarity, here, we use the notation “GPL- X ” for the kernel with the highest BIC obtained after X depth levels of the algorithm depicted in Figure 4.8 of the main manuscript. We considered two different sets of training data, the first one where the training data is only from phase III, same as in Figures 4.6 and 4.7, and the second from phases I and II. In the first case, Figure 4.10 illustrates the change in the prediction of the polaron phase diagram depth levels of the algorithm depicted in Figure 4.8. As it can be noticed the first kernel GPL-0, k_{MAT} , does predict the change in K_{GS} from zero to π qualitatively. At the third level GPL-3 is $(k_{MAT} + k_{RBF}) \times k_{LIN}$ where the phase diagram is predicted quantitatively. We state that any GPL- X predicts $E(K, \alpha, \beta)$ in the complete range of the Hamiltonian parameters, then we search for K_{GS} which we used to construct the phase diagram.

Here, we also test if GP models are capable to predict quantum observables in-between phases. Predicting quantum observables in between different quantum phases is also a challenging problem. The training data used for this example is sampled from phases I and III. Again, the kernel form, GPL- X , is proposed by the algorithm described in Figure 4.8. GPL-4, in Figure 4.11, is $(k_{MAT} \times k_{LIN} + k_{RBF}) \times k_{LIN}$ and accurately predicts the entire SSH-BM polaron phase diagram.

GPs are a non-parametric method, meaning that the data is one of the key components to train GPs. Here, we illustrate that even with kernel combinations, different data lead to a different combination of kernels. With only considering data from phase III, as in Figure 4.10, and used Section 4.3.2 algorithm to predict the polaron phase diagram. We consider two different training sets, one where the training points are closer to the transition line and one where the points are further away. Each set has the same amount of points, 900, but different values of α and β . For each pair of values of α and β we consider 20 points distributed $0 < K < \pi$. The predicted phase diagram with a GP trained with both set of points is illustrated in Figure 4.12.

For both training sets, the first selected kernel is k_{MAT} , Figure 4.12. A GP model with a single kernel trained with data closer to the phase transition is capable to qualitatively predict the existence of phase II where $K_{GS} > 0$. Additionally, neither of both kernels can predict the existence of phase III, $K_{GS} = 0$. The right column panels of Figure 4.12 are the predicted phase diagrams with GPL-2 for both training sets. $k_{MAT} + k_{RBF} + k_{LIN}$ is the GPL-2 for where the training data is

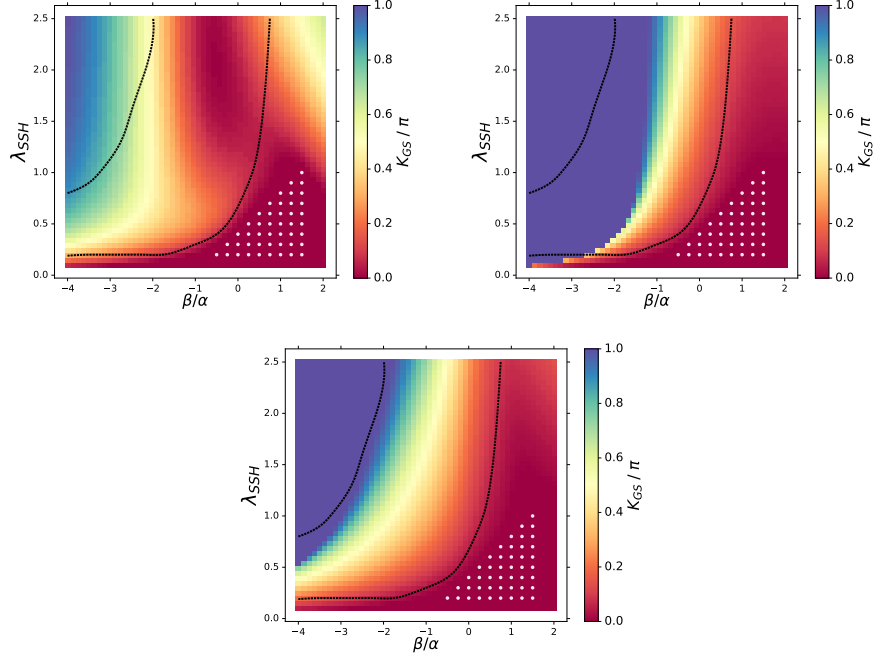


Figure 4.10: K_{GS} for the mixed model (4.3) as a function of β/α for $\lambda_{SSH} = 2\alpha^2/t\hbar\omega$. The dotted curves are the quantum calculations from Ref. [87]. The color map is the prediction of the GP models. Each panel illustrates the improvement of the predicted phase diagram. The panels correspond to the optimized kernels GPL-0 (left), GPL-1 (right), GPL-2 (centre), where “GPL- X ” denotes the optimal kernel obtained after X depth levels in the algorithm depicted in Figure 4.8. The models are trained by the dispersion of the polarons at the locations represented by the white dots.

the furthest from the first phase transitions, while GPL-2 is $(k_{MAT} + k_{RBF})k_{LIN}$ for the data closer to the transition. Both kernels are capable to predict the existence of phase I where $K_{GS} = \pi$. The left column of Figure 4.12 shows that the two sets of training data contain different physical information.

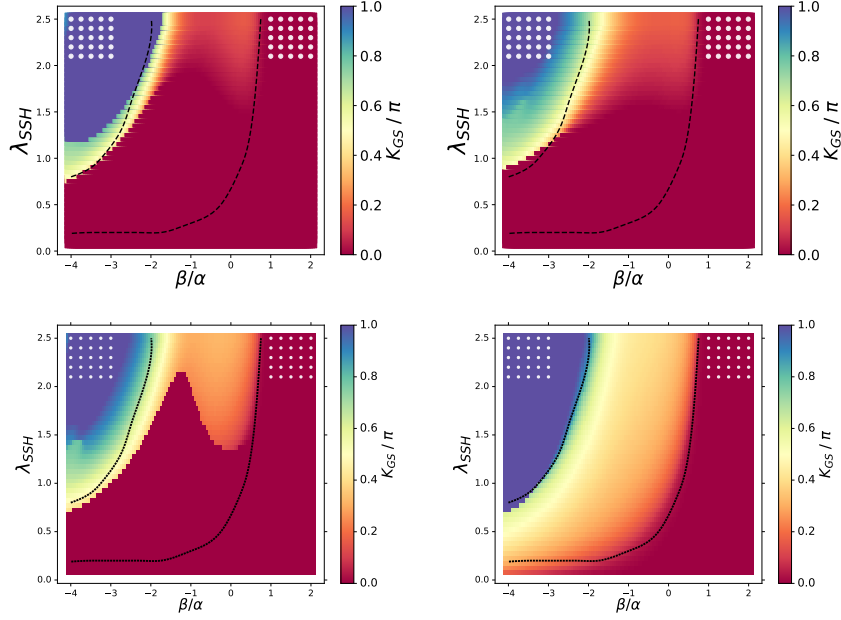


Figure 4.11: K_{GS} for the mixed model (4.3) as a function of β/α for $\lambda_{SSH} = 2\alpha^2/t\hbar\omega$. The dotted curves are the quantum calculations from Ref. [87]. The color map is the prediction of the GP models. Each panel illustrates the improvement of the predicted phase diagram. The panels correspond to the optimized kernels GPL-0 (left), GPL-1 (right), GPL-2 (centre), where “GPL- X ” denotes the optimal kernel obtained after X depth levels in the algorithm depicted in Figure 4.8. The models are trained by the dispersion of the polarons at the locations represented by the white dots.

4.5 Discussion

Predicting phase transitions using ML learning algorithms is currently a ‘hot’ research topic in the field of many-body physics. One of the most known works was proposed the use of NNs with multiple layers to learn how to classify different spin configurations in order to characterize the ferromagnetic or antiferromagnetic phase by J. Carrasquilla *et al.* in Ref [35]. Another influential work was proposed in 2015 by L.-F. Arsenault *et al.* where it is showed that ML algorithms can also classified different phases of matter and interpolate quantum observables such as

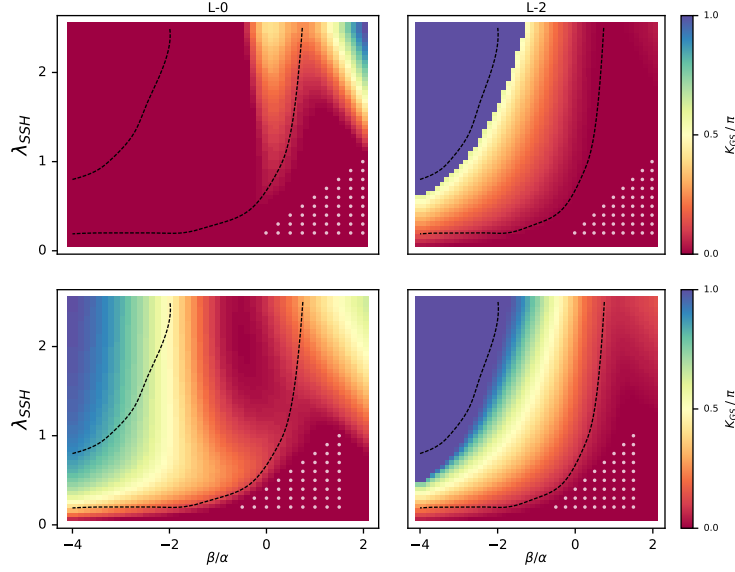


Figure 4.12: K_{GS} for the mixed model (4.3) as a function of β/α and $\lambda_{SSH} = 2\alpha^2/t\hbar\omega$. The black dashed curves are the calculations from Ref. [87]. The color map is the prediction of the GP models with the fully optimized kernels. The models are trained by the dispersion of the polarons at two different locations represented by the white dots. Left column panels are the predicted phase diagram with GPL-0 for both training data sets and right column panels are with GPL-2.

the quasi particle weight in the Hubbard model [8]. The interpolation of the quantum observables was done using KRR. Identifying different phases of matter in the Ising model can also be done with unsupervised learning as L. Wang showed in Ref. [188].

The approach proposed over this chapter illustrates how GP models are capable of robust interpolation and extrapolation of quantum observables to construct phase diagrams. We also state that the learning capacity of GP models can be enhanced using multiple kernels. We highlight the impact that different combination of kernels have in the prediction of GP models. It is important to emphasize that combining a large number of kernels does not ensure a more accurate model, as many of the kernels predicted incorrect phase diagrams in Figures 4.5 to 4.7

Over the course of this chapter, we illustrate that a hand-crafted kernel com-

bination, like $k_{MAT} \times k_{LIN}$, can predict quantum observables beyond training data under some regimes. This kernel form is robust enough to even predict the change of form in the energy dispersion of the SSH polaron model as a function of λ_{SSH} or α . The prediction of GP models intrinsically depend on the number of training points since they are non-parametric methods; Figures 4.2, 4.3 and 4.4 display the increase in the accuracy in the GPs extrapolation when more training data is used. The prediction of $E(\cdot)$ in the low phonon frequency regime is computationally demanding as a result of the Hilbert space dimensionality required. Section 4.4.1 presents that quantum observables can be accurately learned where traditional numeric methods have a lack of convergence.

Constructing the most appropriate combination of kernels to study different systems is a complex problem on its own, and a lack of test data increases the difficulty of the problem. We examine the use of BIC to reduce the combinatorial space of kernels; specifically, we used *greedy* selections. However, different strategies can be used to construct the most accurate combination of kernels. For example, using auto-econders we can reformulate the search of the ‘best’ kernel to a continuous space as R. Gómez-Bombarelli *et al.* did for molecules in Ref. [71]. The advantage of searching for the most accurate kernel combination in a continuous space is the possibility to use optimization algorithms as BO discussed in Chapter 3.

The BIC is the selecting criteria used to pick the most appropriate kernel combination to overcome the lack of test data for comparing the robustness of each model. Further work should be carried out to study the impact of different metrics, like the Akaike information criterion [132], for constructing different kernel combinations,

$$AIC(\mathcal{M}_i) = \log p(\mathbf{y}|X, \hat{\theta}, \mathcal{M}_i) - |\mathcal{M}_i|, \quad (4.12)$$

where the first term is the logarithm of the marginal likelihood for an optimized kernel and $|\mathcal{M}_i|$ is the number of kernel parameters. It is important to mention that there can be more than one single kernel that can accurately predict phase diagram, as it is shown in Figures 4.6 and 4.7.

As we discussed in Chapter 2, when non-optimal kernel parameters are used for prediction, GP models have a lack of accuracy. The algorithm introduced in Section 4.3.1 sequentially constructs more robust kernels using the BIC. The BIC depends on the optimal values of θ , meaning that it is possible to select a wrong kernel if the kernel optimization is not converged. GP models with extremely complex kernels cannot be used since the optimization of such kernels is challenging and could lead to inaccurate predictions.

The prediction of phase diagrams is one of the key problems in condensed-matter physics, however, we can use ML to push the boundaries of a physical problem and try to infer new phases of a diagram without knowing its existence. With the rise of more complex ML algorithms like deep-NN [72], bayesian-NN [134], GPs of GPs (deep-GPs) [53] or even with different kernel functions [191, 192], it may be possible to discover new quantum systems or study the quantum world in new regimes.

Chapter 5

Accelerating quantum walks

*If you are receptive and humble,
mathematics will lead you by the hand.*

— Paul A. M. Dirac

This chapter presents a study for the spreading of a quantum particle placed in a single site of a lattice or binary tree with the Hamiltonian permitting particle number changes. We show that particle number-changing interactions can accelerate the spreading beyond the ballistic expansion limit by inducing off-resonant Rabi oscillations between states of different numbers of particles. We consider the effect of perturbative number-changing couplings on Anderson localization in one-dimensional disordered lattices and show that they lead to a decrease of localization. The effect of these couplings is shown to be larger at larger disorder strength, which is a consequence of the disorder-induced broadening of the particle dispersion bands. Results presented in this chapter are published in Ref. [185]

5.1 Introduction

An important class of quantum computing algorithms is based on quantum walks (QWs) [89, 187], the quantum analog of random walks [46]. Random walks on lattices and graphs are powerful mathematical objects that can be used as algorithmic tools for a variety of problems, including optimization, search and classification. The efficiency of many such algorithms is determined by ‘hitting’ and ‘mixing’

times, quantifying how long it takes random walks to explore the underlying graphs [4]. Depending on the Hamiltonian, quantum walks can be accelerated by dynamical interferences and have potential to offer polynomial or, for some problems, exponential, computation speedup [63]. The role of interferences in QWs is perhaps best exemplified by the ballistic expansion of a quantum particle with time T in a periodic lattice [45, 47, 89, 171], leading to the $\propto T$ growth of the probability distribution, compared to the $\propto \sqrt{T}$ expansion of the classical random walk. QWs have been proven to offer the \sqrt{N} speed-up of spatial search over N items arranged in a d -dimensional lattice, with $d > 4$ [44].

With recent advances in the experiments on controlling atoms [19, 57, 66, 103, 143, 184], molecules [32, 111, 176], photons [9] and arrays of superconducting qubits [17, 31, 80, 98, 136], it has become possible to engineer lattice Hamiltonians. This, combined with the importance of the speed of QWs for the quantum computing algorithms and for the study of the fundamental limits of the velocity of quantum correlation propagations [117, 149], raises the question if and how lattice or graph Hamiltonians can be engineered to accelerate quantum walks. The effect of Hamiltonian engineering on quantum walks has been studied in many different contexts. For example, Giraud *et al.* [70] showed that Anderson localization impeding quantum walks in disordered systems can be mitigated by adding hopping terms, which provide shortcuts in circular graphs. QWs can also be accelerated by coupling a Hamiltonian system to an external bath. While the general belief is that particle-environment interactions destroy the coherence of quantum walks leading to transport suppression in ordered systems, multiple recent studies showed that interactions with certain non-Markovian baths provide new pathways for interferences [130, 144, 148]. The range of particle hopping and particle interactions are also known to determine the speed of quantum information propagation [39, 54, 149, 153].

Engineering many-particle (as opposed to a single particle) quantum walks is becoming an important research goal [139]. As shown by Childs *et al.*, QWs on a sparse graph can be used to efficiently simulate any quantum circuit [43] and interacting quantum walks are capable of universal quantum computation [48]. Quantum walks of interacting pairs can be used to determine if graphs are isomorphic [67]. Particle correlations can be exploited to change the directionality of

quantum walks [28]. Quantum walks of interacting particles can be used to realize quantum Hash schemes [116]. Two-body or multi-particle correlations have been shown to affect quantum walks of few- and many-particle systems in interesting ways [18, 28, 38, 39, 68, 108, 109, 145, 154, 155, 157, 162, 179, 189]. These studies consider particle correlations arising either as a consequence of direct density-density interactions or particle quantum statistics.

This chapter presents an alternative mechanism for accelerating quantum walks, namely quantum walks in a dynamical system governed by a Hamiltonian allowing particle number changes. Such Hamiltonians can be engineered with quasiparticles, such as excitons [34, 84, 128, 201], or with ultracold atoms trapped in optical lattices and immersed in a condensate [40, 99]. They are also of significant experimental and theoretical interest due to the relation to the topologically protected states and their possible use in quantum computing [106]. The present work shows that the particle-number-changing interactions lead to Rabi oscillations, which significantly accelerate the spreading of quantum wave packets in ideal lattices and binary trees. Also considers the effect of such terms on Anderson localization and show that they lead to decrease of the inverse participation ratio in disordered systems. This work also shows that the effect of number-changing interactions on the participation ratio becomes stronger with increasing disorder strength.

5.2 Models

Consider the quantum dynamics governed by the following lattice Hamiltonian:

$$\hat{\mathcal{H}} = \sum_i \omega_i \hat{c}_i^\dagger \hat{c}_i + t \sum_{\langle i,j \rangle} \hat{c}_j^\dagger \hat{c}_i + v \sum_{\langle i,j \rangle} c_i^\dagger c_i c_j^\dagger c_j + \hat{V}_{\text{nc}}, \quad (5.1)$$

where

$$\hat{V}_{\text{nc}} = \Delta \sum_{\langle i,j \rangle} (\hat{c}_i^\dagger \hat{c}_j^\dagger + \hat{c}_i \hat{c}_j) + \gamma \sum_i (\hat{c}_i^\dagger + \hat{c}_i), \quad (5.2)$$

\hat{c}_i is the operator that removes the particle from site i , the quantities ω_i , t , Δ and v are the Hamiltonian parameters, and the angular brackets indicate that the hopping

and interactions are only permitted between nearest neighbour sites. The on-site energy ω_i is defined as $\omega_i = \Delta\epsilon + \epsilon_i$, where $\Delta\epsilon$ is a constant and ϵ_i is varied in the calculations for lattices with on-site disorder (more details below and in Chapter 6). The term \hat{V}_{nc} couples different particle-number states.

Model (5.1) is a special case of the full Hamiltonian for the Frenkel excitons in an ensemble of coupled two-level systems [2]. At $\Delta = 0$, $\gamma = 0$ and $\nu = 0$, this Hamiltonian reduces to the tight-binding model. At $\Delta = 0$ and $\gamma \neq 0$, the model describes the quantum annealer setup of D-wave [80], where currents in interacting superconducting qubits are mapped onto spin states.

This work considers the few-particle limit of Hamiltonian (5.1) and calculates the dynamics of quantum walks by diagonalizing the Hamiltonian and constructing the full time evolution operator from the complete set of the corresponding eigenvectors, as was done, for example, in Ref. [199]. In order to describe properly the dynamics governed by the models with $\Delta \neq 0$ and/or $\gamma \neq 0$, the Hilbert space must include multiple particle-number states. The Hilbert space is truncated and only includes one and three particles for the case $\Delta \neq 0, \gamma = 0$. When $\Delta = 0, \gamma \neq 0$, the Hilbert space includes the vacuum state (zero particles), one, two, and three particles. As discussed below, this chapter considers the Hamiltonian parameters, for which the multiple-particle states have high energy. Since the energy of such states increases with the number of particles and the couplings can only change the number of particles by one or two, the contribution of such states decreases with the number of particles. Previous calculations done, verified that for a lattice with 19 sites that including the states of five particles does not change the results for the Hamiltonian parameters considered here, Figure 5.2.

The on-site energy $\Delta\epsilon + \epsilon_i$ determines the energy separation between states with different numbers of particles. Throughout this work, we consider the limit $\Delta, \gamma \ll \Delta\epsilon$. For ideal lattices, $\epsilon_i = 0$. For disordered lattices, ϵ_i is drawn from a uniform distribution of random numbers. In this limit, the state corresponding to one particle at zero time becomes weakly dressed with higher particle-number states. The effect of the dressing can be accounted for by the Schrieffer-Wolf transformation [167], which in first order leads to the appearance of next-nearest-neighbour hopping terms, as shown in Appendix B. Including higher order terms resulting from the transformation induces longer-range hopping. The particle can effec-

tively hop by undergoing virtual transitions to higher particle-number states and back. Note that in models with $\gamma \neq 0$, the particle can also hop by virtual transitions to the vacuum state (the state of no particles) and back.

5.2.1 Ideal 1D lattices

The first considered case is the well-studied problem of ballistic spreading in an ideal one-dimensional (1D) lattice. At $\Delta = 0$ and $v = 0$, a particle placed in an individual lattice site expands as shown by the solid black line in Figure 5.1. This spreading is much faster than the expansion of the area covered by the classical random walk, illustrated in Figure 5.1 by the dotted curve. Figure 5.1 shows that the quantum dynamics of a single particle initially placed in a single lattice site is drastically different from both the random walk result and the ballistic spreading when governed by the model (5.1) with $\Delta \neq 0$. In particular, the width of the wave packet oscillates at short times, approaching the ballistic-expansion-like behaviour at long times. These calculations are performed for the 1D lattice with $N = 41$ lattice sites with open boundary conditions. As can be seen from Figure 5.1, the effect of the boundaries is not important until time reaches $\approx 11 t^{-1}$.

In order to understand the origin of the oscillations, we plot in the insets of Figure 5.1 the average number of particles $\langle n \rangle$ as a function of time. It can be seen that $\langle n \rangle$ oscillates with the same period as the wave packet size. Thus conclude that the oscillations observed in Figure 5.1 are due to off-resonant Rabi flopping between the state of one particle and the states of multiple particles induced by \hat{V}_{nc} . Figure 5.1 shows that these coherent oscillations accelerate quantum walks beyond the ballistic limit. Note that $\langle n \rangle$ in Figure 5.1 is an average of one and three particles. For $\Delta\epsilon/t = 20$, $\langle n \rangle < 1.2$, which illustrates that the three-particle subspace remains largely unpopulated at all times.

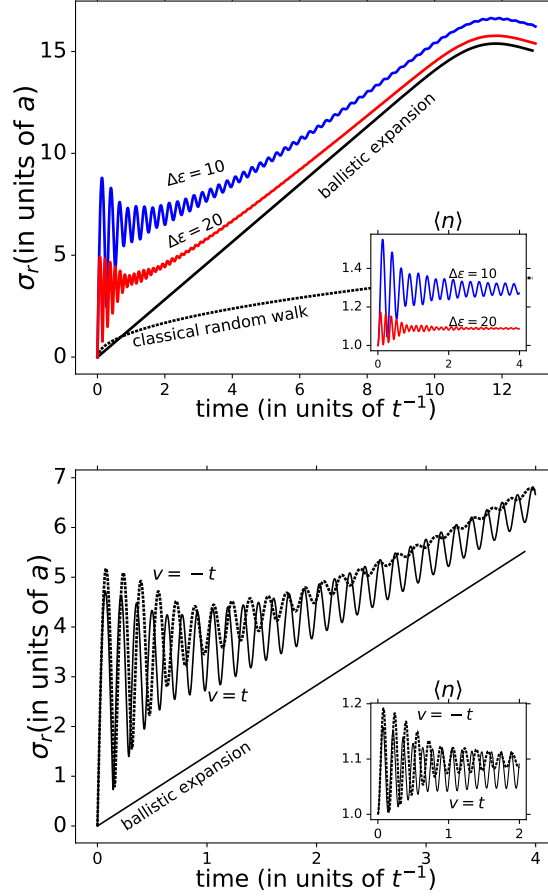


Figure 5.1: Time dependence of the standard deviation (in units of lattice constant) of the wave packet for a particle initially placed in a single site of a one-dimensional ideal lattice. The solid black curves represent the ballistic expansion governed by the Hamiltonian (5.1) with $v = 0$ and $\hat{V}_{nc} = 0$. Upper panel: The oscillating curves show the size of the wave packets governed by the Hamiltonian (5.1) with $v = 0$, $\gamma = 0$, $\Delta/t = 1$, $t = 1$ and two values of $\Delta\epsilon$: $\Delta\epsilon = 10/t$ (blue) and $\Delta\epsilon = 20/t$ (red). Lower panel: The oscillating curves show the size of the wave packets governed by the Hamiltonian (5.1) with $v = \pm 1$, $\Delta/t = 1$, $t = 1$ and $\Delta\epsilon = 20/t$. The insets show the average number of particles $\langle n \rangle$ as a function of time for the corresponding Hamiltonian parameters. Notice that for $\Delta\epsilon = 20/t$, $\langle n \rangle$ stays below 1.2 at all times. Figure from reference [185].

Since the $c_i^\dagger c_j^\dagger$ term generates pairs of particles in adjacent sites, it is important to consider the role of inter-site interactions v . Such interactions appear in extended Hubbard models, leading to non-trivial properties of the lattice systems [135, 164, 184] and inducing correlations in quantum walks [39]. Here, they are transient as the multiple-particle subspaces are populated only virtually. The inset of Figure 5.1 illustrates that repulsive interactions stabilize the oscillations at long times, while the short-time dynamics appears to be largely unaffected by the density-density interactions.

The $\Delta \neq 0$ term couples the subspaces with the odd number of particles. Thus, the state of a single particle is coupled to a state of three-particles, but not to the state of two particles or the vacuum state. By contrast, the $\gamma \neq 0$ term couples subspaces differing in the number of particles by one. To illustrate the effect of such couplings on the dynamics of quantum walks, we compare two models: (i) $\Delta = 0, \gamma = t$ and (ii) $\Delta = t, \gamma = 0$. The results shown in Figure 5.3 illustrate that the couplings in case (i) have a much stronger effect, leading to larger amplitudes of the oscillations and the persistence of the oscillations for much longer time.

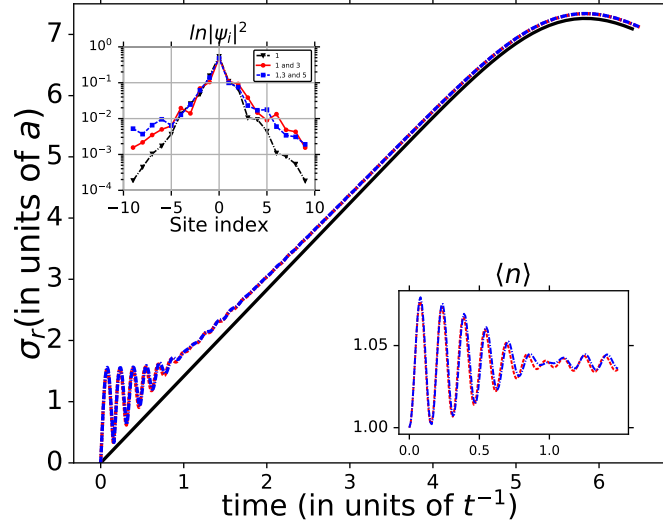


Figure 5.2: Time dependence of the standard deviation (in units of lattice constant) of the wave packet for a particle initially placed in a single site of a one-dimensional ideal lattice. The solid black curves represent the ballistic expansion governed by the Hamiltonian (5.1) with $v = 0$ and $\hat{V}_{nc} = 0$. The oscillating curves show the size of the wave packets governed by the Hamiltonian (5.1) with $v = 0$, $\gamma = 0$, $\Delta/t = 1$, $t = 1$, $\Delta\varepsilon = 20/t$ and a Hilbert space with different number of particles, 1 particle (black), 1 and 3 particles (red) and 1,3 and 5 (blue). The upper inset shows the logarithm of the particle probability distributions in a disordered 1D lattice with 19 sites. The results are averaged over 50 realizations of disorder and are time-independent with $w = 10/t$. And the lower inset depicts the average number of particles $\langle n \rangle$ as a function of time for the three different wave packets. Figure from reference [185].

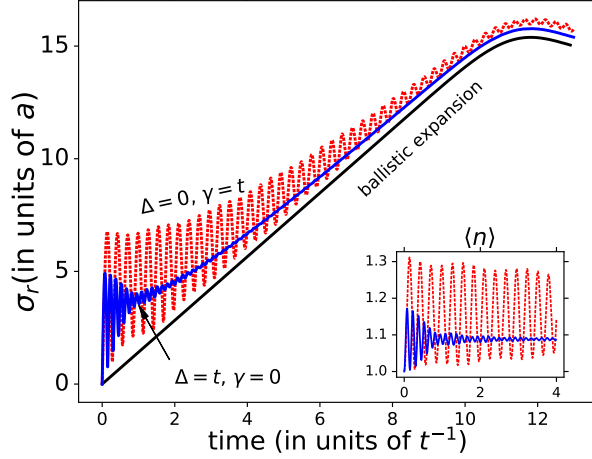


Figure 5.3: Time dependence of the standard deviation (in units of lattice constant) of the wave packet for a particle initially placed in a single site of a one-dimensional ideal lattice. The solid black line represents the ballistic expansion governed by the Hamiltonian (5.1) with $v = 0$, $\Delta = 0$ and $\gamma = 0$. For the dotted red curve $\Delta = 0$ and $\gamma = t$ while for the blue solid curve $\Delta = t$ and $\gamma = 0$. For all of these calculations, $\Delta\epsilon = 20/t$. The inset shows the average number of particles $\langle n \rangle$ as a function of time for $\Delta = 0$ and $\gamma = t$ (dotted red curve) and $\Delta = t$ and $\gamma = 0$ (blue solid curve). Figure from reference [185].

5.2.2 Disordered 1D lattices

We next consider disordered 1D lattices. The disorder is generated by randomizing the on-site energy ϵ_i by drawing the random values from a uniform distribution $[w/2, w/2]$, where w quantifies the strength of disorder. Non-interacting particles are exponentially localized in 1D disordered systems [7]. Our goal is to explore the role of the $\Delta \neq 0$ interactions on the localization.

In all of the disordered models we consider $\gamma = 0$, $\Delta/t \leq 1$ and $\Delta\epsilon/t = 20$. Notice that for the ideal lattice with $\Delta/t = 1$ illustrated in Figure 1, this value of $\Delta\epsilon$ ensures that the average number of particles $\langle n \rangle < 1.2$ at all times. The three-particle sub-space is thus far off-resonant and contributes to the dynamics perturbatively.

Upper panel of Figure 5.4 shows the average lattice population distributions illustrating the localization. To obtain these distributions, we place a particle in a single lattice site, propagate the wave packet to long time and average the resulting probability distribution over 100 random instances of disorder. We have verified that this number of disorder realizations ensures converged results. The averaging removes the time-dependence in the long-time limit. The results show that the term \hat{V}_{nc} induces non-exponential wings of the distribution, which rise with the magnitude of Δ . To illustrate the quantitative contribution of these wings, we compute the inverse participation ratio (IPR) defined as

$$I(t) = \sum_i \left(\frac{|\psi_i(t)|^2}{\sum_i |\psi_i(t)|^2} \right)^2, \quad (5.3)$$

where $|\psi_i(t)|^2$ is the probability of the population of lattice site i at time t . The value of the IPR ranges from $1/N$ for the state completely delocalized over the lattice with N sites to 1 for the state localized in a single lattice site. We find that the couplings with $\Delta/t = 1$ decrease the IPR, indicating decrease of localization. Surprisingly, the effect of these couplings increases with increasing disorder strength. This phenomenon is reminiscent of noise-induced delocalization [49, 130, 140, 148]. Here, the variation of on-site energy due to disorder brings the energy of the different particle-number states for random lattice sites closer together, thereby enhancing the effect of the couplings induced by \hat{V}_{nc} . With increasing disorder strength w , the probability of the different number states becoming closer in energy increases, leading to more and stronger high-order hopping terms, thereby decreasing localization more significantly. Note that this result applies only in the limit $\Delta \ll \Delta\epsilon$, i.e. in the limit where the number-changing interactions are much weaker than the energy separation between the number subspaces.

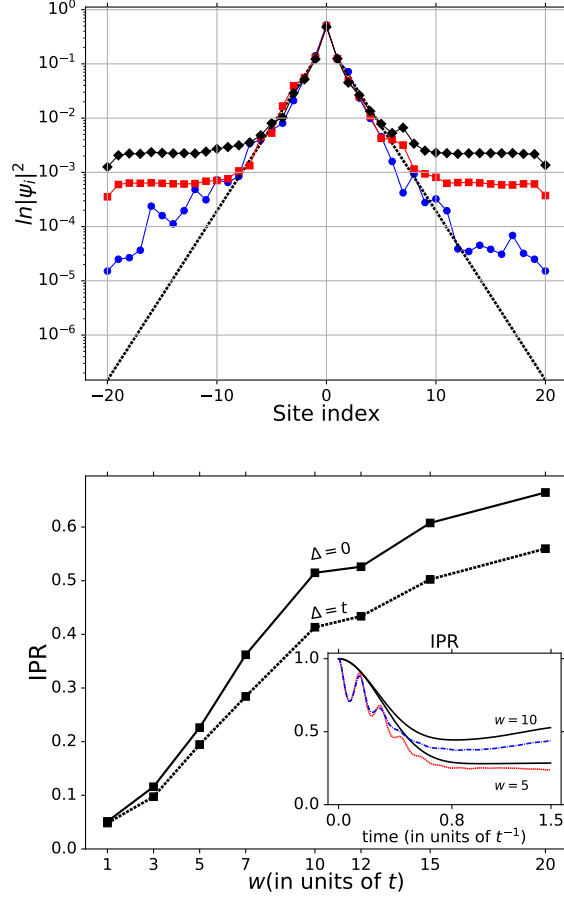


Figure 5.4: Upper panel: The logarithm of the particle probability distributions in a disordered 1D lattice with 41 sites: diamonds – $\Delta/t = 1$, squares – $\Delta/t = 1/2$, circles – $\Delta/t = 1/10$. The results are averaged over 100 realizations of disorder and are time-independent. The dashed line is an exponential fit to the $\Delta/t = 1/10$ results. Lower panel: the long-time limit of the IPR defined in Eq. (5.3) averaged over 100 instances of disorder as a function of the disorder strength w : solid line – $\Delta = 0$, dashed line – $\Delta/t = 1$. The inset shows the IPR averaged over 100 realizations of disorder for two disorder strengths $w = \{5/t, 10/t\}$ as functions of time: the solid black curves – $\Delta = 0$; the dotted and dot-dashed curves – $\Delta = t$. Figure from reference [185].

5.2.3 Binary trees

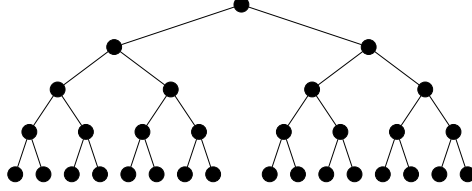


Figure 5.5: Schematic diagram of an ideal binary tree with depth-5 ($\mathcal{G}5$).

This section considers quantum walks in binary trees. A binary tree is characterized by the number of layers and the connectivity of the lattice sites. Here, we consider the binary tree $\mathcal{G}5$ with five layers schematically depicted in Figure 5.5. The model (5.1) adapted to binary trees becomes

$$\begin{aligned} \mathcal{H} = & \sum_{i=1}^{2^g-1} (\Delta\epsilon + \epsilon_i) \hat{c}_i^\dagger \hat{c}_i + \sum_{i=1}^{2^g-1} \sum_{j=1}^{2^g-1} t_{ij} \left(\hat{c}_i^\dagger \hat{c}_j + \hat{c}_j^\dagger \hat{c}_i \right) \\ & + \sum_{i=1}^{2^g-1} \sum_{j=1}^{2^g-1} \Delta \left(\hat{c}_i^\dagger \hat{c}_j^\dagger + \hat{c}_j \hat{c}_i \right) \end{aligned} \quad (5.4)$$

$\gamma=0$ is set for all binary tree calculations. Each node of the binary tree is connected to three nodes: its father, the left child $2i$ and the right child $2i+1$, so

$$\mathcal{H} = \sum_{i=1}^{2^g-1} (\Delta\epsilon + \epsilon_i) \hat{c}_i^\dagger \hat{c}_i + t \sum_{i=1}^{2^g-1} \left(\hat{c}_i^\dagger \hat{c}_{2i} + \hat{c}_i^\dagger \hat{c}_{2i+1} + h.c. \right) + \hat{V}_{\text{nc},\text{tree}} \quad (5.5)$$

where

$$\hat{V}_{\text{nc},\text{tree}} = \Delta \sum_{i=1}^{2^g-1} \left(\hat{c}_i^\dagger \hat{c}_{2i}^\dagger + \hat{c}_i^\dagger \hat{c}_{2i+1}^\dagger + \hat{c}_i \hat{c}_{2i} + \hat{c}_i \hat{c}_{2i+1} + h.c. \right) \quad (5.6)$$

The spread of the quantum wave packets in such trees can be described by,

$$\sigma(t) = \sqrt{\langle \mathbf{v}^2 \rangle - \langle \mathbf{v} \rangle^2} = \sqrt{\sum_{\mathbf{v}=1}^{2^g-1} \mathbf{v}^2 p_{\mathbf{v}}(t) - (\mathbf{v} p_{\mathbf{v}}(t))^2} \quad (5.7)$$

Quantum walks started at the root of a graph and comparison between the dynamics

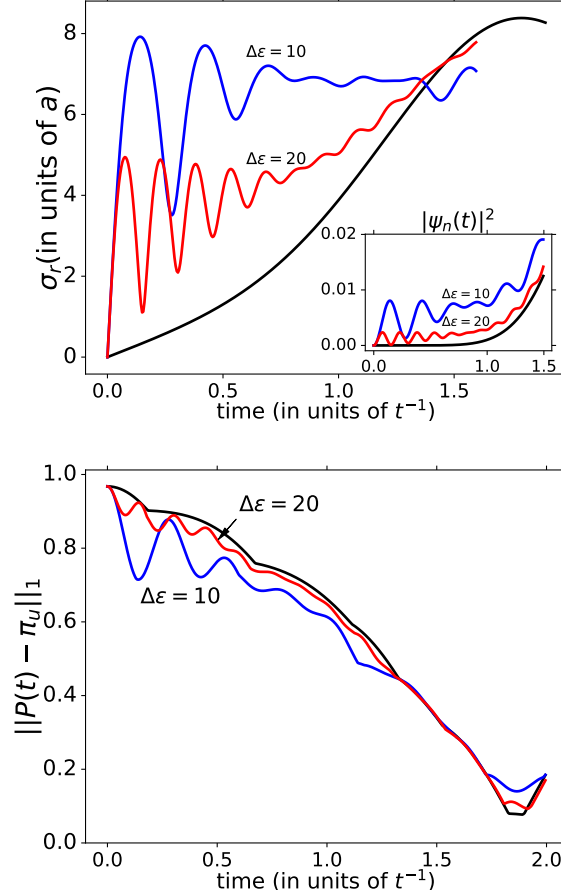


Figure 5.6: Upper panel: The growth of the wave packet for a single particle placed at the root of the tree (\mathcal{G}_5): black line – $\Delta = 0$; the oscillating curves – $\Delta/t = 1$ with $\Delta\epsilon = 10/t$ (blue) and $\Delta\epsilon = 20/t$ (red). The inset shows the probability of reaching the last node (in layer 5) of the binary tree as a function of time. The curves are color-coded in the same ways as in the main plot. Lower panel: The L_1 norm between the probability distribution of the wave packet and the uniform distribution as a function of time. The uniform distribution is defined by the value $1/\dim(\mathcal{G}_5)$ for each node. The solid black curve is for $\Delta = 0$. The oscillating curves are for $\Delta = t$ with $\Delta\epsilon = 10/t$ (blue) or $\Delta\epsilon = 20/t$ (red). For all curves in both figures $\gamma = 0$. Figure from reference [185].

of models with $\Delta = 0$ and $\Delta \neq 0$ is carried.

Upper panel of Figure 5.6 shows that the couplings $\hat{V}_{\text{nc,tree}}$ accelerate quantum walks on the tree. To quantify the effect of $\Delta \neq 0$ on quantum walks, we compute the mixing time defined as

$$M_\varepsilon = \min \{T | \forall t \geq T : \|P(t) - \pi\|_1 \leq \varepsilon\} \quad (5.8)$$

where $P(t)$ is the probability distribution at time t , π is a distribution that the quantum system is expected to approach, and $\|\cdot\|_1$ is the $L1$ norm.

Two distributions π are considered, the uniform distribution π_U and the stationary distribution π_S . The uniform distribution is characterized by the same value of probability for each node. The stationary distribution is defined by the following values of the probability for node v

$$\pi_S(v) = \lim_{T \rightarrow \infty} \bar{p}_v(T), \quad (5.9)$$

where $\bar{p}_v(T)$ is the time average of the probability of populating node v ,

$$\bar{p}_v(T) = \frac{1}{T} \sum_{t=0}^{T-1} p_v(t) = \frac{1}{T} \sum_{t=0}^{T-1} |\langle v | \Psi(t) \rangle|^2 \quad (5.10)$$

$$\begin{aligned} &= \frac{1}{T} \sum_{t=0}^{T-1} \langle v | \Psi(t) \rangle \langle \Psi(t) | v \rangle \\ &= \frac{1}{T} \sum_{t=0}^{T-1} \left\{ \sum_{\lambda} e^{-iE_{\lambda}t} \langle v | \lambda \rangle \langle \lambda | \Psi(0) \rangle \right\} \\ \bar{p}_v(T) &= \left\{ \sum_{\lambda'} e^{iE_{\lambda'}t} \langle \Psi(0) | \lambda' \rangle \langle \lambda' | v \rangle \right\} \end{aligned} \quad (5.11)$$

Here, $\hbar = 1$ is set. Defining $\langle \lambda | \Psi(0) \rangle$ as $c_{n_0}^\lambda$,

$$\begin{aligned}
\bar{p}_v(T) &= \frac{1}{T} \sum_{t=0}^{T-1} \left\{ \sum_{\lambda} e^{-iE_{\lambda}t} c_{n_0}^{\lambda} \langle v | \lambda \rangle \right\} \left\{ \sum_{\lambda'} e^{iE_{\lambda'}t} c_{n_0}^{\lambda'*} \langle \lambda' | v \rangle \right\} \\
&= \frac{1}{T} \sum_{t=0}^{T-1} \sum_{\lambda} |c_{n_0}^{\lambda}|^2 |\langle v | \lambda \rangle|^2 \\
&\quad + \frac{1}{T} \sum_{t=0}^{T-1} \sum_{\lambda, \lambda'} \left(c_{n_0}^{\lambda} c_{n_0}^{\lambda'*} e^{i(E_{\lambda'} - E_{\lambda})t} \langle v | \lambda \rangle \langle \lambda' | v \rangle \right). \tag{5.12}
\end{aligned}$$

In the limit of long time $T \rightarrow \infty$ the imaginary part of $\bar{p}_v(T)$ tends to zero. The factor $\frac{1}{T}$ in the real part of $\bar{p}_v(T)$ cancels because $\sum_{t=0}^{T-1} e^{i(E_{\lambda} - E_{\lambda})t} = T$. We can thus rewrite $\pi_S(v)$ as

$$\pi_S(v) = \sum_{\lambda} |c_{n_0}^{\lambda}|^2 |\langle v | \lambda \rangle|^2. \tag{5.13}$$

From Eq. (5.13) it can be observed that $\pi_S(v)$ depends on the initial condition ($|\Psi(t=0)\rangle$).

Lower panel of Figure 5.6 illustrates the effect of the couplings $\hat{V}_{\text{nc,tree}}$ on the speed of approaching the uniform distribution $\pi_u(v)$ and Figure 5.7 the effect of the couplings $\hat{V}_{\text{nc,tree}}$ on the stationary distribution $\pi_S(v)$. The approach to the uniform distribution is accelerated by the $\hat{V}_{\text{nc,tree}}$ terms at short times. As can be seen from Figure 5.6, the couplings $\hat{V}_{\text{nc,tree}}$ enhance the stationary distribution, illustrating that the graph is explored more efficiently by the dynamics with the $\hat{V}_{\text{nc,tree}}$ couplings.

5.2.4 Glued binary trees

If two binary trees of Figure 5.5 are joined together as shown in Figure 5.8, one obtains a glued binary tree. Transport through glued binary trees represents an important class of problems [96, 104]. Of particular interest is the probability of transfer from the head node to the bottom node in disordered glued trees. Studies of such processes have been used to understand the consequences of quantum localization for the application of quantum walks for quantum computing and quantum communication algorithms [96, 104].

To study quantum walks in a glued binary tree, model (5.4) was used but with

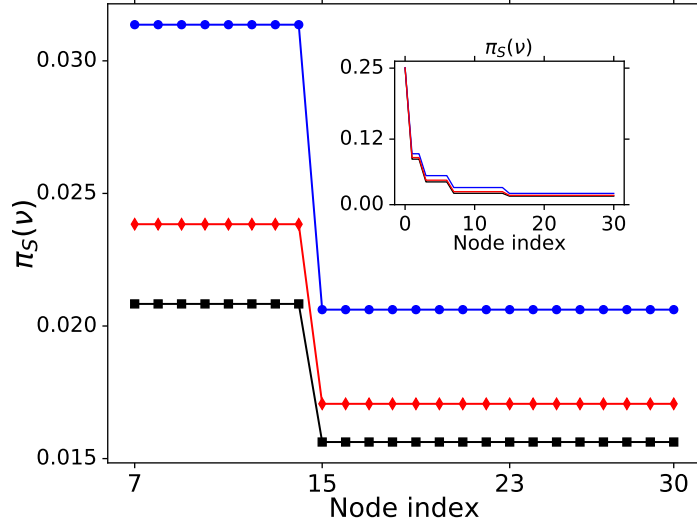


Figure 5.7: The squares represent the stationary distribution $\pi_S(v)$ for a QW on the $\mathcal{G}5$ tree with $\Delta = 0$. The circles and diamonds are the stationary distributions $\pi_S(v)$ for quantum walks with $\Delta/t = 1$ and $\Delta\epsilon = 10/t$ and $\Delta\epsilon = 20/t$, respectively. Figure from reference [185].

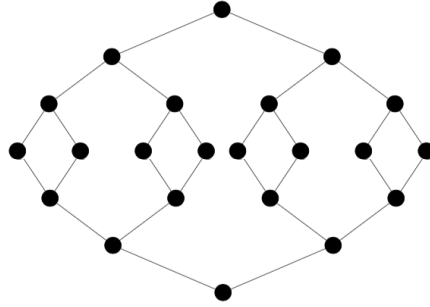


Figure 5.8: Schematic diagram of an ideal glued binary trees with depth-4 ($\mathcal{GBT} - 4$).

an adapted summation index to the tree shown in Figure 5.8. Figures 5.10 and 5.9 illustrate the effect of the particle-number fluctuations on quantum walks through a glued binary tree. The results shown in the insets of 5.10 are for a single particle

placed at zero time in the head node of an ideal glued tree depicted in Figure 5.8. The upper panel of 5.10 shows the probability

$$p_j(t) = \sum_{j=0}^{2^d} |\langle j | \psi(t) \rangle|^2 \quad (5.14)$$

summed over all nodes of depth level $j = 3$. The lower panel of Figure 5.10 is the probability of particle density transfer between the two ends of the glued tree. Interestingly, while the $\Delta \neq 0$ interactions affect the population of the $j = 3$ level, it can observed a small effect of these interactions on the head-to-bottom transfer of the particle for times $< 10 t^{-1}$ (see the inset of the lower panel of Figure 5.10).

In contrast, the same interactions have a much stronger effect on the head-to-bottom transfer through a disordered tree. As can be seen from Figure 5.10, the $\Delta \neq 0$ interactions accelerate the efficiency of particle transfer through the disordered tree, especially at short times by inducing oscillations as in the case of an open-ended binary tree discussed above. Figure 5.9 shows that these oscillations survive averaging over 100 disorder realizations. The disorder strength was set to $w = 5/t$ for these calculations.

While the methodology used here limits the size of the glued tree to seven levels, results indicate that the localization of quantum particles in disordered glued trees must be affected by the couplings between particle number subspaces. It would be interesting to see if the head-to-bottom transfer remains insensitive to these interactions and how the localization length is affected by such interactions in larger trees. To treat such problems, it is necessary to develop approximate computation techniques for few-particle systems in glued trees.

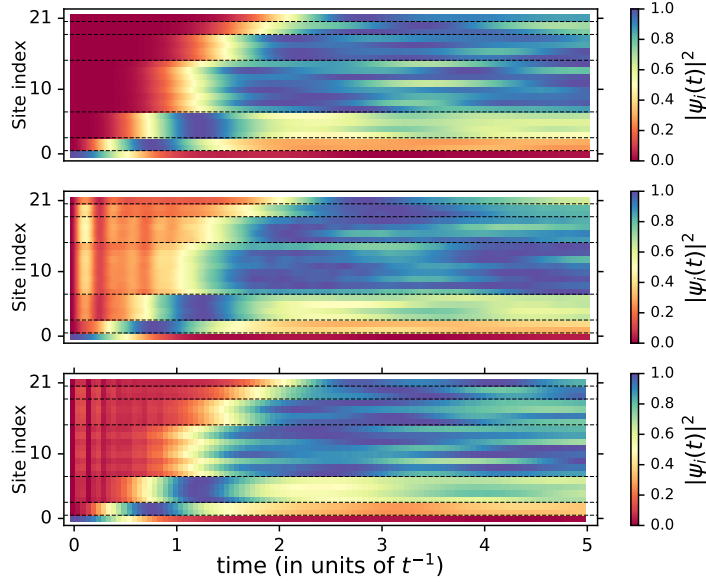


Figure 5.9: Average particle probability distribution in a disordered $\mathcal{GBT}4$ graph: upper panel – $\Delta/t = 0$, middle panel – $\Delta\epsilon = 10/t$ and lower panel – $\Delta\epsilon = 20/t$. For all panels we consider 100 realizations of disorder with a strength of $w = 5/t$ and $\gamma = 0$. The wave packet for a particle is initially placed in the head node of a $\mathcal{GBT}4$ graph. Figure from reference [185].

5.3 Conclusion

In this work, coherent quantum dynamics governed by the lattice Hamiltonians with number non-conserving interactions in the few-body limit was considered. As it is illustrated, the couplings between particle-number subspaces, even if much smaller than the energy separation between these subspaces, accelerate the dynamics of quantum walks in ideal lattices and binary trees and increase the localization length in disordered lattices. Effectively, these couplings provide new degrees of freedom, increasing the range of hopping due to virtual excitations and/or transient elimination of a single particle due to coupling to the vacuum state. As is showed, the number-changing interactions decrease the mixing and hitting times for quantum walks on binary trees.

Results show that the inverse participation ratio in disordered one-dimensional lattices decreases in the presence of number-changing interactions, signalling decrease of localization. This effect increases with increasing disorder strength, leading to larger changes of the inverse participation ratio in lattices with stronger on-site disorder. This is a direct consequence of the disorder-induced broadening of the particle energy bands. This broadening brings different particle number subspaces closer in energy, increasing the effect of the number-changing couplings and, consequently, the effective range of particle hopping.

Engineering lattice Hamiltonians to accelerate quantum dynamics has been of much recent interest due to potential applications in quantum computing and the study of the fundamental limits of the speed of correlation propagations in quantum many-body systems. Also of much interest is the localization dynamics of particles with long-range hopping in disordered lattices and graphs. This work illustrates that models of the type (5.1) can be used to study the effect of hopping range on Anderson localization and quantum walks spreading faster than ballistic expansion.

While non-interacting particles are known to be always localized in disordered 1D lattices, there is a localization - diffusion transition in 3D lattices [156]. Results indicate that the number-changing interactions must affect this transition. It would be interesting in future work to explore the quantitative effect of such interactions on the localization transition in 3D disordered lattices. It would also be interesting to explore the effect of such interactions on localization in 2D lattices. While non-interacting particles with short-range hopping are known to be always localized in 2D disordered lattices, particle interactions may lead to delocalization. Since the $\Delta \neq 0$ terms considered here create pairs of interacting particles in adjacent sites, these interactions may have non-trivial consequences on the localization in disordered 2D lattices.

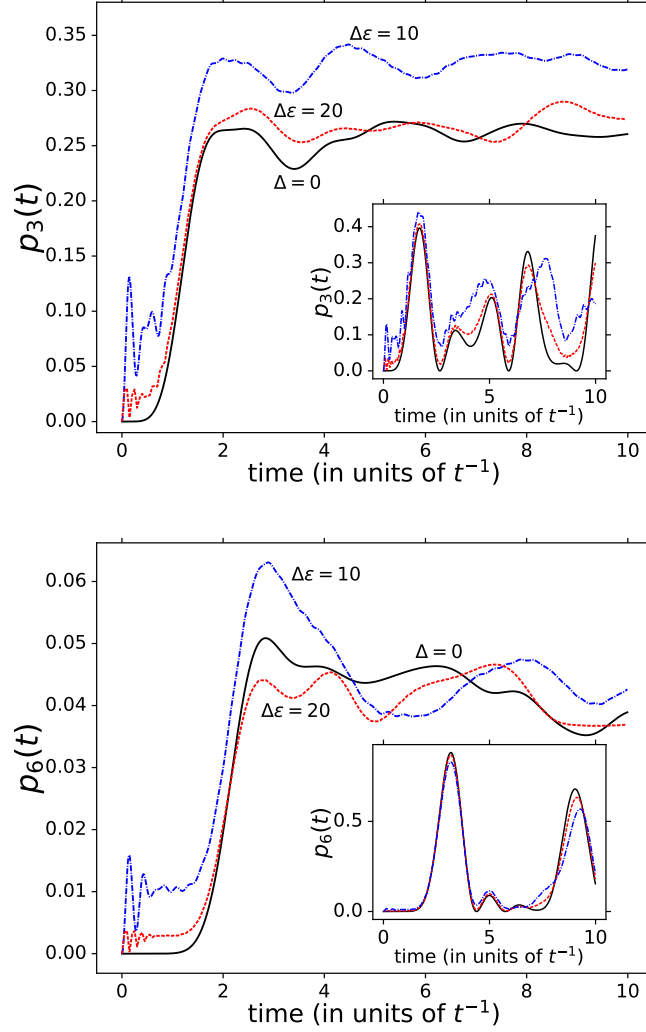


Figure 5.10: Average particle probability distribution in a disordered $\mathcal{GBT} - 4$ graph with strength $w = 5/t$. The wave packet for a single particle is initially placed in the head node of a $\mathcal{GBT} - 4$ graph shown in Figure 5.8. The upper panel displays the probability (5.14) summed over all nodes of layer $j = 3$; the lower panel shows $p_j(t)$ for the bottom node $j = 6$. Insets: The particle probability distributions for an ideal $\mathcal{GBT} - 4$ graph with $\Delta\epsilon = 10/t$ (blue dot-dashed) and $\Delta\epsilon = 20/t$ (red dashed). The broken curves show the results obtained with $\Delta/t = 1$ and the full black lines – $\Delta/t = 0$. Figure from reference [185].

Chapter 6

Quantum simulators with highly magnetic atoms

Nature isn't classical, dammit, and if you want to make a simulation of nature, you'd better make it quantum mechanical, and by golly it's a wonderful problem, because it doesn't look so easy.

– Richard P. Feynman

We show that Zeeman excitations of ultracold Dy atoms trapped in an optical lattice can be used to engineer extended Hubbard models with tunable inter-site and particle number-non-conserving interactions. We show that the ratio of the hopping amplitude and inter-site interactions in these lattice models can be tuned in a wide range by transferring the atoms to different Zeeman states. We propose to use the resulting controllable models for the study of the effects of direct particle interactions and particle number-non-conserving terms on Anderson localization. Results presented in this chapter are published in Ref. [184].

6.1 Introduction

There is currently growing interest in engineering lattice Hamiltonians with ultracold atoms and molecules [115]. Of particular interest are extended Hubbard models, which include interactions between particles in different lattice sites. Such models exhibit rich physics and have been used to explain the role of long-range

interactions in the context of superfluid - Mott insulator transitions [74], antiferromagnetism [27, 101], high-Tc superconductivity [56], twisted superfluidity [173], supersolids [141], self-trapping of bipolarons [175]. Extended Hubbard models are very difficult to solve numerically, especially for two- and three-dimensional lattices. Hence, the need to build experiments, where a many-body quantum system is described by an extended Hubbard model, whose parameters (in particular, the ratio of the hopping amplitude and the inter-site interaction energy) can be tuned by varying external fields, and where the particle densities can be imaged preferably with single site resolution. Tuning the parameters of the model, one could use such experiments to map out the phase diagrams.

There are many proposals for realizing lattice models, including extended Hubbard models [11, 29, 129, 135, 164, 202], with ultracold atoms or molecules trapped in optical lattices. However, if ultracold atoms or molecules are used as probe particles of such models, the inter-site interactions are usually very weak. Therefore, the measurements of the phase diagrams require extremely low temperatures and extremely long coherence times, which are often difficult to achieve in current experiments. A more promising approach is to trap ultracold molecules in an optical lattice in a Mott insulator phase (with one molecule per site) and use rotational excitations of trapped molecules as probe particles of lattice models [73, 84, 86, 110, 121, 128, 138, 198, 201]. Such excitations can be transferred between molecules in different sites due to dipole - dipole interactions. The dynamics of the excitations as well as their interactions can be controlled by external dc electric and/or microwave fields, leading to lattice models with tunable parameters. Experiments using excitations as probe particles of lattice models can tolerate much higher temperatures of atomic or molecular motion. However, it is currently not possible to create an optical lattice filled uniformly with molecules. On the other hand, ultracold atoms can be trapped in optical lattices with nearly uniform filling [74, 195]. Thus, it would be desirable to engineer extended Hubbard models with internal excitations of atoms (instead of molecules) trapped in a Mott insulator phase.

A series of experiments have recently demonstrated the cooling of highly magnetic Cr [75], Dy [119, 120], and Er [5, 65] atoms to quantum degeneracy. Such atoms interact via long-range magnetic dipole interactions and one can envision en-

engineering the same lattice models with magnetic atoms as with ultracold molecules. However, the internal level structure of magnetic atoms is more complex than the rotational structure of molecules and the nature of magnetic dipole interactions is different from that of electric dipole interactions. Motivated by the experiments on magnetic atoms and the work with ultracold molecules, we explore here the possibility of engineering extended Hubbard models with internal Zeeman excitations of ultracold magnetic atoms, such as Dy, trapped in a Mott insulator phase. Exploiting the unique nature of magnetic dipole interactions, we show that, for Zeeman excitations, the ratio of the hopping amplitude and inter-site interaction energy in the resulting lattice models can be tuned in a wide range by transferring the atoms to different Zeeman states. We discuss the advantages of using Zeeman excitations of magnetic atoms over rotational excitations of ultracold molecules. In particular, we show that the hopping of the Zeeman excitations in the lattice is insensitive to the magnitude of the magnetic field, which makes the coherent dynamics of excitations robust to field fluctuations. We show that Zeeman excitations in a diluted lattice of Dy atoms undergo Anderson localization over time scales less than one second and propose the models derived here for the study of the role of interactions and particle number fluctuations on Anderson localization.

6.2 Lattice Hamiltonian with Zeeman excitations

We consider an ensemble of open-shell atoms with non-zero electron spin (\mathcal{S}) and orbital angular momentum (\mathcal{L}) trapped in an optical lattice in the presence of an external DC magnetic field. We assume that the atoms fill the lattice uniformly with one atom per lattice site and that the atoms are not allowed to tunnel between different lattice sites. Thus, the atoms are separated by a large distance (≥ 260 nm) equal to half the wavelength of the trapping field. At such separations, the dominant interaction between the atoms in sites i and j is the magnetic dipole - dipole interaction \hat{V}_{ij} . For simplicity, we assume that the atoms are arranged in a one-dimensional array along the z -axis of the space-fixed coordinate frame. In this case,

$$\hat{V}_{ij} = \frac{\alpha}{r_{ij}^3} \left\{ \frac{1}{2} [\hat{J}_{i,+}\hat{J}_{j,-} + \hat{J}_{i,-}\hat{J}_{j,+}] - 2\hat{J}_{i,z}\hat{J}_{j,z} \right\}. \quad (6.1)$$

where \hat{J}_z and \hat{J}_\pm are the z -component and the raising/lowering operators of the total angular momentum $\mathbf{J} = \mathbf{L} + \mathbf{S}$, acting on the space of the eigenstates $|JM\rangle$ of \mathbf{J}^2 and \hat{J}_z , and α is the fine structure constant. Equation 6.1 is derived in Appendix C. The full Hamiltonian of the many-atom system is

$$\hat{\mathcal{H}} = \sum_i \{A\mathbf{L}_i \cdot \mathbf{S}_i + \mu_B(\mathbf{L}_i + 2\mathbf{S}_i) \cdot \mathbf{B}\} + \frac{1}{2} \sum_i \sum_{j \neq i} \hat{V}_{ij} \quad (6.2)$$

where A is the constant of the spin-orbit interaction, μ_B is the Bohr magneton and \mathbf{B} is the vector of an external magnetic field.

We assume that all atoms are initially prepared in the Zeeman state $|g\rangle$ and a small number of atoms is then transferred to another Zeeman state $|e\rangle$. Note that the state $|e\rangle$ can be lower or higher in energy than the state $|g\rangle$. Following the approach described in Refs. [2] (see also [196]), we derive the second-quantized Hamiltonian describing the Zeeman transitions in this system:

$$\hat{H}_{\text{ex}} = \nu_g + \sum_i \sum_{e'} \left\{ \varepsilon_{e'} - \varepsilon_g + \sum_{j \neq i} [\langle e'_i | \langle g_j | \hat{V}_{ij} | e'_i \rangle | g_j \rangle - \langle g_i | \langle g_j | \hat{V}_{ij} | g_i \rangle | g_j \rangle] \right\} \hat{c}_{i,e'}^\dagger \hat{c}_{i,e'} \quad (6.3)$$

$$+ \sum_{i,j \neq i} \sum_{e',e''} \langle g_i | \langle e'_j | \hat{V}_{ij} | e''_j \rangle | g_j \rangle \hat{c}_{i,e''}^\dagger \hat{c}_{j,e'} + \sum_{i,j \neq i} \sum_{e',e''} (1 - \delta_{e',e''}) \langle e'_i | \langle g_j | \hat{V}_{ij} | e''_i \rangle | g_j \rangle \hat{c}_{i,e'}^\dagger \hat{c}_{i,e''} \quad (6.4)$$

$$+ \frac{1}{2} \sum_{i,j \neq i} \sum_{e',e''} \sum_{f',f''} [\delta_{e',e''} \delta_{f',f''} \langle g_i | \langle g_j | \hat{V}_{ij} | g_i \rangle | g_j \rangle + \langle e'_i | \langle f'_j | \hat{V}_{ij} | e''_i \rangle | f''_j \rangle - 2\delta_{f',f''} \langle e'_i | \langle g_j | \hat{V}_{ij} | e''_i \rangle | g_j \rangle] \hat{c}_{i,e'}^\dagger \hat{c}_{i,e''} \hat{c}_{j,f'}^\dagger \hat{c}_{j,f''} \quad (6.5)$$

$$+ \sum_{i,j \neq i} \sum_{e'} [\langle g_i | \langle g_j | \hat{V}_{ij} | g_i \rangle | e'_j \rangle \hat{c}_{j,e'} + \langle g_i | \langle e'_j | \hat{V}_{ij} | g_i \rangle | g_j \rangle \hat{c}_{j,e'}^\dagger] \quad (6.6)$$

$$+ \frac{1}{2} \sum_{i,j \neq i} \sum_{e',e''} [\langle g_i | \langle g_j | \hat{V}_{ij} | e'_i \rangle | e''_j \rangle \hat{c}_{i,e'} \hat{c}_{j,e''} + \langle e'_i | \langle e''_j | \hat{V}_{ij} | g_i \rangle | g_j \rangle \hat{c}_{i,e'}^\dagger \hat{c}_{j,e''}^\dagger] \quad (6.7)$$

$$+ \frac{1}{2} \sum_{i,j \neq i} \sum_{e',e'',f'} [\langle e'_i | \langle g_j | \hat{V}_{ij} | e''_i \rangle | f'_j \rangle - \delta_{e',e''} \langle g_i | \langle g_j | \hat{V}_{ij} | g_i \rangle | f'_j \rangle] \hat{c}_{i,e'}^\dagger \hat{c}_{i,e''} \hat{c}_{j,f'} \quad (6.8)$$

$$+ \frac{1}{2} \sum_{i,j \neq i} \sum_{e',e'',f'} [\langle e'_i | \langle f'_j | \hat{V}_{ij} | e''_i \rangle | g_j \rangle - \delta_{e',e''} \langle g_i | \langle f'_j | \hat{V}_{ij} | g_i \rangle | g_j \rangle] \hat{c}_{i,e'}^\dagger \hat{c}_{i,e''} \hat{c}_{j,f'}^\dagger \quad (6.9)$$

where

$$v_g = N\varepsilon_g + \frac{1}{2} \sum_i \sum_{j \neq i} V_{ij}^{gg}, \quad (6.10)$$

N is the number of atoms, ε_g and $\varepsilon_{e'}$ are the energies of the atomic states $|g\rangle$ and $|e'\rangle$, and

$$V_{ij}^{gg} = \langle g_i | \langle g_j | \hat{V}_{ij} | g_i \rangle | g_j \rangle. \quad (6.11)$$

Here, we assume that the Zeeman states $e', e'', f', f'' \neq g$ and use the operators $\hat{c}_{i,e'}^\dagger$ and $\hat{c}_{i,e'}$ defined by $\hat{c}_{i,e'}^\dagger |g_j\rangle = \delta_{ij} |e'_j\rangle$ and $\hat{c}_{i,e'} |e'_j\rangle = \delta_{ij} |g_j\rangle$. For the purposes of this work, it is convenient to rewrite this complex Hamiltonian as

$$\hat{H}_{\text{ex}} = v_g + \sum_i (\Delta\varepsilon_{eg} + d_i) \hat{c}_i^\dagger \hat{c}_i + \sum_i \sum_{j \neq i} t_{ij} \hat{c}_j^\dagger \hat{c}_i + \quad (6.12)$$

$$\frac{1}{2} \sum_i \sum_{j \neq i} v_{ij} c_i^\dagger c_i c_j^\dagger c_j + \quad (6.13)$$

$$\frac{1}{2} \sum_i \sum_{j \neq i} t_{ij} (\hat{c}_i^\dagger \hat{c}_j^\dagger + \hat{c}_i \hat{c}_j) + \sum_i \sum_{j \neq i} s_{ij} (\hat{c}_i^\dagger + \hat{c}_i) + \sum_i \sum_{j \neq i} p_{ij} (\hat{c}_i^\dagger + \hat{c}_i) \hat{c}_j^\dagger \hat{c}_j \quad (6.14)$$

$$+ \mathcal{H}(e' \neq e, g) \quad (6.15)$$

where the operators \hat{c}_i^\dagger and \hat{c}_i are defined by $\hat{c}_i^\dagger |g_j\rangle = \delta_{ij} |e_j\rangle$ and $\hat{c}_i |e_j\rangle = \delta_{ij} |g_j\rangle$, $\Delta\varepsilon_{eg}$ is the energy separation between the states $|e\rangle$ and $|g\rangle$, and the parameters of the Hamiltonian are

$$d_i = \sum_{j \neq i} d_{ij}, \quad (6.16)$$

$$d_{ij} = \left\{ V_{ij}^{ge} - V_{ij}^{gg} \right\}, \quad (6.17)$$

$$v_{ij} = V_{ij}^{ee} + V_{ij}^{gg} - 2V_{ij}^{eg}, \quad (6.18)$$

$$V_{ij}^{eg} = V_{ij}^{ge} = \langle g_i | \langle e_j | \hat{V}_{ij} | g_i \rangle | e_j \rangle \quad (6.19)$$

$$V_{ij}^{ee} = \langle e_i | \langle e_j | \hat{V}_{ij} | e_i \rangle | e_j \rangle \quad (6.20)$$

$$t_{ij} = \langle g_i | \langle e_j | \hat{V}_{ij} | e_i \rangle | g_j \rangle \quad (6.21)$$

$$s_{ij} = \langle e_i | \langle g_j | \hat{V}_{ij} | g_i \rangle | g_j \rangle \quad (6.22)$$

and

$$p_{ij} = \langle e_i | \langle g_j | \hat{V}_{ij} | e_i \rangle | e_j \rangle - \langle e_i | \langle g_j | \hat{V}_{ij} | g_i \rangle | g_j \rangle. \quad (6.23)$$

The terms (6.12), (6.13) and (6.14) are a part of the full Hamiltonian that describes the Zeeman transitions only within the four-state subspace $|a\rangle|b\rangle$ with both $|a\rangle$ and $|b\rangle$ being either $|g\rangle$ or $|e\rangle$. If the energy gap for the $|g\rangle \rightarrow |e\rangle$ transition were far detuned from all other energy gaps in the Zeeman level spectrum, it would be sufficient to consider the part of the Hamiltonian given by Eqs. (6.12), (6.13) and (6.14). It is important to note that for highly magnetic atoms it may be necessary to consider Zeeman states outside of this subspace. Figure 1a shows that the Zeeman states of a Dy atom in the ground electronic state form a ladder of nearly equidistant levels at weak magnetic fields. This pattern of energy levels is characteristic of highly magnetic atoms with zero or negligible hyperfine structure. This pattern of energy levels allows for transitions to states outside of the subspace spanned by $|g\rangle$ or $|e\rangle$. For example, two atoms in the $|g\rangle$ state may interact to produce two Zeeman states with energies just above and just below that of $|g\rangle$. Such interactions are induced by the matrix elements in Eq. (6.7). The full Hamiltonian must also include the terms that describe the interactions of two atoms in states $e', e'' \neq g$ to produce atoms in other states $f', f'' \neq g, e$. Since the majority of atoms are in a particular state $|g\rangle$, we assume that such interactions are unlikely and neglect them.

Various lattice models can be engineered by controlling the magnitude of the different matrix elements of the magnetic dipole interaction entering Eqs. (6.3) - (6.9).

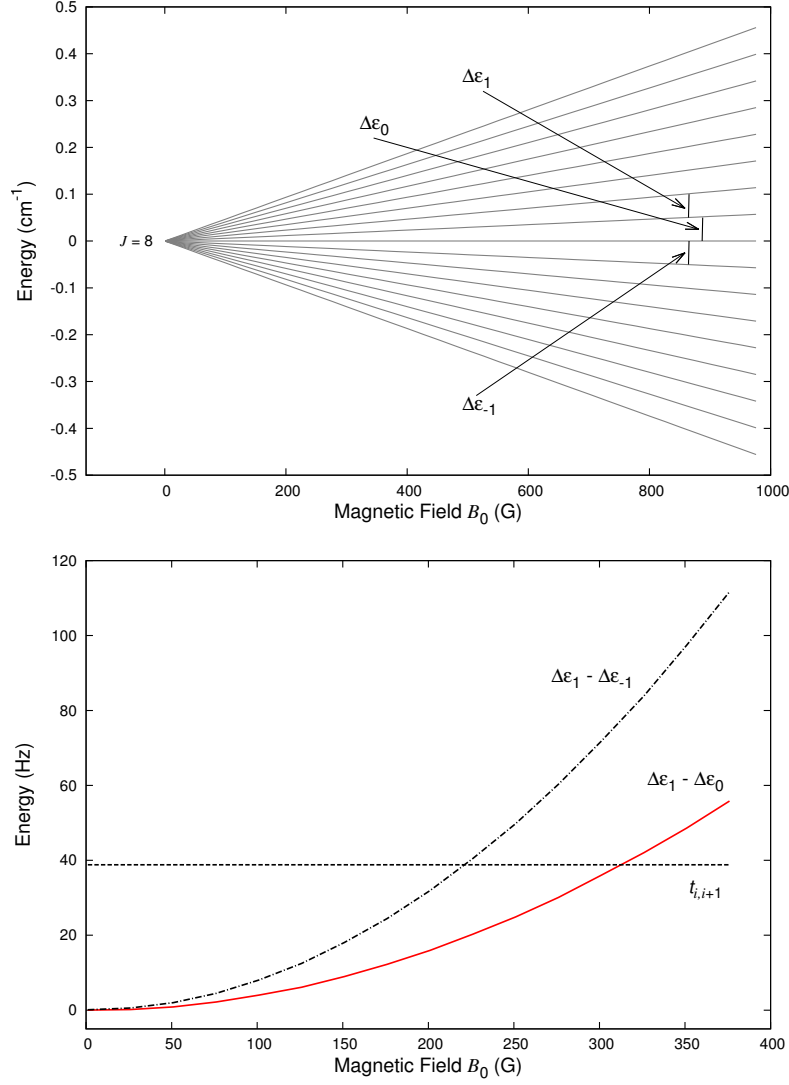


Figure 6.1: Upper panel: Zeeman levels of a $\text{Dy}(^5I)$ atom in the lowest-energy spin-orbit state characterized by $J = 8$ in a magnetic field $\mathbf{B} = B_0 \hat{z}$. Lower panel: the solid curve – difference of the energy gaps $(\epsilon_{M=2} - \epsilon_{M=1}) - (\epsilon_{M=1} - \epsilon_{M=0})$; the dot-dashed curve – difference of the energy gaps $(\epsilon_{M=2} - \epsilon_{M=1}) - (\epsilon_{M=0} - \epsilon_{M=-1})$. The horizontal dashed line shows the magnitude of the matrix element $t_{i,i+1}$ in Eq. (6.21) for Dy atoms with $|g\rangle = |J = 8, M = 0\rangle$ and $|e\rangle = |J = 8, M = 1\rangle$ in an optical lattice with $a = 266$ nm. Figure from reference [184].

6.3 Engineering lattice models

In this section we show (i) how to simplify the lattice Hamiltonian presented in Section II by applying magnetic fields; and (ii) how to tune the relative magnitudes of the parameters of the resulting lattice models by transferring atoms into different states. We illustrate the tunable range of the parameters by calculating the model parameters for the specific example of Dy atoms in an optical lattice.

6.3.1 $t - V$ model

Eqs. (6.12) and (6.13) represents a $t - V$ model [37], an extended single band, Hubbard model for hard-core bosons [76, 135, 141]. This model can be studied with the Zeeman excitations if the effect of the terms (6.14) and (6.15) are suppressed. As we show below, this can be achieved by applying a finite magnetic field and introducing a small admixture of different M -states into the eigenstates $|JM\rangle$.

Eqs. (6.3) - (6.9) and (6.12) – (6.14) can be separated into terms that conserve the number of excitations (Eqs. 6.3 – 6.5, 6.12 and 6.13) as well as particle number-non-conserving terms (Eqs. 6.6 – 6.9 and 6.14). If the Zeeman states form a ladder of equidistant states, the particle number-non-conserving terms can be further separated into energy-conserving (some terms in Eq. 6.7) and energy-non-conserving terms (Eqs. 6.6 – 6.9, 6.14). The effect of the energy-non-conserving terms can be eliminated by applying a finite magnetic field such that the energy difference between the Zeeman levels is significantly larger than the magnitude of the matrix elements appearing in Eqs. (6.6) – (6.9) and (6.14).

In order to eliminate the effect of all terms in Eq. (6.15), it is necessary to make the energy gap for the $|g\rangle \rightarrow |e\rangle$ transition unique, i.e. different from the energy gaps in the Zeeman spectrum just below and just above the states $|g\rangle$ and $|e\rangle$. This can be achieved by applying a magnetic field strong enough to shift the Zeeman levels due to couplings between different total angular momentum states. As illustrated in the lower panel of Figure 6.1, these couplings introduce a differential in the energy gaps between different Zeeman states. To illustrate this, we plot in Figure 6.1 the of the energy gaps between the states correlating with the states $|J = 8, M = -1\rangle$ and $|J = 8, M = 0\rangle$; states $|J = 8, M = 0\rangle$ and $|J = 8, M = +1\rangle$ and states $|J = 8, M = +1\rangle$ and $|J = 8, M = +2\rangle$, as functions of

B_0 . As Figure 6.1 shows, the magnetic field with $B_0 \approx 200 - 300$ G produces the differential of the energy gaps equal to the matrix elements $t_{i,i+1}$ for Dy atoms on an optical lattice with $a = 266$ nm. At fields with $B_0 > 300$ G, the difference in the energy gaps becomes much larger than any of the matrix elements in Eq. (6.15) so the Hamiltonian (6.12) – (6.15) reduces to the $t - V$ model.

The parameters of the $t - V$ model can be tuned by transferring atoms into different Zeeman states. If the $|g\rangle$ and $|e\rangle$ states are the Zeeman states $|g\rangle = |JM\rangle$ and $|e\rangle = |JM'\rangle$, the matrix elements (6.11) and (6.21) of the operator (6.1) can be written as follows:

$$d_{ij} = V_{ij}^{ge} - V_{ij}^{gg} = \frac{2\alpha}{r_{ij}^3} (M^2 - M'M) \quad (6.24)$$

and

$$t_{ij} = \frac{\alpha}{2r_{ij}^3} \left[a_+^i b_-^j \delta_{M',M+1}^i \delta_{M',M-1}^j + a_-^i b_+^j \delta_{M',M-1}^i \delta_{M',M+1}^j \right], \quad (6.25)$$

with

$$a_\pm^i = [J(J+1) - M(M \pm 1)]^{1/2} \quad (6.26)$$

$$b_\pm^j = [J(J+1) - M'(M' \pm 1)]^{1/2} \quad (6.27)$$

The interaction between the Zeeman excitations (6.18) can be written as

$$v_{ij} = - \left[(V_{ij}^{eg} - V_{ij}^{gg}) + (V_{ij}^{eg} - V_{ij}^{ee}) \right] = - \frac{2\alpha}{r_{ij}^3} (M - M')^2 \quad (6.28)$$

These equations show that the diagonal matrix elements V_{ij}^{gg} and V_{ij}^{ee} , and hence d_{ij} and v_{ij} are non-zero, provided both $M \neq 0$ and $M' \neq 0$. This is different from the case of the electric dipole - dipole interaction between molecules [85]. The electric dipole interaction must couple states of the opposite parity. Therefore, if $|g\rangle$ and $|e\rangle$ are the eigenstates of a molecular Hamiltonian in the absence of electric fields, the matrix elements d_{ij} and v_{ij} of the electric dipole - dipole interaction vanish. These interactions can be induced in an ensemble of polar molecules by

applying an external electric field that mixes the rotational states with different parity [85, 197]

In contrast, the matrix elements of the magnetic dipole - dipole interaction (6.24) and (6.25) should not be expected to vary significantly with an external magnetic field. This will be illustrated and discussed in the following section, using the example of Dy atoms on an optical lattice. As follows from Eqs. (6.24) and (6.25), the relative weights of the two couplings can be tuned by choosing different Zeeman states $|JM\rangle$ as the $|g\rangle$ and $|e\rangle$ states. Notice, for example, that for the particular case of $|g\rangle$ being the state $|J, M = 0\rangle$, the magnitudes of d_{ij} , and consequently d_i , vanish.

6.3.2 $t - V$ model with Dy atoms

We illustrate the range of controllability of the parameters of the $t - V$ models using an example of Dy atoms in an optical lattice. The absolute magnitudes of d_{ij} , t_{ij} and v_{ij} increase with J as the square of the magnetic moment. The ground electronic state of Dy is characterized by the total angular momentum $J = 8$ so Dy atoms have a large magnetic moment (10 Bohr magnetons) and a manifold of Zeeman states displayed in Figure 6.1. The Zeeman structure of Dy allows for the possibility of using the state $|M = 0\rangle$ as the $|g\rangle$ state, leading to the value $d_{ij} = 0$.

If the states for the Zeeman excitations in an ensemble of Dy atoms are chosen to be well-defined angular momentum states $|g\rangle = |JM\rangle$ and $|e\rangle = |JM'\rangle$, Eq. (6.25) shows that $t_{ij} = 0$ unless $|M - M'| = 1$. Eq. (6.28) shows that the interaction v_{ij} is $\propto (M - M')^2$ so it is independent of M and M' , if $|M - M'| = 1$. However, the parameter t_{ij} is sensitive to the magnitudes of M and M' . This is illustrated in the upper panel of Figure 6.2. The ratio t_{ij}/v_{ij} can thus be tuned by transferring atoms into the Zeeman states with different M , as illustrated in the lower panel of Figure 6.2. Notice that the ratio t_{ij}/v_{ij} is always negative, which means that the interactions between the excitations are always effectively attractive. The largest magnitude of the ratio $t_{ij}/v_{ij} \approx -18$ can be achieved when the atoms are prepared in the Zeeman state with $M = 0$ and excited to the Zeeman state with $M = +1$, while the smallest magnitude of the ratio $t_{ij}/v_{ij} \approx -4$ can be achieved by preparing the atoms in the maximally stretched state $|J = 8, M = -8\rangle$ or $|J = 8, M = +8\rangle$.

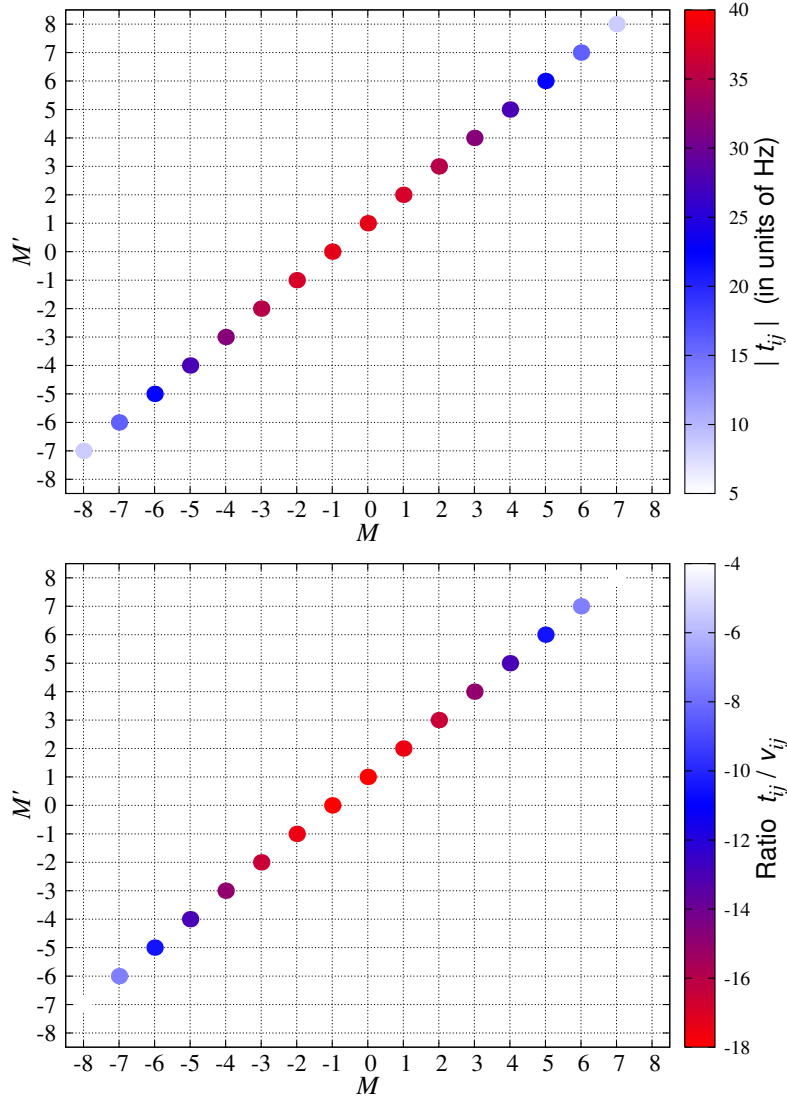


Figure 6.2: The magnitudes of the coupling constants t_{ij} (upper panel) and the ratio t_{ij}/v_{ij} (lower panel) with $j = i \pm 1$ for the Zeeman states of Dy corresponding to $|g\rangle \Rightarrow |JM\rangle$ and $|e\rangle \Rightarrow |JM'\rangle$. The calculations are for the magnetic field $\mathbf{B} = B_0(0.1\hat{x} + \hat{z})$ with $B_0 = 100$ G. The Zeeman states in this magnetic field retain 96% of the eigenstates of \mathbf{J}^2 and \mathbf{J}_z . Figure from reference [184].

As illustrated in Figure 6.3 the absolute magnitude of v_{ij} can be tuned if the atoms are prepared in coherent superpositions of states with different M . Consider for example the superpositions $|g\rangle = \alpha|JM\rangle + \beta|J, M + \delta\rangle$ and $|e\rangle = \alpha'|JM'\rangle + \beta'|J, M' + \delta'\rangle$. For the parameter t_{ij} to be non-zero, either $|M - M'|$ or $|M - M' + \delta - \delta'|$ must be 1. However, there is no such restriction on the matrix elements determining the magnitude of v_{ij} . As follows from Eq. (6.28), the magnitude of v_{ij} is expected to increase with increasing the difference between the angular momentum projections of the states participating in the excitation. This is graphically illustrated in Figure 6.3, showing that the magnitude of v_{ij} can reach 600 Hz, if $M - M' = 16$. This suggests that the ratio t_{ij}/v_{ij} can be tuned by preparing the atoms in the coherent superpositions of the following kind: $|g\rangle = \alpha|JM\rangle + \beta|J, M + \delta\rangle$ and $|e\rangle = \alpha'|JM + 1\rangle + \beta'|J, M + \delta'\rangle$, Figure 6.4. The parameters t_{ij} and v_{ij} for these states are both non-zero and the magnitude of v_{ij} can be modified by varying the value of $|\delta - \delta'|$.

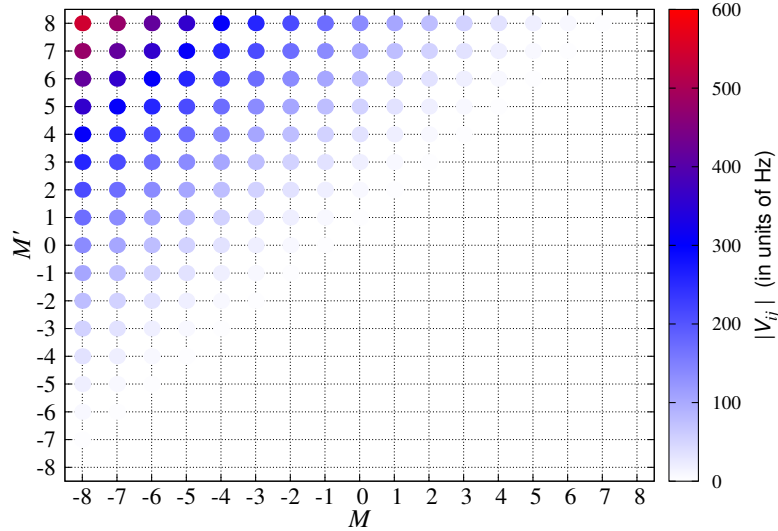


Figure 6.3: The magnitude of the coupling constant v_{ij} with $j = i \pm 1$ for the Zeeman states of Dy corresponding to $|g\rangle \Rightarrow |JM\rangle$ and $|e\rangle \Rightarrow |JM'\rangle$. The calculations are for the magnetic field $\mathbf{B} = B_0(0.1\hat{x} + \hat{z})$ with $B_0 = 100$ G. The Zeeman states in this magnetic field retain 96% of the eigenstates of \mathbf{J}^2 and \mathbf{J}_z . Figure from reference [184].

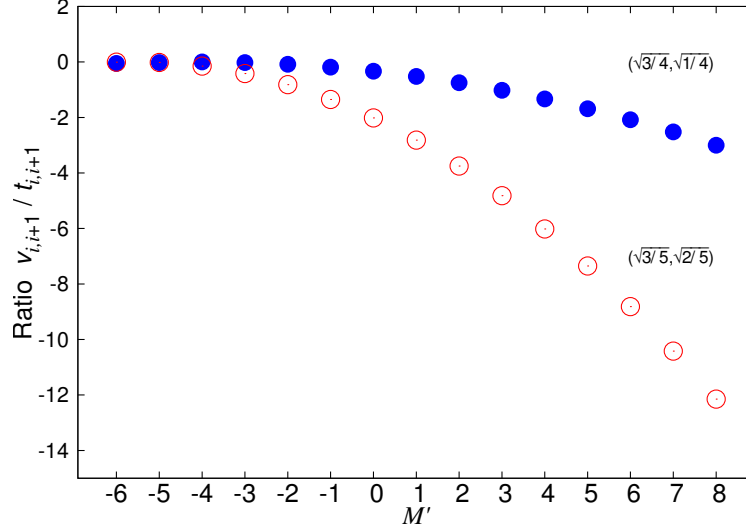


Figure 6.4: The ratio $v_{i,i+1}/t_{i,i+1}$ for the Zeeman states of Dy corresponding to $|g\rangle = |J=8, M=-7\rangle$ and $|e\rangle = a|J=8, M=-8\rangle + b|J=8, M'\rangle$: full circles – $a = \sqrt{3/4}, b = \sqrt{1/4}$; open circles – $a = \sqrt{3/5}, b = \sqrt{2/5}$. The calculations are for the magnetic field $\mathbf{B} = B_0(0.1\hat{x} + \hat{z})$ with $B_0 = 100$ G. The Zeeman states in this magnetic field retain 96% of the eigenstates of \mathbf{J}^2 and J_z . Figure from reference [184].

The interaction of atoms with a magnetic field couples states with different total angular momenta J , which may - in principle - modify the atomic states $|g\rangle$ and $|e\rangle$, and, consequently, the lattice model parameters. It is important to examine the effect of an external magnetic field on the lattice model parameters. To do this we diagonalized the full Hamiltonian of the Dy atom in a magnetic field $\mathbf{B} = B_0(0.1\hat{x} + \hat{z})$ and used the eigenstates to evaluate the model parameters in Eqs. (6.12) – (6.13). Since the states of different J in the Dy atom are separated by large energy gaps ($> 1000 \text{ cm}^{-1}$) due to the spin-orbit interaction, the eigenstates of Dy in a magnetic field are nearly identical to the angular momentum states $|JM\rangle$. Figure 6.4 shows the nearest-neighbour coupling parameters $t_{i,i+1}$ and $v_{i,i+1}$ for a one-dimensional array of Dy atoms on an optical lattice with the lattice site separation $a = 266 \text{ nm}$ computed for two pairs of Zeeman states at different magnetic fields. The results shown in Figure 6.5 illustrate that the Hamiltonian parameters

do not change with the magnetic field in the interval of field strengths between zero and 5000 G. This is important because it shows that the magnetic field can be used to separate the Zeeman states in order to create isolated two-level systems or tuned to the limit of vanishing field where the terms in Eq. (6.14) become important, without affecting the parameters of excitation interactions.

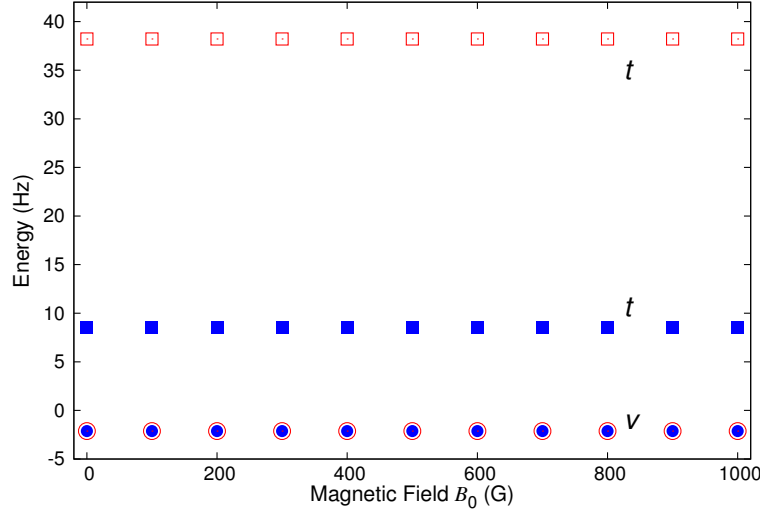


Figure 6.5: The magnetic field dependence of the quantities $t_{i,i+1}$ (squares) and $v_{i,i+1}$ (circles) defined in Eqs. (6.24) and (6.25) for two different pairs of the Zeeman state of Dy($J = 8$) atoms: the full symbols – the results for $|g\rangle = |J = 8, M = -8\rangle$ and $|e\rangle = |J = 8, M = -7\rangle$; the open symbols – the results for $|g\rangle = |J = 8, M = 0\rangle$ and $|e\rangle = |J = 8, M = +1\rangle$. The magnetic field is given by $\mathbf{B} = B_0 (0.1\hat{x} + \hat{z})$. The Zeeman states in such a magnetic field retain 96% of the eigenstates of \mathbf{J}^2 and \mathbf{J}_z . Figure from reference [184].

6.3.3 Particle number-non-conserving interactions

In the limit of weak magnetic fields, as $\Delta\epsilon_{eg} \rightarrow 0$, the energy separation between different particle number states of the model (6.12) decreases to the minimum of d_i . As follows from Eq. (6.24), this parameter can be eliminated if the ground state $|g\rangle$ is chosen to be $|J, M = 0\rangle$. At weak magnetic fields, the particle number-

non-conserving terms (6.14) as well as the terms in Eq. (6.15) must be included in the Hamiltonian. Number non-conserving interactions may mediate effective long-range hopping (for example, a particle can move in a lattice by virtual transitions to the three-particle subspace and back). As such, these interactions may have non-trivial effects on the dynamics of quantum walks and localization in disordered lattices. Such interactions arise in the context of excitons in molecular crystals [3, 12]. However, they are usually negligibly small and difficult to investigate. As shown below, number non-conserving interactions can be made significant in the system considered here.

We first note that if the array of atoms is arranged along the magnetic field direction, the matrix elements of the operator (6.1) that determine the parameters s_{ij} and p_{ij} in Eq. (6.14) vanish. This simplifies the resulting lattice models to include only the first of the particle number-non-conserving terms in Eq. (6.14). If desired, the terms s_{ij} and p_{ij} can be tuned to finite values if the magnetic field direction is changed or the atoms are prepared in coherent superpositions of different M -states. For example, if $|g\rangle = |J, M = 0\rangle$ and $|e\rangle = \alpha|J = 8, M = 0\rangle + \beta|J = 8, M = 1\rangle$, all of t_{ij} , s_{ij} and p_{ij} become non-zero. Here, we assume that the magnetic field is directed along the atomic array and that $s_{ij} = 0$ and $p_{ij} = 0$.

Care must be taken when considering the limit $\Delta\epsilon_{eg} \rightarrow 0$. In this limit, multiple Zeeman states become degenerate and it may be necessary to consider interband couplings determined by Eq. (6.15). This may be useful if complicated models, including multiple excitations of different kind, are desired. Note, however, that if $|g\rangle$ and $|e\rangle$ are states with well-defined M and M' , a two-atom state $|M\rangle|M'\rangle$ can only be coupled to the same state, the state $|M'\rangle|M\rangle$ or a state $|M \pm 1\rangle|M' \mp 1\rangle$. The matrix elements of the dipole - dipole interaction $\langle M, M' | \hat{V}_{ij} | M \pm 1, M' \mp 1 \rangle$ change the number and type of excitations in the atomic ensemble. These processes can be eliminated if the state $|g\rangle$ is chosen to be $|J, M = \pm J\rangle$. In this case, the effective lattice model describing the dynamics of $|g\rangle \rightarrow |e\rangle$ excitations is

$$\begin{aligned}
\hat{H}_{\text{ex}} = & v_g + \sum_i (\Delta\epsilon_{eg} + d_i) \hat{c}_i^\dagger \hat{c}_i + \sum_i \sum_{j \neq i} t_{ij} \hat{c}_j^\dagger \hat{c}_i + \\
& \frac{1}{2} \sum_i \sum_{j \neq i} v_{ij} c_i^\dagger c_i c_j^\dagger c_j + \frac{1}{2} \sum_i \sum_{j \neq i} t_{ij} \left(\hat{c}_i^\dagger \hat{c}_j^\dagger + \hat{c}_i \hat{c}_j \right)
\end{aligned} \tag{6.29}$$

It is important to note that this model is valid as long as $\Delta\epsilon_{eg}$ (which is determined by the magnitude of the magnetic field) is significantly larger than t_{ij} . In this limit, the effect of the number-non-conserving terms is perturbative, i.e. a single excitation remains predominantly in the single-particle subspace, undergoing virtual transitions to the three-particle subspace. If the energy gap $\Delta\epsilon_{eg}$ is so small that the interactions (6.29) lead to the creation of multiple particles, other terms in Eq. (6.7) must be included, making the Hamiltonian more complex.

If the effects of the interactions v_{ij} are to be removed, one can choose the states $|g\rangle = |J, M=0\rangle$ and $|e\rangle = |J, M=1\rangle$. In this case, $|t_{ij}| \gg |v_{ij}|$ (see Figure 4). However, the lattice model for these excitations is also affected by terms in Eq. (6.7), which lead to leaking of the $|e\rangle$ -state populations to other Zeeman states of higher energy. These terms lead to the spontaneous creation of atoms in Zeeman states above and below the energy of the state with $M=0$, as well as the inverse process. The Zeeman state populations must eventually return to states e and g , as the total number of the Zeeman states is finite and small. These terms thus serve as an additional source of particle number-non-conserving interactions that generate atoms in state e .

6.3.4 Anderson localization of Zeeman excitations

Until now, we assumed that the atoms populate the optical lattice uniformly. If the lattice is populated partially (which is more often the case in experiments), the empty lattice sites serve as impurities that can scatter the Zeeman excitations. Since the distribution of empty sites is random, the Zeeman excitations thus propagate in a randomly diluted lattice. Tuning the models as described above suggests an interesting opportunity to explore the role of direct particle interactions and number non-conserving forces on Anderson localization in disordered lattices [7, 62]. In

addition, the ability to design optical lattices with various dimensionalities and geometries can be used to verify the scaling hypothesis of Anderson localization [1] as well as Anderson localization of particles with long-range hopping in various geometries [118]. Here, we explore if the parameters of the models based on Zeeman excitations of Dy are significant enough to allow Anderson localization over experimentally feasible time- and length-scales.

We consider an isolated Zeeman excitation in a one-dimensional array of 1000 Dy atoms trapped in an optical lattice with $a = 266$ nm containing 20 % of empty lattice sites. We use the parameters corresponding to the $|J = 8, M = 0\rangle \rightarrow |J = 8, M = +1\rangle$ excitation and compute the dynamics of quantum walk for the Zeeman excitation placed at $t = 0$ on a single atom in the middle of the lattice. The wave packet of the excitations is propagated by computing the time-evolution operator, as described in detail in Ref. [200]. The results of each dynamical propagation are averaged over 100 disorder realizations (random distributions of empty lattice sites).

The results shown in Figure 6.6 illustrate that the Zeeman excitation forms an exponentially localized spatial distribution within one second. The width of the distribution characterized as the length L containing 90 % of the excitation probability exhibits a short-time oscillation which is likely an effect of coherent back scattering and approaches the value of ~ 20 lattice sites in the limit of long time. These results can be directly mapped onto the results describing Anderson localization for rotational excitations in an ensemble of polar molecules [200] and the electronic excitations in an ensemble of Rydberg atoms [150].

6.4 Conclusion

In this work, we consider Zeeman excitations in an ensemble of highly magnetic atoms (such as Dy) trapped in an optical lattice, with one atom per lattice site. The Zeeman excitations can travel in the lattice due to energy transfer between the atoms. The most important results of this work can be summarized as follows:

- We show that superpositions of the Zeeman excitations can be used to simulate the $t - V$ model (the single-band, extended Bose-Hubbard model for hard-core bosons). The parameters of the model (most importantly, the ra-

tio of the hopping amplitude and the inter-site interaction energy) can be tuned by preparing the atoms in different Zeeman states. For an ensemble of Dy atoms on an optical lattice with $a = 266$ nm, we show that the inter-site interaction can be engineered to be as large as 600 Hz.

- We illustrate that the parameters of the model (hopping amplitudes and inter-site interactions) are insensitive to the magnetic field. This has two significant consequences. First, an external magnetic field can be used to uncouple the electron degrees of freedom from nuclear spins, thereby removing complications associated with the hyperfine structure of atoms and the degeneracies of the Zeeman states. Second, an external magnetic field can be used to separate the Zeeman states, leading to suppression of energy- and particle number-non-conserving terms.
- We show that the same Hamiltonian can be used to simulate a lattice model with significant $c_i^\dagger c_j^\dagger$ terms, leading to particle number interactions. These interactions mediate effective interactions modifying the hopping of particles and can be used to produce entangled pairs [88].
- Since the lattice with randomly distributed empty sites leads to a quantum percolation model for the Zeeman excitations, we propose to apply the models derived here for the study of Anderson localization induced by off-diagonal disorder. In particular, our results suggest the possibility of studying the role of inter-site interactions and particle number fluctuations on quantum localization in diluted lattices. We show that for an optical lattice with $a = 266$ nm partially populated with Dy atoms, Anderson localization of excitations placed on individual atoms occurs over timescales less than a second.

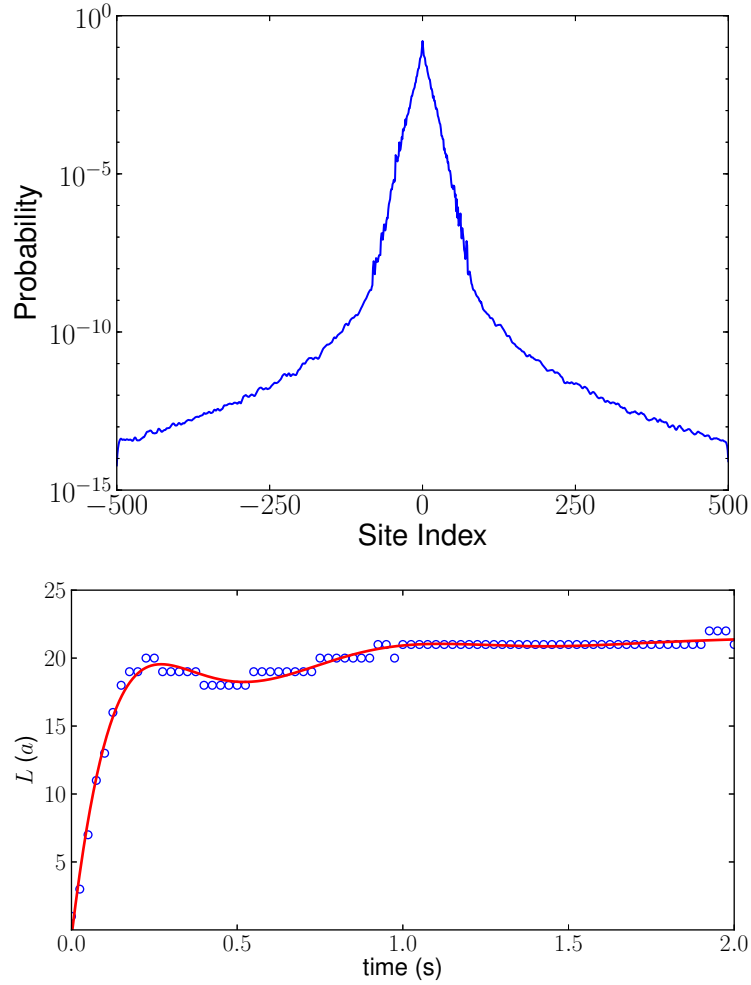


Figure 6.6: Anderson localization of the $|J = 8, M = 0\rangle \rightarrow |J = 8, M = +1\rangle$ excitation in a one-dimensional array of Dy atoms on an optical lattice with $a = 266$ nm and 20 % of the lattice sites empty. The upper panel shows the probability distribution for the atoms in the corresponding site to be in the excited state at $t = 2$ seconds formed by a single excitation placed at $t = 0$ in the middle of a lattice with 1000 sites. The lower panel shows the width of the excitation probability distribution as a function of time. Figure from reference [184].

Chapter 7

Conclusion

*Our intelligence is what makes us human,
and AI is an extension of that quality.*

— Yann LeCun

The journey to understand quantum phenomena has been challenging. In the last decades, knowledge about the quantum world has gained a huge boost both by building quantum simulators and by the growth in the computational power. For example, the field of theoretical physics has been able to study larger systems as the computational power has grown over the years. Furthermore, quantum simulators have also improved in recent years due to progress in the precision and power of electronics used in the experiments. This thesis has explored new theoretical and numerical possibilities to study novel quantum phenomena; for example, using machine learning algorithms that allow the discovery of new phases of matter or a possibility to accelerate the transport of quantum particles. This dissertation also proposed an experimental scheme to study many-body physics.

7.1 Summary of the thesis

The first part of the dissertation illustrated that quantum observables, like the electronic energy of a molecule or the polaron dispersion energy, can be learned using supervised learning algorithms. The second part of the thesis introduced a novel approach to accelerate quantum walks in various graphs and the possibility to sim-

ulate extended Bose-Hubbard models using magnetic atoms trapped in an optical lattice.

All the ML algorithms that were introduced in this dissertation are *supervised* learning methods, where the only goal is to learn the map from an input to an output,

$$y_i = \mathcal{F}(\mathbf{x}_i), \quad (7.1)$$

where \mathbf{x}_i is the input and y_i the output. For example, in the case of PESs \mathbf{x}_i is the position of the nuclei and y_i is the electronic energy. In the case of the polaron Hamiltonians, \mathbf{x}_i is the parameters of the Hamiltonian and y_i is the dispersion energy. Here it was showed that given some training data, $\mathcal{F}(\cdot)$ can be approximated using GP models, where the only assumption made is the analytic form of the kernel function. Chapter 2 compared the accuracy of GP models with another famous supervised learning algorithm, Neural networks. As it is pointed out in Chapter 2, when the dimensionality of \mathbf{x}_i is low, GP models are a more accurate interpolation algorithm.

As was illustrated in this thesis, the main message of Chapter 3 is the use of GPmodels as an optimization algorithm. In quantum physics, there are numerous problems that can be formulated as an optimization problem. The goal for any optimization algorithm is to find the minimizer of a function,

$$\mathbf{x}_* = \arg \min_{\mathbf{x}} f(\mathbf{x}), \quad (7.2)$$

where the minimum of function $f(\cdot)$ is at \mathbf{x}_* . For example, the optimization of experimental setups to increase the trapping time of an ion in a surface-electrode trap [21, 166]. A great variety of optimization algorithms are gradient based methods. However, in quantum physics the function that describes the phenomena is not analytic and $\nabla f(\mathbf{x})$ can not be evaluated. On the other hand, the ability of GP models to accurately approximate $f(\mathbf{x})$ given some training data is what makes them a suitable tool for BO. Also, GP models can compute the uncertainty of the prediction, something simple NNs can not do. The uncertainty is key in BO since we can use it to explore the volume space of $f(\cdot)$ to find \mathbf{x}_* .

GPs are a non-parametric supervised method and the choice of the kernel function is fundamental. For interpolation problems, GP models can predict accurately using distance-based kernel functions as $e^{-(x_i-x_j)^2}$. Chapter 4 introduced an algorithm that finds more robust kernels by combining simple kernel functions, challenging the idea that one can quantitatively and qualitatively predict quantum observables beyond the original observation range. Furthermore, it is demonstrated that by extrapolating continuous quantum observables, like the polaron dispersion, one can discover new phases of matter.

The second part of this dissertation presented the possibility to enhance the diffusion of quantum walks and an experimental proposal to simulate Bose-Hubbard type models. Quantum walks are used in different algorithms due to their ability to spread faster than a random walk. Major efforts have been devoted to enhancing the linear time dependence of quantum walks diffusion, $\mathcal{O}(t)$. One possibility is to allow quantum walks to interact with an external environment [130, 148]. However, we decided to tackle the same problem by allowing the quantum walks to create/annihilate more walks in a graph. The possibility of making the number of initial particles be a non-conservative quantity made the quantum walk spread faster than the ballistic expansion at short times, $\mathcal{O}(t^n)$ where $n > 1$. We also showed that in the presence of disorder a quantum walk is less localized if particle number-changing interactions are considered. These conclusions are important since the diffusion and localization of quantum walks is key in quantum computing [47, 89, 171] and quantum transport [49, 140].

The Bose-Hubbard Hamiltonian was proposed to describe the electrons in a solid state in 1963 [77, 93]. It is practically impossible to computationally simulate quantum systems with a large number of particles; therefore, the ability to experimentally study them is key. Chapter 6 presented an experimental proposal to study Bose-Hubbard type Hamiltonians using magnetic atoms trapped in an optical lattice. This illustrates the possibility to tune various Bose-Hubbard type Hamiltonians by experimentally preparing magnetic atoms in different Zeeman states. Additionally, it was explained in Chapter 6, in the limit of low external magnetic field particle a Bose-Hubbard model with number-changing interactions can be described.

7.2 Future work

This section discusses some of the future possible research directions given the results presented in this dissertation.

ML has recently caught the attention in the fields of theoretical chemistry and physics. However, in the field of theoretical chemistry, the first paper where a NN was used to construct PESs was published in 2006 [124]. Furthermore, if one were to construct the PES of a big polymer using GP models is practically impossible due to the computational resources needed to invert the covariance matrix. As it is mentioned, one of the limitations of GPs is the scalability for large dimensional systems. Nevertheless, there have been new proposals in the ML field to improve the scalability of GPs models for large dimensional systems [59, 192]. The computational bottleneck for GPs’ prediction is the inversion of the covariance matrix, which scales as $\mathcal{O}(N^3)$, where N is the number of training points. In principle, GPs are dimensionally independent; however, the volume space increases as a function of the dimensions and so does the training data required to fill the volume. Thus, the need for robust kernel functions to learn the intrinsic similarity of the data it is crucial.

Another salient point introduced in this dissertation is the optimization of black-box type functions. Black-box type functions do not have an analytic form and only output a value given a specific input. Optimization of black-box type functions is an active research field in computer science and could also have a great impact in experimental physics. Chapter 3 explains how BO works to find the minimizer of black-box functions. The experimental challenges in ultracold physics can benefit by the use of BO type algorithms to search for better parameters that could increase, for example, the lifetime of quantum particles. In a current collaboration with Prof. Tobias Schaetz, from the University of Freiburg, we are exploring the use of BO to increase the trapping time of ions in surface-electrode traps by simulating the lifetime of an ion in a trap with different voltages. The assumption made here is that there exists a function $f(\mathbf{v})$ that maps a set of voltages, \mathbf{v} , applied to the trap to the trapping time of an ion. Thus, one can use BO to search for \mathbf{v}_* that will increase the trapping time of an ion.

In the field of chemistry, BO already has had a great impact. Recently, BO was

introduced in the field of experimental chemistry to improve the yield of chemistry reactions [82, 152]. Moreover, as explained in Appendix A, BO can optimize hybrid-exchange functionals for a benchmark of molecules. Most of the hybrid-exchange functionals are optimized by computing the square error between the predicted property of a molecule and the accurate one,

$$\mathcal{L} = \sqrt{\frac{1}{M} \sum_{i=1}^M (E_{DFT} - \hat{E})^2}, \quad (7.3)$$

where M is the total number of molecular properties, E_{DFT} is the predicted property using density functional theory (DFT) and \hat{E} is the accurate molecular property. These types of functions cannot be minimized with gradient descent based methods since the gradient of a molecular property given different values of exchange-correlation functional parameters, a_0, a_x and a_c do not exist. However, the error function \mathcal{L} does depend on the values of a_0, a_x , and a_c [105, 151], $\mathcal{L} = f(a_0, a_x, a_c)$. Thus, the minimizer of \mathcal{L} is the values of a_0, a_x , and a_c that best reproduce the molecular properties.

Chapter 4 introduced how GP models are capable of extrapolating continuous variables if the right kernel function is selected. This has a great implication since one can study quantum problems where the observables cannot be accessed due to convergence problems or lack of computational power. For example, it is illustrated that GP models can extrapolate the Holstein polaron dispersion for low phonon frequencies, where traditional condensed matter methods have a lack of convergence due to the dimensionality of the Hilbert space. This is not the only possibility, one can think of extrapolating quantum observables over time $\langle \hat{O}(t) \rangle$. For example, F. Häse *et al.* [81] used a deepNN to interpolate for different times the excitation energy transfer properties for different Fenna-Matthews-Olson (FMO) pigment-protein complex with eight chlorophyll pigments. Along these lines, one could use GP models with a combination of kernels to extrapolate the same time-dependent observables and reduce the computational cost, since the time-dependent calculations for longer time periods are computationally demanding.

The constructing procedure to combine simple kernels proposed by D. K. Duvenaud *et al.* in Ref. [60] is based on the maximization of the BIC. The BIC

allows the comparison of different kernel combinations without the need for test data, which makes it a powerful tool to extrapolate quantum observables. BIC is unbiased to models with a large number of parameters since it also penalizes kernels with a larger number of parameters to avoid memorization and enhance the learning. One of the problems with the methodology presented in Chapter 4 is the lack of criteria to stop the search for the optimal kernel combination. In theory, one can run the algorithm and create complex kernel forms with a large number of parameters but this does not guarantee a better extrapolation prediction. An interesting research direction is to understand why BIC prefers certain types of kernel combinations. This will allow the use of GPs to extrapolate quantum observables with a more physical intuition.

As Jeff Hawkins said: The key to artificial intelligence has always been the representation. The most remarkable point of this dissertation is the possibility for scientists to break free from the old mathematical analytics and study more complex or new problems, whose representation can be better captured and described with machine learning algorithms.

Bibliography

- [1] E. Abrahams, P. W. Anderson, D. C. Licciardello, and T. V. Ramakrishnan. Scaling theory of localization: Absence of quantum diffusion in two dimensions. *Phys. Rev. Lett.*, 42:673–676, Mar 1979. doi:10.1103/PhysRevLett.42.673. URL <https://link.aps.org/doi/10.1103/PhysRevLett.42.673>. → page 113
- [2] V. M. Agranovich. *Excitations in Organic Solids*. Oxford: Oxford University Press, 2009. → pages 80, 100
- [3] V. M. Agranovich and D. M. Basko. Frenkel excitons beyond the Heitler-London approximation. *The Journal of Chemical Physics*, 112(18): 8156–8162, 2000. → page 111
- [4] D. Aharonov, A. Ambainis, J. Kempe, and U. Vazirani. Quantum walks on graphs. In *Proceedings of the Thirty-third Annual ACM Symposium on Theory of Computing*, STOC '01, pages 50–59, New York, NY, USA, 2001. ACM. ISBN 1-58113-349-9. → pages 8, 78
- [5] K. Aikawa, A. Frisch, M. Mark, S. Baier, A. Rietzler, R. Grimm, and F. Ferlaino. Bose-Einstein condensation of erbium. *Phys. Rev. Lett.*, 108: 210401, May 2012. → page 98
- [6] Y. Akinaga and S. Ten-no. Range-separation by the Yukawa potential in long-range corrected density functional theory with Gaussian-type basis functions. *Chemical Physics Letters*, 462(4):348 – 351, 2008. ISSN 0009-2614. URL <http://www.sciencedirect.com/science/article/pii/S0009261408010609>. → page 147
- [7] P. W. Anderson. Absence of diffusion in certain random lattices. *Phys. Rev.*, 109:1492–1505, Mar 1958. → pages 85, 112

- [8] L.-F. Arsenault, O. Anatole von Lilienfeld, and A. J. Millis. Machine learning for many-body physics: efficient solution of dynamical mean-field theory. *ArXiv e-prints*, June 2015. → pages 1, 54, 74
- [9] A. Aspuru-Guzik and P. Walther. Photonic quantum simulators. *Nature Physics*, 8:285 EP –, Apr 2012. URL <http://dx.doi.org/10.1038/nphys2253>. Review Article. → page 78
- [10] R. Baer and D. Neuhauser. Density functional theory with correct long-range asymptotic behavior. *Phys. Rev. Lett.*, 94:043002, Feb 2005. doi:10.1103/PhysRevLett.94.043002. URL <https://link.aps.org/doi/10.1103/PhysRevLett.94.043002>. → page 147
- [11] S. Baier, M. J. Mark, D. Petter, K. Aikawa, L. Chomaz, Z. Cai, M. Baranov, P. Zoller, and F. Ferlaino. Extended Bose-Hubbard models with ultracold magnetic atoms. *Science*, 352(6282):201–205, 2016. → page 98
- [12] L. D. Bakalis and J. Knoester. Optical properties of one-dimensional exciton systems: Beyond the Heitler-London approximation. *The Journal of Chemical Physics*, 106(17):6964–6976, 1997. → page 111
- [13] A. P. Bartók and G. Csányi. Gaussian approximation potentials: A brief tutorial introduction. *International Journal of Quantum Chemistry*, 115(16):1051–1057, 2015. doi:10.1002/qua.24927. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.24927>. → page 25
- [14] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.*, 104:136403, Apr 2010. doi:10.1103/PhysRevLett.104.136403. URL <https://link.aps.org/doi/10.1103/PhysRevLett.104.136403>. → page 24
- [15] J. Behler. Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. *Phys. Chem. Chem. Phys.*, 13:17930–17955, 2011. doi:10.1039/C1CP21668F. URL <http://dx.doi.org/10.1039/C1CP21668F>. → page 25
- [16] M. Berciu. Green’s function of a dressed particle. *Phys. Rev. Lett.*, 97:036402, Jul 2006. doi:10.1103/PhysRevLett.97.036402. URL <https://link.aps.org/doi/10.1103/PhysRevLett.97.036402>. → pages 59, 69
- [17] A. J. Berkley, M. W. Johnson, P. Bunyk, R. Harris, J. Johansson, T. Lanting, E. Ladizinsky, E. Tolkacheva, M. H. S. Amin, and G. Rose. A scalable

readout system for a superconducting adiabatic quantum optimization system. *Superconductor Science and Technology*, 23(10):105014, 2010. URL <http://stacks.iop.org/0953-2048/23/i=10/a=105014>. → page 78

- [18] S. D. Berry and J. B. Wang. Two-particle quantum walks: Entanglement and graph isomorphism testing. *Phys. Rev. A*, 83:042317, Apr 2011. doi:10.1103/PhysRevA.83.042317. URL <https://link.aps.org/doi/10.1103/PhysRevA.83.042317>. → page 79
- [19] J. Billy, V. Josse, Z. Zuo, A. Bernard, B. Hambrecht, P. Lugan, D. Clément, L. Sanchez-Palencia, and A. Bouyer, Philippe and Aspect. Direct observation of Anderson localization of matter waves in a controlled disorder. *Nature*, 453:891 EP –, Jun 2008. URL <http://dx.doi.org/10.1038/nature07000>. → page 78
- [20] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006. ISBN 0387310738. → pages 11, 67
- [21] R. Blatt and C. F. Roos. Quantum simulations with trapped ions. *Nature Physics*, 8:277 EP –, Apr 2012. URL <http://dx.doi.org/10.1038/nphys2252>. Review Article. → pages 8, 117
- [22] J. Bonča, T. Kstrašnik, and S. A. Trugman. Mobile bipolaron. *Phys. Rev. Lett.*, 84:3153–3156, Apr 2000. doi:10.1103/PhysRevLett.84.3153. URL <https://link.aps.org/doi/10.1103/PhysRevLett.84.3153>. → pages xvii, 69, 70
- [23] A. I. Boothroyd, W. J. Keogh, P. G. Martin, and M. R. Peterson. A refined H3 potential energy surface. *The Journal of Chemical Physics*, 104(18): 7139–7152, 1996. → pages xiii, xiv, 24, 40, 41, 42, 44, 47
- [24] B. J. Braams and J. M. Bowman. Permutationally invariant potential energy surfaces in high dimensionality. *International Reviews in Physical Chemistry*, 28(4):577–606, 2009. → page 40
- [25] E. Brochu, V. M. Cora, and N. de Freitas. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. *ArXiv e-prints*, Dec. 2010. → pages 34, 37, 38, 51
- [26] F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller. Bypassing the Kohn-Sham equations with machine learning.

Nature Communications, 8(1):872, 2017. ISSN 2041-1723.
doi:10.1038/s41467-017-00839-3. URL
<https://doi.org/10.1038/s41467-017-00839-3>. → page 146

- [27] P. Bruno. Absence of spontaneous magnetic order at nonzero temperature in one- and two-dimensional Heisenberg and XY systems with long-range interactions. *Phys. Rev. Lett.*, 87:137203, Sep 2001. → page 98
- [28] A. Buarque and W. Dias. Unidirectional quantum walk of two correlated particles: Manipulating bound-pair and unbound wave-packet components. *Physics Letters A*, 381(37):3173 – 3177, 2017. ISSN 0375-9601.
doi:<https://doi.org/10.1016/j.physleta.2017.08.016>. URL
<http://www.sciencedirect.com/science/article/pii/S0375960117307417>. → page 79
- [29] A. Bühler and H. P. Büchler. Supersolid phase in atomic gases with magnetic dipole interaction. *Phys. Rev. A*, 84:023607, Aug 2011. → page 98
- [30] I. Buluta and F. Nori. Quantum simulators. *Science*, 326(5949):108–111, 2009. ISSN 0036-8075. doi:10.1126/science.1177838. URL
<http://science.sciencemag.org/content/326/5949/108>. → pages 1, 8
- [31] P. I. Bunyk, E. M. Hoskinson, M. W. Johnson, E. Tolkacheva, F. Altomare, A. J. Berkley, R. Harris, J. P. Hilton, T. Lanting, A. J. Przybysz, and J. Whittaker. Architectural considerations in the design of a superconducting quantum annealing processor. *IEEE Transactions on Applied Superconductivity*, 24(4):1–10, Aug 2014. ISSN 1051-8223.
doi:10.1109/TASC.2014.2318294. → page 78
- [32] S. Cantrill. Water behind walls. *Nature Chemistry*, 3:753 EP –, Sep 2011. URL <http://dx.doi.org/10.1038/nchem.1168>. → page 78
- [33] G. Carleo and M. Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017. ISSN 0036-8075. doi:10.1126/science.aag2302. URL
<http://science.sciencemag.org/content/355/6325/602>. → page 1
- [34] L. D. Carr, D. DeMille, R. V. Krems, and J. Ye. Cold and ultracold molecules: science, technology and applications. *New Journal of Physics*, 11(5):055049, 2009. URL
<http://stacks.iop.org/1367-2630/11/i=5/a=055049>. → page 79

- [35] J. Carrasquilla and R. G. Melko. Machine learning phases of matter. *Nature Physics*, 13:431 EP –, Feb 2017. URL <http://dx.doi.org/10.1038/nphys4035>. → pages 1, 54, 73
- [36] S. Carter, N. C. Handy, and J. Demaison. The rotational levels of the ground vibrational state of formaldehyde. *Molecular Physics*, 90(5): 729–738, 1997. → page 26
- [37] M. A. Cazalilla, R. Citro, T. Giamarchi, E. Orignac, and M. Rigol. One dimensional bosons: From condensed matter systems to ultracold gases. *Rev. Mod. Phys.*, 83:1405–1466, 2011. → page 104
- [38] C. M. Chandrashekar and T. Busch. Quantum walk on distinguishable non-interacting many-particles and indistinguishable two-particle. *Quantum Information Processing*, 11(5):1287–1299, Oct 2012. ISSN 1573-1332. doi:10.1007/s11128-012-0387-6. URL <https://doi.org/10.1007/s11128-012-0387-6>. → page 79
- [39] T. Chattaraj and R. V. Krems. Effects of long-range hopping and interactions on quantum walks in ordered and disordered lattices. *Phys. Rev. A*, 94:023601, Aug 2016. doi:10.1103/PhysRevA.94.023601. URL <https://link.aps.org/doi/10.1103/PhysRevA.94.023601>. → pages 78, 79, 83
- [40] R. H. Chaviguri, T. Comparin, V. S. Bagnato, and M. A. Caracanhas. Phase transition of ultracold atoms immersed in a Bose-Einstein-condensate vortex lattice. *Phys. Rev. A*, 95:053639, May 2017. doi:10.1103/PhysRevA.95.053639. URL <https://link.aps.org/doi/10.1103/PhysRevA.95.053639>. → page 79
- [41] L. Che, Z. Ren, X. Wang, W. Dong, D. Dai, X. Wang, D. H. Zhang, X. Yang, L. Sheng, G. Li, H.-J. Werner, F. Lique, and M. H. Alexander. Breakdown of the Born-Oppenheimer approximation in the $F + o\text{-D}_2 \rightarrow DF + D$ reaction. *Science*, 317(5841):1061–1064, 2007. → page 40
- [42] J. Chen, X. Xu, X. Xu, and D. H. Zhang. A global potential energy surface for the $H_2 + OH \rightarrow H_2O + H$ reaction using neural networks. *The Journal of Chemical Physics*, 138(15):154301, 2013. → pages xiv, 40, 41, 42, 45, 46
- [43] A. M. Childs. Universal computation by quantum walk. *Phys. Rev. Lett.*, 102:180501, May 2009. doi:10.1103/PhysRevLett.102.180501. URL <https://link.aps.org/doi/10.1103/PhysRevLett.102.180501>. → page 78

- [44] A. M. Childs and J. Goldstone. Spatial search and the Dirac equation. *Phys. Rev. A*, 70:042312, Oct 2004. doi:10.1103/PhysRevA.70.042312. URL <https://link.aps.org/doi/10.1103/PhysRevA.70.042312>. → page 78
- [45] A. M. Childs and J. Goldstone. Spatial search by quantum walk. *Phys. Rev. A*, 70:022314, Aug 2004. doi:10.1103/PhysRevA.70.022314. URL <https://link.aps.org/doi/10.1103/PhysRevA.70.022314>. → pages 8, 78
- [46] A. M. Childs, E. Farhi, and S. Gutmann. An example of the difference between quantum and classical random walks. *Quant. Info. Proc.*, 1(1): 35–43, Apr 2002. → pages 6, 77
- [47] A. M. Childs, R. Cleve, E. Deotto, E. Farhi, S. Gutmann, and D. A. Spielman. Exponential algorithmic speedup by a quantum walk. In *Proceedings of the Thirty-fifth Annual ACM Symposium on Theory of Computing*, STOC '03, pages 59–68. ACM, 2003. ISBN 1-58113-674-9. → pages 8, 78, 118
- [48] A. M. Childs, D. Gosset, and Z. Webb. Universal computation by multiparticle quantum walk. *Science*, 339(6121):791–794, 2013. ISSN 0036-8075. doi:10.1126/science.1229957. URL <http://science.sciencemag.org/content/339/6121/791>. → page 78
- [49] A. W. Chin, A. Datta, F. Caruso, S. F. Huelga, and M. B. Plenio. Noise-assisted energy transfer in quantum networks and light-harvesting complexes. *New Journal of Physics*, 12(6):065002, 2010. URL <http://stacks.iop.org/1367-2630/12/i=6/a=065002>. → pages 86, 118
- [50] Y.-S. Cho and R. J. L. Roy. Full empirical potential curves for the $X^1\Sigma^+$ and $A^1\Pi$ states of CH^+ from a direct-potential-fit analysis. *The Journal of Chemical Physics*, 144(2):024311, 2016. → page 48
- [51] M. A. Collins. Molecular potential-energy surfaces for chemical reaction dynamics. *Theoretical Chemistry Accounts*, 108(6):313–324, Dec 2002. → page 40
- [52] J. Cui and R. V. Krems. Efficient non-parametric fitting of potential energy surfaces for polyatomic molecules with Gaussian processes. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 49(22):224001, 2016. → page 25
- [53] A. Damianou and N. Lawrence. Deep Gaussian processes. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and*

Statistics, volume 31 of *Proceedings of Machine Learning Research*, pages 207–215. PMLR, 2013. → page 76

- [54] F. A. B. F. de Moura, A. V. Malyshev, M. L. Lyra, V. A. Malyshev, and F. Domínguez-Adame. Localization properties of a one-dimensional tight-binding model with nonrandom long-range intersite interactions. *Phys. Rev. B*, 71:174203, May 2005. doi:10.1103/PhysRevB.71.174203. URL <https://link.aps.org/doi/10.1103/PhysRevB.71.174203>. → page 78
- [55] W. Dong, C. Xiao, T. Wang, D. Dai, X. Yang, and D. H. Zhang. Transition-state spectroscopy of partial wave resonances in the F + HD reaction. *Science*, 327(5972):1501–1502, 2010. → page 40
- [56] S. Doniach and M. Inui. Long-range Coulomb interactions and the onset of superconductivity in the high- t_c materials. *Phys. Rev. B*, 41:6668–6678, Apr 1990. → page 98
- [57] W. Dür, R. Raussendorf, V. M. Kendon, and H.-J. Briegel. Quantum walks in optical lattices. *Phys. Rev. A*, 66:052319, Nov 2002. doi:10.1103/PhysRevA.66.052319. URL <https://link.aps.org/doi/10.1103/PhysRevA.66.052319>. → page 78
- [58] D. Duvenaud. *Automatic Model Construction with Gaussian Processes*. PhD thesis, Computational and Biological Learning Laboratory, University of Cambridge, 2014. → pages v, 57, 58, 67, 68
- [59] D. Duvenaud, H. Nickisch, and C. E. Rasmussen. Additive Gaussian processes. In *Advances in Neural Information Processing Systems 24*, pages 226–234, 2011. → page 119
- [60] D. Duvenaud, J. R. Lloyd, R. Grosse, J. B. Tenenbaum, and Z. Ghahramani. Structure discovery in nonparametric regression through compositional kernel search. In *Proceedings of the 30th International Conference on Machine Learning*, pages 1166–1174, 2013. → pages v, 57, 58, 68, 120
- [61] A. Ernesti and J. M. Hutson. Non-additive intermolecular forces from the spectroscopy of van der waals trimers: A comparison of Ar₂-HF and Ar₂-HCl, including H/D isotope effects. *The Journal of Chemical Physics*, 106(15):6288–6301, 1997. → pages 39, 48
- [62] F. Evers and A. D. Mirlin. Anderson transitions. *Rev. Mod. Phys.*, 80: 1355–1417, Oct 2008. → page 112

- [63] E. Farhi and S. Gutmann. Quantum computation and decision trees. *Phys. Rev. A*, 58:915–928, Aug 1998. → pages 8, 78
- [64] R. P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6):467–488, Jun 1982. ISSN 1572-9575. doi:10.1007/BF02650179. URL <https://doi.org/10.1007/BF02650179>. → pages 1, 8
- [65] A. Frisch, K. Aikawa, M. Mark, A. Rietzler, J. Schindler, Z. E., R. Grimm, and F. Ferlaino. Narrow-line magneto-optical trap for erbium. *Phys. Rev. A*, 85:051401, May 2012. → page 98
- [66] T. Fukuhara, P. Schauß, M. Endres, S. Hild, M. Cheneau, I. Bloch, and C. Gross. Microscopic observation of magnon bound states and their dynamics. *Nature*, 502:76 EP –, Sep 2013. URL <http://dx.doi.org/10.1038/nature12541>. → page 78
- [67] J. K. Gamble, M. Friesen, D. Zhou, R. Joynt, and S. N. Coppersmith. Two-particle quantum walks applied to the graph isomorphism problem. *Phys. Rev. A*, 81:052313, May 2010. doi:10.1103/PhysRevA.81.052313. URL <https://link.aps.org/doi/10.1103/PhysRevA.81.052313>. → page 78
- [68] B. T. Gard, R. M. Cross, P. M. Anisimov, H. Lee, and J. P. Dowling. Quantum random walks with multiphoton interference and high-order correlation functions. *J. Opt. Soc. Am. B*, 30(6):1538–1545, Jun 2013. doi:10.1364/JOSAB.30.001538. URL <http://josab.osa.org/abstract.cfm?URL=josab-30-6-1538>. → page 79
- [69] B. Gerlach and H. Löwen. Analytical properties of polaron systems or: Do polaronic phase transitions exist or not? *Rev. Mod. Phys.*, 63:63–90, Jan 1991. doi:10.1103/RevModPhys.63.63. URL <https://link.aps.org/doi/10.1103/RevModPhys.63.63>. → page 56
- [70] O. Giraud, B. Georgeot, and D. L. Shepelyansky. Quantum computing of delocalization in small-world networks. *Phys. Rev. E*, 72:036203, Sep 2005. doi:10.1103/PhysRevE.72.036203. URL <https://link.aps.org/doi/10.1103/PhysRevE.72.036203>. → page 78
- [71] R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, and A. Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Science*, 4(2):268–276, 2018. → pages 1, 75

- [72] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>. → pages 30, 76
- [73] A. V. Gorshkov, S. R. Manmana, G. Chen, J. Ye, E. Demler, M. D. Lukin, and A. M. Rey. Tunable superfluidity and quantum magnetism with ultracold polar molecules. *Phys. Rev. Lett.*, 107:115301, Sep 2011. → pages 9, 98
- [74] M. Greiner, O. Mandel, T. Esslinger, T. W. Hänsch, and I. Bloch. Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms. *Nature*, 415:39 EP –, Jan 2002. URL <http://dx.doi.org/10.1038/415039a>. Article. → pages 9, 98
- [75] A. Griesmaier, J. Werner, S. Hensler, J. Stuhler, and T. Pfau. Bose-Einstein condensation of chromium. *Phys. Rev. Lett.*, 94:160401, Apr 2005. doi:10.1103/PhysRevLett.94.160401. URL <https://link.aps.org/doi/10.1103/PhysRevLett.94.160401>. → pages 9, 98
- [76] S.-J. Gu, S.-S. Deng, Y.-Q. Li, and H.-Q. Lin. Entanglement and quantum phase transition in the extended Hubbard model. *Phys. Rev. Lett.*, 93: 086402, Aug 2004. doi:10.1103/PhysRevLett.93.086402. URL <https://link.aps.org/doi/10.1103/PhysRevLett.93.086402>. → page 104
- [77] M. C. Gutzwiller. Effect of correlation on the ferromagnetism of transition metals. *Phys. Rev. Lett.*, 10:159–162, Mar 1963. doi:10.1103/PhysRevLett.10.159. URL <https://link.aps.org/doi/10.1103/PhysRevLett.10.159>. → page 118
- [78] C. M. Handley and J. Behler. Next generation interatomic potentials for condensed systems. *The European Physical Journal B*, 87(7):152, Jul 2014. ISSN 1434-6036. doi:10.1140/epjb/e2014-50070-0. URL <https://doi.org/10.1140/epjb/e2014-50070-0>. → page 25
- [79] C. M. Handley and P. L. A. Popelier. Potential energy surfaces fitted by artificial neural networks. *The Journal of Physical Chemistry A*, 114(10): 3371–3383, 2010. → page 40
- [80] R. Harris, M. W. Johnson, T. Lanting, A. J. Berkley, J. Johansson, P. Bunyk, E. Tolkacheva, E. Ladizinsky, N. Ladizinsky, T. Oh, F. Cioata, I. Perminov, P. Spear, C. Enderud, C. Rich, S. Uchaikin, M. C. Thom, E. M. Chapple, J. Wang, B. Wilson, M. H. S. Amin, N. Dickson, K. Karimi, B. Macready, C. J. S. Truncik, and G. Rose. Experimental investigation of an eight-qubit unit cell in a superconducting optimization processor. *Phys. Rev. B*, 82:

024511, Jul 2010. doi:10.1103/PhysRevB.82.024511. URL
<https://link.aps.org/doi/10.1103/PhysRevB.82.024511>. → pages 78, 80

- [81] F. Häse, S. Valleau, E. Pyzer-Knapp, and A. Aspuru-Guzik. Machine learning exciton dynamics. *Chem. Sci.*, 7:5139–5147, 2016. doi:10.1039/C5SC04786B. URL <http://dx.doi.org/10.1039/C5SC04786B>. → page 120
- [82] F. Häse, L. M. Roch, and A. Aspuru-Guzik. Chimera: enabling hierarchy based multi-objective optimization for self-driving laboratories. *Chem. Sci.*, pages –, 2018. doi:10.1039/C8SC02239A. URL <http://dx.doi.org/10.1039/C8SC02239A>. → page 120
- [83] F. Häse, L. M. Roch, C. Kreisbeck, and A. Aspuru-Guzik. PHOENICS: A universal deep Bayesian optimizer. *ArXiv e-prints*, Jan. 2018. → page 51
- [84] K. R. A. Hazzard, B. Gadway, M. Foss-Feig, B. Yan, S. A. Moses, J. P. Covey, N. Y. Yao, M. D. Lukin, J. Ye, D. S. Jin, and A. M. Rey. Many-body dynamics of dipolar molecules in an optical lattice. *Phys. Rev. Lett.*, 113:195302, Nov 2014. doi:10.1103/PhysRevLett.113.195302. URL <https://link.aps.org/doi/10.1103/PhysRevLett.113.195302>. → pages 9, 79, 98
- [85] F. Herrera and R. V. Krems. Tunable Holstein model with cold polar molecules. *Phys. Rev. A*, 84:051401(R), Nov 2011. → pages 105, 106
- [86] F. Herrera, M. Litinskaya, and R. V. Krems. Tunable disorder in a crystal of cold polar molecules. *Phys. Rev. A*, 82:033428, Sep 2010. → pages 9, 98
- [87] F. Herrera, K. W. Madison, R. V. Krems, and M. Berciu. Investigating polaron transitions with polar molecules. *Phys. Rev. Lett.*, 110:223002, May 2013. doi:10.1103/PhysRevLett.110.223002. URL <https://link.aps.org/doi/10.1103/PhysRevLett.110.223002>. → pages xvi, xvii, xviii, 56, 58, 59, 62, 63, 65, 66, 72, 73, 74
- [88] F. Herrera, Y. Cao, S. Kais, and K. B. Whaley. Infrared-dressed entanglement of cold open-shell polar molecules for universal matchgate quantum computing. *New Journal of Physics*, 16(7):075001, 2014. → page 114
- [89] A. P. Hines and P. C. E. Stamp. Quantum walks, quantum gates, and quantum computers. *Phys. Rev. A*, 75:062321, Jun 2007. → pages 6, 77, 78, 118

- [90] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136: B864–B871, Nov 1964. doi:10.1103/PhysRev.136.B864. URL <https://link.aps.org/doi/10.1103/PhysRev.136.B864>. → page 145
- [91] T. Hollebeek, T.-S. Ho, and H. Rabitz. Constructing multidimensional molecular potential energy surfaces from ab initio data. *Annual Review of Physical Chemistry*, 50(1):537–570, 1999. → page 40
- [92] J. M. M. Howson and J. M. Hutson. Morphing the He-OCS intermolecular potential. *The Journal of Chemical Physics*, 115(11):5059–5065, 2001. → pages 39, 48
- [93] J. Hubbard. Electron correlations in narrow energy bands. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 276(1365):238–257, 1963. ISSN 0080-4630. doi:10.1098/rspa.1963.0204. URL <http://rspa.royalsocietypublishing.org/content/276/1365/238>. → page 118
- [94] F. Hutter, H. H. Hoos, and K. Leyton-Brown. Sequential model-based optimization for general algorithm configuration. In C. A. C. Coello, editor, *Learning and Intelligent Optimization*, pages 507–523, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. ISBN 978-3-642-25566-3. → page 51
- [95] H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao. A long-range correction scheme for generalized-gradient-approximation exchange functionals. *The Journal of Chemical Physics*, 115(8):3540–3544, 2001. doi:10.1063/1.1383587. URL <https://doi.org/10.1063/1.1383587>. → page 147
- [96] S. R. Jackson, T. J. Khoo, and F. W. Strauch. Quantum walks on trees with disorder: Decay, diffusion, and localization. *Phys. Rev. A*, 86:022335, Aug 2012. doi:10.1103/PhysRevA.86.022335. URL <https://link.aps.org/doi/10.1103/PhysRevA.86.022335>. → page 91
- [97] D. Jaksch, C. Bruder, J. I. Cirac, C. W. Gardiner, and P. Zoller. Cold bosonic atoms in optical lattices. *Phys. Rev. Lett.*, 81:3108–3111, Oct 1998. doi:10.1103/PhysRevLett.81.3108. URL <https://link.aps.org/doi/10.1103/PhysRevLett.81.3108>. → page 8
- [98] M. W. Johnson, P. Bunyk, F. Maibaum, E. Tolkacheva, A. J. Berkley, E. M. Chapple, R. Harris, J. Johansson, T. Lanting, I. Perminov, E. Ladizinsky, T. Oh, and G. Rose. A scalable control system for a superconducting adiabatic quantum optimization processor. *Superconductor Science and*

Technology, 23(6):065004, 2010. URL

<http://stacks.iop.org/0953-2048/23/i=6/a=065004>. → page 78

- [99] T. H. Johnson, Y. Yuan, W. Bao, S. R. Clark, C. Foot, and D. Jaksch. Hubbard model for atomic impurities bound by the vortex lattice of a rotating Bose-Einstein condensate. *Phys. Rev. Lett.*, 116:240402, Jun 2016. doi:10.1103/PhysRevLett.116.240402. URL <https://link.aps.org/doi/10.1103/PhysRevLett.116.240402>. → page 79
- [100] D. R. Jones. A taxonomy of global optimization methods based on response surfaces. *Journal of Global Optimization*, 21(4):345–383, Dec 2001. → page 37
- [101] G. S. Joyce. Absence of ferromagnetism or antiferromagnetism in the isotropic Heisenberg model with long-range interactions. *Journal of Physics C: Solid State Physics*, 2(8):1531, 1969. → page 98
- [102] A. Kamath, R. A. Vargas-Hernández, R. V. Krems, J. T. Carrington, and S. Manzhos. Neural networks vs Gaussian process regression for representing potential energy surfaces: A comparative study of fit quality and vibrational spectrum accuracy. *The Journal of Chemical Physics*, 148(24):241702, 2018. → pages xiii, 25, 26, 29, 39, 57
- [103] M. Karski, L. Förster, J.-M. Choi, A. Steffen, W. Alt, D. Meschede, and A. Widera. Quantum walk in position space with single optically trapped atoms. *Science*, 325(5937):174–177, 2009. ISSN 0036-8075. doi:10.1126/science.1174436. URL <http://science.sciencemag.org/content/325/5937/174>. → page 78
- [104] J. P. Keating, N. Linden, J. C. F. Matthews, and A. Winter. Localization and its consequences for quantum walk algorithms and quantum communication. *Phys. Rev. A*, 76:012315, Jul 2007. doi:10.1103/PhysRevA.76.012315. URL <https://link.aps.org/doi/10.1103/PhysRevA.76.012315>. → page 91
- [105] K. Kim and K. D. Jordan. Comparison of density functional and MP2 calculations on the water monomer and dimer. *The Journal of Physical Chemistry*, 98(40):10089–10094, 1994. doi:10.1021/j100091a024. URL <https://doi.org/10.1021/j100091a024>. → page 120
- [106] A. Kitaev. Anyons in an exactly solved model and beyond. *Annals of Physics*, 321(1):2 – 111, 2006. ISSN 0003-4916. doi:<https://doi.org/10.1016/j.aop.2005.10.005>. URL

<http://www.sciencedirect.com/science/article/pii/S0003491605002381>.
January Special Issue. → page 79

- [107] W. Koch and M. Holthausen. *A chemist's guide to density functional theory*. Wiley-VCH, 2000. ISBN 9783527299188. URL <https://books.google.es/books?id=rCApAAAAYAAJ>. → page 145
- [108] P. L. Krapivsky, J. M. Luck, and K. Mallick. Interacting quantum walkers: two-body bosonic and fermionic bound states. *Journal of Physics A: Mathematical and Theoretical*, 48(47):475301, 2015. URL <http://stacks.iop.org/1751-8121/48/i=47/a=475301>. → page 79
- [109] P. L. Krapivsky, J. M. Luck, and K. Mallick. Quantum centipedes: collective dynamics of interacting quantum walkers. *Journal of Physics A: Mathematical and Theoretical*, 49(33):335303, 2016. URL <http://stacks.iop.org/1751-8121/49/i=33/a=335303>. → page 79
- [110] K. A. Kuns, A. M. Rey, and A. V. Gorshkov. *d*-wave superfluidity in optical lattices of ultracold polar molecules. *Phys. Rev. A*, 84:063639, Dec 2011. → pages 9, 98
- [111] K. Kurotobi and Y. Murata. A single molecule of water encapsulated in fullerene C₆₀. *Science*, 333(6042):613–616, 2011. ISSN 0036-8075. doi:10.1126/science.1206376. URL <http://science.sciencemag.org/content/333/6042/613>. → page 78
- [112] H. Kushner. A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *J. Basic Eng.*, 86(1):97–106, 1964. → page 34
- [113] B. Lau, M. Berciu, and G. A. Sawatzky. Single-polaron properties of the one-dimensional breathing-mode Hamiltonian. *Phys. Rev. B*, 76:174305, Nov 2007. doi:10.1103/PhysRevB.76.174305. URL <https://link.aps.org/doi/10.1103/PhysRevB.76.174305>. → pages 56, 58
- [114] T. Leininger, H. Stoll, H.-J. Werner, and A. Savin. Combining long-range configuration interaction with short-range density functionals. *Chemical Physics Letters*, 275(3):151 – 160, 1997. ISSN 0009-2614. doi:[https://doi.org/10.1016/S0009-2614\(97\)00758-6](https://doi.org/10.1016/S0009-2614(97)00758-6). URL <http://www.sciencedirect.com/science/article/pii/S0009261497007586>. → page 147

- [115] M. Lewenstein, A. Sanpera, V. Ahufinger, B. Damski, A. Sen(De), and U. Sen. Ultracold atomic gases in optical lattices: mimicking condensed matter physics and beyond. *Advances in Physics*, 56(2):243–379, 2007. → pages 8, 97
- [116] D. Li, J. Zhang, F.-Z. Guo, W. Huang, Q.-Y. Wen, and H. Chen. Discrete-time interacting quantum walks and quantum hash schemes. *Quantum Information Processing*, 12(3):1501–1513, Mar 2013. ISSN 1573-1332. doi:10.1007/s11128-012-0421-8. URL <https://doi.org/10.1007/s11128-012-0421-8>. → page 79
- [117] E. H. Lieb and D. W. Robinson. The finite group velocity of quantum spin systems. *Comm. Math. Phys.*, 28(3):251–257, 1972. URL <https://projecteuclid.org:443/euclid.cmp/1103858407>. → page 78
- [118] D. E. Logan and P. G. Wolynes. Localizability and dephasing of dipolar excitons in topologically disordered systems. *The Journal of Chemical Physics*, 87(12):7199–7207, 1987. → page 113
- [119] M. Lu, N. Q. Burdick, S. H. Youn, and B. L. Lev. Strongly dipolar Bose-Einstein condensate of dysprosium. *Phys. Rev. Lett.*, 107:190401, Oct 2011. → page 98
- [120] M. Lu, N. Q. Burdick, and B. L. Lev. Quantum degenerate dipolar Fermi gas. *Phys. Rev. Lett.*, 108:215301, May 2012. → page 98
- [121] S. R. Manmana, E. M. Stoudenmire, K. R. A. Hazzard, A. M. Rey, and A. V. Gorshkov. Topological phases in ultracold polar-molecule quantum magnets. *Phys. Rev. B*, 87:081106, Feb 2013. → pages 9, 98
- [122] S. Manzhos and J. T. Carrington. Using neural networks to represent potential surfaces as sums of products. *The Journal of Chemical Physics*, 125(19):194105, 2006. → page 24
- [123] S. Manzhos and T. Carrington. Using an internal coordinate Gaussian basis and a space-fixed cartesian coordinate kinetic energy operator to compute a vibrational spectrum with rectangular collocation. *The Journal of Chemical Physics*, 145(22):224110, 2016. → page 28
- [124] S. Manzhos, X. Wang, R. Dawes, and T. Carrington. A nested molecule-independent neural network approach for high-quality potential fits. *The Journal of Physical Chemistry A*, 110(16):5295–5304, 2006. → pages 24, 119

- [125] D. J. J. Marchand, G. De Filippis, V. Cataudella, M. Berciu, N. Nagaosa, N. V. Prokof'ev, A. S. Mishchenko, and P. C. E. Stamp. Sharp transition for single polarons in the one-dimensional Su-Schrieffer-Heeger model. *Phys. Rev. Lett.*, 105:266605, Dec 2010. doi:10.1103/PhysRevLett.105.266605. URL <https://link.aps.org/doi/10.1103/PhysRevLett.105.266605>. → pages 56, 59
- [126] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979. → page 26
- [127] M. Meuwly and J. M. Hutson. Morphing ab initio potentials: A systematic study of Ne-HF. *The Journal of Chemical Physics*, 110(17):8338–8347, 1999. → pages 39, 48
- [128] A. Micheli, G. K. Brennen, and P. Zoller. A toolbox for lattice-spin models with polar molecules. *Nature Physics*, 2:341 EP –, Apr 2006. URL <http://dx.doi.org/10.1038/nphys287>. Article. → pages 9, 79, 98
- [129] R. Micnas, J. Ranninger, and S. Robaszkiewicz. Superconductivity in narrow-band systems with local nonretarded attractive interactions. *Rev. Mod. Phys.*, 62:113–171, Jan 1990. doi:10.1103/RevModPhys.62.113. URL <https://link.aps.org/doi/10.1103/RevModPhys.62.113>. → page 98
- [130] M. Mohseni, P. Rebentrost, S. Lloyd, and A. Aspuru-Guzik. Environment-assisted quantum walks in photosynthetic energy transfer. *The Journal of Chemical Physics*, 129(17):174106, 2008. doi:10.1063/1.3002335. URL <https://doi.org/10.1063/1.3002335>. → pages 78, 86, 118
- [131] S. A. Moses, J. P. Covey, M. T. Miecnikowski, B. Yan, B. Gadway, J. Ye, and D. S. Jin. Creation of a low-entropy quantum gas of polar molecules in an optical lattice. *Science*, 350(6261):659–662, 2015. ISSN 0036-8075. doi:10.1126/science.aac6400. URL <http://science.sciencemag.org/content/350/6261/659>. → page 9
- [132] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT press, 2012. ISBN 9780262018029. → pages 11, 14, 15, 25, 26, 32, 33, 54, 67, 75
- [133] J. N. Murrell, S. Carter, S. C. Farantos, P. Huxley, and A. J. C. Varandas. *Molecular Potential Energy Functions*. Wiley, New York, 1984. → pages 4, 24, 25, 40

- [134] R. M. Neal. *Bayesian Learning for Neural Networks*. Springer, New York, 1996. → pages 38, 76
- [135] M. Ortner, A. Micheli, G. Pupillo, and P. Zoller. Quantum simulations of extended Hubbard models with dipolar crystals. *New Journal of Physics*, 11(5):055045, 2009. → pages 9, 83, 98, 104
- [136] G. Parisi and B. Seoane. Liquid-glass transition in equilibrium. *Phys. Rev. E*, 89:022309, Feb 2014. doi:10.1103/PhysRevE.89.022309. URL <https://link.aps.org/doi/10.1103/PhysRevE.89.022309>. → page 78
- [137] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. → pages vi, 27
- [138] J. Pérez-Ríos, F. Herrera, and R. V. Krems. External field control of collective spin excitations in an optical lattice of $^2\Sigma$ molecules. *New Journal of Physics*, 12(10):103007, 2010. → pages 9, 98
- [139] A. Peruzzo, M. Lobino, J. C. F. Matthews, N. Matsuda, A. Politi, K. Poulios, X.-Q. Zhou, Y. Lahini, N. Ismail, K. Wörhoff, Y. Bromberg, Y. Silberberg, M. G. Thompson, and J. L. O'Brien. Quantum walks of correlated photons. *Science*, 329(5998):1500–1503, 2010. ISSN 0036-8075. doi:10.1126/science.1193515. URL <http://science.sciencemag.org/content/329/5998/1500>. → page 78
- [140] M. B. Plenio and S. F. Huelga. Dephasing-assisted transport: quantum networks and biomolecules. *New Journal of Physics*, 10(11):113019, 2008. URL <http://stacks.iop.org/1367-2630/10/i=11/a=113019>. → pages 86, 118
- [141] L. Pollet, J. D. Picon, H. P. Büchler, and M. Troyer. Supersolid phase with cold polar molecules on a triangular lattice. *Phys. Rev. Lett.*, 104:125302, Mar 2010. → pages 98, 104
- [142] R. Portugal. *Quantum Walks and Search Algorithms*. Quantum Science and Technology. Springer New York, 2013. ISBN 9781461463368. URL <https://books.google.ca/books?id=n5up9BYK1XwC>. → pages 6, 8
- [143] P. M. Preiss, R. Ma, M. E. Tai, A. Lukin, M. Rispoli, P. Zupancic, Y. Lahini, R. Islam, and M. Greiner. Strongly correlated quantum walks in optical lattices. *Science*, 347(6227):1229–1233, 2015. ISSN 0036-8075.

doi:10.1126/science.1260364. URL
<http://science.sciencemag.org/content/347/6227/1229>. → page 78

- [144] N. V. Prokof'ev and P. C. E. Stamp. Decoherence and quantum walks: Anomalous diffusion and ballistic tails. *Phys. Rev. A*, 74:020102, Aug 2006. doi:10.1103/PhysRevA.74.020102. URL <https://link.aps.org/doi/10.1103/PhysRevA.74.020102>. → page 78
- [145] P. Puddleiner and A. Mielke. Interacting bosons in two-dimensional flat band systems. *The European Physical Journal B*, 88(8):207, Aug 2015. ISSN 1434-6036. doi:10.1140/epjb/e2015-60371-3. URL <https://doi.org/10.1140/epjb/e2015-60371-3>. → page 79
- [146] M. Qiu, Z. Ren, L. Che, D. Dai, S. A. Harich, X. Wang, X. Yang, C. Xu, D. Xie, M. Gustafsson, R. T. Skodje, Z. Sun, and D. H. Zhang. Observation of feshbach resonances in the $F + H_2 \rightarrow HF + H$ reaction. *Science*, 311(5766):1440–1443, 2006. → page 40
- [147] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT press, 2006. ISBN 026218253X. → pages 11, 14, 15, 16, 56, 57
- [148] P. Rebentrost, M. Mohseni, I. Kassal, S. Lloyd, and A. Aspuru-Guzik. Environment-assisted quantum transport. *New Journal of Physics*, 11(3): 033003, 2009. URL <http://stacks.iop.org/1367-2630/11/i=3/a=033003>. → pages 78, 86, 118
- [149] P. Richerme, Z.-X. Gong, A. Lee, C. Senko, J. Smith, M. Foss-Feig, S. Michalakis, A. V. Gorshkov, and C. Monroe. Non-local propagation of correlations in quantum systems with long-range interactions. *Nature*, 511: 198 EP –, Jul 2014. URL <http://dx.doi.org/10.1038/nature13450>. → page 78
- [150] F. Robicheaux and N. M. Gill. Effect of random positions for coherent dipole transport. *Phys. Rev. A*, 89:053429, May 2014. → page 113
- [151] L. M. Roch and K. K. Baldridge. General optimization procedure towards the design of a new family of minimal parameter spin-component-scaled double-hybrid density functional theory. *Phys. Chem. Chem. Phys.*, 19: 26191–26200, 2017. doi:10.1039/C7CP04125J. URL <http://dx.doi.org/10.1039/C7CP04125J>. → page 120

- [152] L. M. Roch, F. Häse, C. Kreisbeck, T. Tamayo-Mendoza, L. P. E. Yunker, J. E. Hein, and A. Aspuru-Guzik. Chemos: An orchestration software to democratize autonomous discovery. *ChemRxiv e-prints*, march 2018,3. → pages 51, 120
- [153] A. Rodríguez, V. A. Malyshev, G. Sierra, M. A. Martín-Delgado, J. Rodríguez-Laguna, and F. Domínguez-Adame. Anderson transition in low-dimensional disordered systems driven by long-range nonrandom hopping. *Phys. Rev. Lett.*, 90:027404, Jan 2003. doi:10.1103/PhysRevLett.90.027404. URL <https://link.aps.org/doi/10.1103/PhysRevLett.90.027404>. → page 78
- [154] J. P. J. Rodriguez, Z. J. Li, and J. B. Wang. Discord and entanglement of two-particle quantum walk on cycle graphs. *Quantum Information Processing*, 14(1):119–133, Jan 2015. ISSN 1573-1332. doi:10.1007/s11128-014-0859-y. URL <https://doi.org/10.1007/s11128-014-0859-y>. → page 79
- [155] P. P. Rohde, A. Schreiber, M. Štefaňák, I. Jex, and C. Silberhorn. Multi-walker discrete time quantum walks on arbitrary graphs, their properties and their photonic implementation. *New Journal of Physics*, 13(1):013001, 2011. URL <http://stacks.iop.org/1367-2630/13/i=1/a=013001>. → page 79
- [156] L. J. Root and J. L. Skinner. Localization phase diagram for the energetically and substitutionally disordered anderson/quantum percolation model. *The Journal of Chemical Physics*, 89(5):3279–3284, 1988. doi:10.1063/1.454933. URL <https://doi.org/10.1063/1.454933>. → page 95
- [157] K. Rudinger, J. K. Gamble, M. Wellons, E. Bach, M. Friesen, R. Joynt, and S. N. Coppersmith. Noninteracting multiparticle quantum random walks applied to the graph isomorphism problem for strongly regular graphs. *Phys. Rev. A*, 86:022334, Aug 2012. doi:10.1103/PhysRevA.86.022334. URL <https://link.aps.org/doi/10.1103/PhysRevA.86.022334>. → page 79
- [158] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533, 1986. → page 33
- [159] M. Rupp. Machine learning for quantum mechanics in a nutshell. *International Journal of Quantum Chemistry*, 115(16):1058–1073, 2015. doi:10.1002/qua.24954. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.24954>. → page 25

- [160] S. Sachdev. *Quantum Phase Transitions*. Cambridge University Press, 2 edition, 2011. → pages 4, 55
- [161] J. Sakurai and S. Tuan. *Modern Quantum Mechanics*. Addison-Wesley, 1994. ISBN 9780201537307. URL <https://books.google.ca/books?id=Nol-AQAACAAJ>. → page 2
- [162] L. Sansoni, F. Sciarrino, G. Vallone, P. Mataloni, A. Crespi, R. Ramponi, and R. Osellame. Two-particle Bosonic-Fermionic quantum walk via integrated photonics. *Phys. Rev. Lett.*, 108:010502, Jan 2012. doi:10.1103/PhysRevLett.108.010502. URL <https://link.aps.org/doi/10.1103/PhysRevLett.108.010502>. → page 79
- [163] A. Savin. *Beyond the Kohn-Sham Determinant*, pages 129–153. World Scientific, Singapore, 1995. doi:10.1142/9789812830586_0004. URL https://www.worldscientific.com/doi/abs/10.1142/9789812830586_0004. → page 147
- [164] V. W. Scarola and S. Das Sarma. Quantum phases of the extended Bose-Hubbard Hamiltonian: Possibility of a supersolid state of cold atoms in optical lattices. *Phys. Rev. Lett.*, 95:033003, Jul 2005. doi:10.1103/PhysRevLett.95.033003. URL <https://link.aps.org/doi/10.1103/PhysRevLett.95.033003>. → pages 83, 98
- [165] G. C. Schatz. The analytical representation of electronic potential-energy surfaces. *Rev. Mod. Phys.*, 61:669–688, 1989. → pages 24, 25
- [166] C. Schneider, D. Porras, and T. Schaetz. Experimental quantum simulations of many-body physics with trapped ions. *Reports on Progress in Physics*, 75(2):024401, 2012. URL <http://stacks.iop.org/0034-4885/75/i=2/a=024401>. → pages 8, 117
- [167] J. R. Schrieffer and P. A. Wolff. Relation between the Anderson and Kondo Hamiltonians. *Phys. Rev.*, 149:491–492, Sep 1966. doi:10.1103/PhysRev.149.491. URL <https://link.aps.org/doi/10.1103/PhysRev.149.491>. → page 80
- [168] G. Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6(2): 461–464, 03 1978. doi:10.1214/aos/1176344136. URL <https://doi.org/10.1214/aos/1176344136>. → page 67
- [169] A. Seko, A. Takahashi, and I. Tanaka. Sparse representation for a potential energy surface. *Phys. Rev. B*, 90:024101, Jul 2014.

doi:10.1103/PhysRevB.90.024101. URL
<https://link.aps.org/doi/10.1103/PhysRevB.90.024101>. → page 25

- [170] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, Jan 2016. → pages 34, 38, 51
- [171] N. Shenvi, J. Kempe, and K. B. Whaley. Quantum random-walk search algorithm. *Phys. Rev. A*, 67:052307, May 2003. → pages 8, 78, 118
- [172] J. Snoek, O. Rippel, K. Swersky, R. Kiros, N. Satish, N. Sundaram, M. M. A. Patwary, P. Prabhat, and R. P. Adams. Scalable bayesian optimization using deep neural networks. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37, ICML’15*, pages 2171–2180, 2015. → pages 38, 51
- [173] P. Soltan-Panahi, D.-S. Lühmann, J. Struck, P. Windpassinger, and K. Sengstock. Quantum phase transition to unconventional multi-orbital superfluidity in optical lattices. *Nature Physics*, 8:71 EP –, Oct 2011. → page 98
- [174] J. Sous, M. Berciu, and R. V. Krems. Bipolarons bound by repulsive phonon-mediated interactions. *Phys. Rev. A*, 96:063619, Dec 2017. doi:10.1103/PhysRevA.96.063619. URL
<https://link.aps.org/doi/10.1103/PhysRevA.96.063619>. → page 59
- [175] J. Sous, M. Chakraborty, C. P. J. Adolphs, R. V. Krems, and M. Berciu. Phonon-mediated repulsion, sharp transitions and (quasi)self-trapping in the extended Peierls-Hubbard model. *Scientific Reports*, 7(1):1169, 2017. → pages 59, 98
- [176] T. Sowiński, O. Dutta, P. Hauke, L. Tagliacozzo, and M. Lewenstein. Dipolar molecules in optical lattices. *Phys. Rev. Lett.*, 108:115301, Mar 2012. doi:10.1103/PhysRevLett.108.115301. URL
<https://link.aps.org/doi/10.1103/PhysRevLett.108.115301>. → page 78
- [177] S. Sra, S. Nowozin, and S. J. Wright. *Optimization for Machine Learning*. The MIT Press, 2011. → pages 31, 32
- [178] N. Srinivas, A. Krause, S. M. Kakade, and M. W. Seeger. Information-theoretic regret bounds for Gaussian process optimization in the bandit setting. *IEEE Transactions on Information Theory*, 58(5): 3250–3265, 2012. → pages 38, 51

- [179] M. Štefaňák, S. M. Barnett, B. Kollr, T. Kiss, and I. Jex. Directional correlations in quantum walks with two particles. *New Journal of Physics*, 13(3):033029, 2011. URL <http://stacks.iop.org/1367-2630/13/i=3/a=033029>. → page 79
- [180] N. Q. Su, J. Chen, Z. Sun, D. H. Zhang, and X. Xu. H + H₂ quantum dynamics using potential energy surfaces based on the XYG3 type of doubly hybrid density functionals: Validation of the density functionals. *The Journal of Chemical Physics*, 142(8):084107, 2015. → pages 40, 42
- [181] R. S. Sutton and A. G. Barto. *Reinforcement Learning, An Introduction*. MIT Press, 1 edition, 2016. → pages 11, 43, 51, 68
- [182] K. Swersky, J. Snoek, and R. P. Adams. Multi-task bayesian optimization. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems* 26, pages 2004–2012. Curran Associates, Inc., 2013. URL <http://papers.nips.cc/paper/5086-multi-task-bayesian-optimization.pdf>. → page 51
- [183] D. G. Truhlar, R. Steckler, and M. S. Gordon. Potential energy surfaces for polyatomic reaction dynamics. *Chemical Reviews*, 87(1):217–236, 1987. → pages 24, 25
- [184] R. A. Vargas-Herández and R. V. Krems. Engineering extended Hubbard models with Zeeman excitations of ultracold Dy atoms. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 49(23):235501, 2016. URL <http://stacks.iop.org/0953-4075/49/i=23/a=235501>. → pages xxi, xxii, xxiii, 78, 83, 97, 103, 107, 108, 109, 110, 115
- [185] R. A. Vargas-Hernández and R. V. Krems. Quantum walks assisted by particle-number fluctuations. *Phys. Rev. A*, 98:022107, Aug 2018. doi:10.1103/PhysRevA.98.022107. URL <https://link.aps.org/doi/10.1103/PhysRevA.98.022107>. → pages xviii, xix, xx, xxi, 77, 82, 84, 85, 87, 89, 92, 94, 96
- [186] A. Vázquez-Sánchez and J. G. Ávila-Zárraga. A formal synthesis of (+/-) parvifoline by an aromatic Cope rearrangement of a trans-1-aryl-2-ethenylcyclobutanecarbonitrile. *Tetrahedron Letters*, 58(10): 981 – 984, 2017. → page 25
- [187] S. E. Venegas-Andraca. Quantum walks: a comprehensive review. *Quant. Info. Proc.*, 11(5):1015–1106, Oct 2012. → pages 6, 77

- [188] L. Wang. Discovering phase transitions with unsupervised learning. *Phys. Rev. B*, 94:195105, Nov 2016. doi:10.1103/PhysRevB.94.195105. URL <https://link.aps.org/doi/10.1103/PhysRevB.94.195105>. → page 74
- [189] Q. Wang and Z.-J. Li. Repelling, binding, and oscillating of two-particle discrete-time quantum walks. *Annals of Physics*, 373:1 – 9, 2016. ISSN 0003-4916. doi:<https://doi.org/10.1016/j.aop.2016.06.015>. URL <http://www.sciencedirect.com/science/article/pii/S0003491616300914>. → page 79
- [190] T. Wang, J. Chen, T. Yang, C. Xiao, Z. Sun, L. Huang, D. Dai, X. Yang, and D. H. Zhang. Dynamical resonances accessible only by reagent vibrational excitation in the $F + HD \rightarrow HF + D$ reaction. *Science*, 342 (6165):1499–1502, 2013. → page 40
- [191] A. Wilson and R. Adams. Gaussian process kernels for pattern discovery and extrapolation. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 1067–1075. PMLR, 2013. → page 76
- [192] A. G. Wilson, E. Gilboa, A. Nehorai, and J. P. Cunningham. Fast kernel learning for multidimensional pattern extrapolation. In *Advances in Neural Information Processing Systems 27*, pages 3626–3634. Curran Associates, Inc., 2014. → pages 76, 119
- [193] R. Winkle. Quasi-degenerate perturbation theory. In R. Winkler, editor, *Spin–Orbit Coupling Effects in Two-Dimensional Electron and Hole Systems*, Springer Tracts in Modern Physics, chapter B, page 201. Springer-Verlag Berlin Heidelberg, Springer, Berlin, Heidelberg, 2003. → page 149
- [194] P. Wittek. *Quantum Machine Learning*. Elsevier, Boston, 2014. ISBN 978-0-12-800953-6. doi:<https://doi.org/10.1016/B978-0-12-800953-6.00015-3>. URL <http://www.sciencedirect.com/science/article/pii/B9780128009536000153>. → page 5
- [195] P. Würtz, T. Langen, T. Gericke, A. Koglbauer, and H. Ott. Experimental demonstration of single-site addressability in a two-dimensional optical lattice. *Phys. Rev. Lett.*, 103:080404, Aug 2009. → pages 9, 98
- [196] P. Xiang. *Quantum control of dynamics of quasiparticles in periodic and disordered lattice potentials*. PhD thesis, University of British Columbia,

2014. URL <https://open.library.ubc.ca/collections/24/items/1.0165942>.
Text. → page 100
- [197] P. Xiang, M. Litinskaya, and R. V. Krems. Tunable exciton interactions in optical lattices with polar molecules. *Phys. Rev. A*, 85:061401(R), Jun 2012. → page 106
- [198] P. Xiang, M. Litinskaya, E. A. Shapiro, and R. V. Krems. Non-adiabatic control of quantum energy transfer in ordered and disordered arrays. *New Journal of Physics*, 15(6):063015, 2013. → pages 9, 98
- [199] T. Xu and R. V. Krems. Quantum walk and Anderson localization of rotational excitations in disordered ensembles of polar molecules. *New Journal of Physics*, 17(6):065014, 2015. URL <http://stacks.iop.org/1367-2630/17/i=6/a=065014>. → page 80
- [200] T. Xu and R. V. Krems. Quantum walk and Anderson localization of rotational excitations in disordered ensembles of polar molecules. *New Journal of Physics*, 17(6):065014, 2015. → page 113
- [201] B. Yan, S. A. Moses, B. Gadway, J. P. Covey, K. R. A. Hazzard, A. M. Rey, D. S. Jin, and J. Ye. Observation of dipolar spin-exchange interactions with lattice-confined polar molecules. *Nature*, 501:521 EP –, Sep 2013. URL <http://dx.doi.org/10.1038/nature12483>. → pages 9, 79, 98
- [202] J. Zaanen. Quantum phase transitions in cuprates: stripes and antiferromagnetic supersolids. *Physica C: Superconductivity*, 317-318:217 – 229, 1999. → page 98
- [203] R. Zare. *Angular momentum: understanding spatial aspects in chemistry and physics*. George Fisher Baker non-resident lectureship in chemistry at Cornell University. Wiley, 1988. ISBN 9780471858928. URL <https://books.google.ca/books?id=kDQ31MBZxioC>. → pages 152, 156

Appendix A

Bayesian optimization for hybrid-density functionals

Electronic structure methods developed by quantum chemists are designed to evaluate the ground state energy, E_{GS} , of a given ensemble of electrons and fixed position nuclei,

$$E_{GS} [\hat{V}, N] = \min_{\Psi \rightarrow N} [\langle \Psi | \hat{H} | \Psi \rangle], \quad (\text{A.1})$$

where N is the number of electrons, \hat{V} is the potential energy that electrons feel by the fixed position nuclei and Ψ is the N -electrons wave function. The Hamiltonian, \hat{H} , describe the total energy $\hat{H} = \hat{T} + \hat{U} + \hat{V}$, where first and second terms are the kinetic energy operator and electron-electron interaction potential. The minimization of E_{GS} is still an active research problem due to the complexity of the problem.

In 1964 a new methodology, known as Density functional theory (DFT), was introduced to evaluate Equation A.1 [90]. DFT assumes that for N -electrons produced by \hat{V} there is a single electronic density, $n(\mathbf{r})$, [107]. Given \hat{V} and $n(\mathbf{r})$ one can compute E_{GS} as,

$$E_{GS} [\hat{V}, N] = \min_{n \rightarrow N} \left[F[n(\mathbf{r})] + \int n(\mathbf{r}) v(\mathbf{r}) d\mathbf{r}^3 \right], \quad (\text{A.2})$$

where $F[n(\mathbf{r})]$ is the universal functional and the second term is the nuclear attraction potential energy function for a electron density $n(\mathbf{r})$ given fixed position nuclei, also denoted as $V_{Ne}[n(\mathbf{r})]$. More condensed notation can be used to rewrite $E_{GS}[\hat{V}, N]$,

$$E_{GS}[\hat{V}, N] = F[n(\mathbf{r})] + V_{Ne}[n(\mathbf{r})] = T[n(\mathbf{r})] + V_{ee}[n(\mathbf{r})] + V_{Ne}[n(\mathbf{r})], \quad (\text{A.3})$$

where $T[\cdot]$ is the kinetic energy functional and $V_{ee}[\cdot]$ is the electron-electron interaction functional. The form of $F[n(\mathbf{r})]$ is still unknown and is still an open research problem; recently $F[\cdot]$ was constructed using ML [26]. The functional $V_{ee}[\cdot]$ plays a key role in DFT since it described the electron-electron interaction, also denoted as $E_{XC}[n(\mathbf{r})]$. Over the last decades, DFT researchers have proposed more accurate descriptors of $E_{XC}[n(\mathbf{r})]$; one of the most successful ones is,

$$E_{XC} = a_0 E_X^{HF} + a_1 E_X^{PBE} + a_2 E_C^{PBE}, \quad (\text{A.4})$$

where E_X^{HF} is the Hartree-Fock exact exchange functional, E_X^{PBE} is the PBE exchange functional and E_C^{PBE} is the PBE correlation functional. This type of functionals are known as *hybrid exchange-correlation* functionals, the trade off between the three terms in hybrid E_{XC} is described by the a_i constants. \bar{a} denotes collectively the parameters of E_{XC} . It is out of the scope of this dissertation to give a more complete explanation about DFT.

From Equations A.3 and A.4, one can observe that the accuracy in DFT's prediction directly depend on the values of \bar{a} . In most of the cases, the values of \bar{a} are set by minimizing an error-type function, where E_{GS} is compared with an accurate value, \hat{E}_{GS} , for a given set of molecules or atoms.

For illustrative purposes, this chapter shows that DFT parameters can be tuned using BO. A similar type of exchange-correlation functional is considered here, range-separated exchange functionals where V_{ee} is segmented in short and long-

range interactions [6, 10, 95, 114, 163],

$$V_{ee} \propto \frac{1}{r_{i,j}} = SR + LR = \frac{\omega(\gamma, r_{i,j})}{r_{i,j}} + \frac{1 - \omega(\gamma, r_{i,j})}{r_{i,j}} \quad (\text{A.5})$$

where $r_{i,j}$ is the distance between two electrons, ω is a function that depends on $r_{i,j}$ and a parameter, denoted by γ . The idea behind the ω function is to switch from short-range (SR) to long-range (LR) interactions to achieve a better description. Given an analytical form for ω , the only free parameter in E_{XC} is γ . In the case of range-separated exchange functionals, γ is tuned by minimizing the a loss function,

$$\mathcal{L}(\gamma|\mathcal{M}) = \sum_i^{|\mathcal{M}|} |\epsilon_i^{HOMO}(\gamma) + IP_i| \quad (\text{A.6})$$

where the summation is over different molecules, $\epsilon_i^{HOMO}(\gamma)$ is the energy of the highest occupied molecular orbital (HOMO) for molecule i computed with a given value of γ and IP_i is the ionization potential of molecule i . $|\mathcal{M}|$ is the total number of molecules or atoms considered. Equation A.6 is an optimization problem under the eyes of ML,

$$\gamma_* = \arg \min_{\gamma} \mathcal{L}(\gamma|\mathcal{M}), \quad (\text{A.7})$$

where γ_* is the minimizer of \mathcal{L} that best describes the IP of each \mathcal{M}_i . Equation A.6 cannot be minimized using gradient descent methods; however, as we explained in Chapter 3, low dimensional black-box type functions can be optimized using BO. The example considered here is the optimization of γ for hydrogen molecule which IP is 16.2 eV. The value of the optimal $\hat{\gamma}$ is 1.2 for an *LCY* – *PBE* functional. For illustrative purposes a *UCB* acquisition function was used, Equation 3.9, to find the minimizer of \mathcal{L} for the hydrogen molecule. Figure A.1 illustrates the first fourth iterations of BO. The first two points were sampled randomly and at the fifth iteration $\gamma = 1.24$.

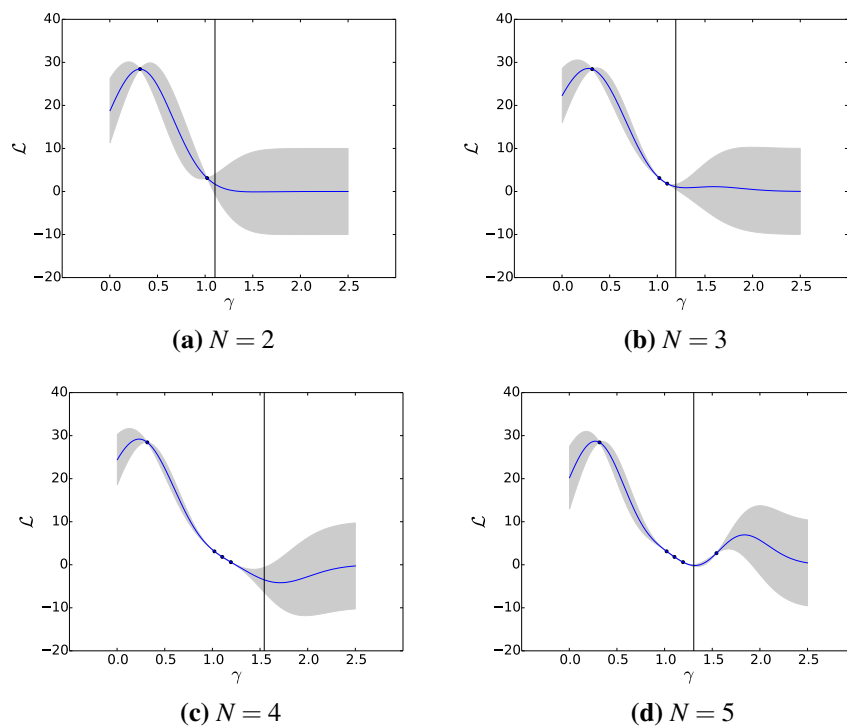


Figure A.1: The iterations of BO to find the minimizer of \mathcal{L} for the H_2 molecule. The black vertical line is the next proposed point by BO. The blue markers are the training data used to construct the acquisition function. The blue-solid curve is the mean of the GP model and the grey shaded area is $\sigma(\cdot)$.

Appendix B

Schrieffer–Wolff transformation

Here, we show that the particle-changing interactions in Eq. (5.2), if perturbative, modify the range of particle hopping. In particular, we show that, to leading order, the couplings (5.2) lead to next-nearest-neighbour hopping.

It is clear that the Hamiltonian (5.1) is not block diagonal in the site representation basis due to couplings in Eq. (5.2). Using the Schrieffer–Wolff (SW) transformation, it is possible to block diagonalize this Hamiltonian. We follow the notation in Ref. [193]. The total Hamiltonian is defined as $\hat{\mathcal{H}} = \hat{\mathcal{H}}_0 + \hat{\mathcal{H}}'$, where $\hat{\mathcal{H}}_0$ contains all the operators that commute with the particle number operator, and $\hat{\mathcal{H}}' = \hat{V}_{\text{nc}}$. The SW transformation assumes that the transformed Hamiltonian

$$\tilde{\mathcal{H}} = e^{-S} \hat{\mathcal{H}} e^S \quad (\text{B.1})$$

can be written as

$$\tilde{\mathcal{H}} = H^{(0)} + H^{(1)} + H^{(2)} + \dots, \quad (\text{B.2})$$

with the different terms defined by the following matrix elements:

$$H_{mm'}^{(0)} = H_{mm'}^0 \quad (\text{B.3})$$

$$H_{mm'}^{(1)} = H'_{mm'} \quad (\text{B.4})$$

$$H_{mm'}^{(2)} = \frac{1}{2} \sum_{\ell} \left[\frac{1}{E_m - E_{\ell}} + \frac{1}{E'_m - E_{\ell}} \right] H'_{m\ell} H'_{\ell m'}. \quad (\text{B.5})$$

Here, the indices m and m' refer to any one-particle state, while ℓ is an index of a three-particle state. H^0 is the Hamiltonian (5.1) with $\hat{V}_{\text{nc}} = 0$. All matrix elements of $H^{(1)}$ are zero: $H_{mm'}^{(1)} = 0$. The first correction to \mathcal{H} appears in $H^{(2)}$ whose matrix elements depend on \hat{V}_{nc} . Here we only consider the case of a 1D lattice with $\gamma = 0$, $\Delta \neq 0$, and nearest-neighbour interactions. The matrix elements $H_{mm'}^{(2)}$ depend on the matrix elements of the Δ -dependent term in \hat{V}_{nc} ,

$$\begin{aligned} H'_{m\ell} &= \langle m | \Delta \sum_i (\hat{c}_i^\dagger \hat{c}_{i\pm 1}^\dagger + \hat{c}_i \hat{c}_{\pm 1}) | abc \rangle = \Delta \sum_i \langle m | \hat{c}_i \hat{c}_{i\pm 1} | abc \rangle \\ &= \Delta [\delta_{m,c} (\delta_{b\pm 1,a} + \delta_{a\pm 1,b}) + \delta_{m,b} (\delta_{c\pm 1,a} + \delta_{a\pm 1,c}) \\ &\quad + \delta_{m,a} (\delta_{c\pm 1,b} + \delta_{b\pm 1,c})] \end{aligned} \quad (\text{B.6})$$

where a , b and c are the lattice indices of the three particles. In the case of an ideal system, Eq. (B.5) can be rewritten as,

$$H_{mm'}^{(2)} = -\frac{1}{2\Delta\epsilon} \sum_{\ell} H'_{m\ell} H'_{\ell m'} \quad (\text{B.7})$$

where the summation \sum_{ℓ} is over all possible combinations of lattice indices for three particles. Inserting Eq. (B.6) into Eq. (B.5) for both $H'_{m\ell}$ and $H'_{\ell m'}$ we obtain,

$$\begin{aligned} H_{mm'}^{(2)} &= -\frac{1}{2\Delta\epsilon} \sum_{\ell} \Delta^2 [\delta_{m,c} (\delta_{b\pm 1,a} + \delta_{a\pm 1,b}) + \delta_{m,b} (\delta_{c\pm 1,a} + \delta_{a\pm 1,c}) \\ &\quad + \delta_{m,a} (\delta_{c\pm 1,b} + \delta_{b\pm 1,c})] \times [\delta_{c,m'} (\delta_{a,b\pm 1} + \delta_{b,a\pm 1}) \\ &\quad + \delta_{b,m'} (\delta_{a,c\pm 1} + \delta_{c,a\pm 1}) + \delta_{a,m'} (\delta_{b,c\pm 1} + \delta_{c,b\pm 1})] \end{aligned} \quad (\text{B.8})$$

The diagonal elements of $H^{(2)}$ are

$$H_{mm}^{(2)} = -\frac{\Delta^2}{\Delta\epsilon} m^*, \quad (\text{B.9})$$

where m^* is the total number of nearest-neighbour lattice sites without considering

the site m . The value of m^* can be computed as,

$$m^* = \sum_{i=0}^{m-1} \sum_{j=i+1}^{m-1} \delta_{i,j\pm 1} + \sum_{i=m+1}^N \sum_{j=i+1}^N \delta_{i,j\pm 1}, \quad (\text{B.10})$$

where the first summation is over all pairs of lattice site interactions for $i < m$, and the second summation is for $i > m$. For example, when m is any of a 1D lattice site edges, $m^* = N - 2$, where N is the total number of lattice sites. In the case when $m \neq m'$,

$$H_{mm'}^{(2)} = -\frac{3\Delta^2}{2\Delta\epsilon} [\delta_{m,m'\pm 2} + \delta_{m\pm 2,m'}], \quad (\text{B.11})$$

which leads to next-nearest-neighbour hopping with $t' = -\frac{3\Delta^2}{2\Delta\epsilon}$. Combining Eqs. (B.9) and (B.11) we see that the SW transformation, to first order, leads to the following one-particle Hamiltonian:

$$\tilde{\mathcal{H}} = \sum_i \omega'_i \hat{c}_i^\dagger \hat{c}_i + \sum_i t \left(\hat{c}_{i\pm 1}^\dagger \hat{c}_i \right) + t' \left(\hat{c}_{i\pm 2}^\dagger \hat{c}_i \right), \quad (\text{B.12})$$

where $\omega'_i = \Delta\epsilon - \frac{\Delta^2}{\Delta\epsilon} m^*$ and $t' = -\frac{3\Delta^2}{2\Delta\epsilon}$.

Appendix C

Magnetic dipole–dipole interaction

The potential energy between two dipoles \mathbf{d}_1 and \mathbf{d}_2 is

$$\hat{V}_{dd} = \frac{C_d}{4\pi} \frac{\mathbf{d}_1 \cdot \mathbf{d}_2 - 3(\mathbf{d}_1 \cdot \mathbf{r})(\mathbf{d}_2 \cdot \mathbf{r})}{r_{12}^3}, \quad (\text{C.1})$$

where \mathbf{d}_i is any of the dipoles, and \mathbf{r} is a unit vector in the direction of \mathbf{r}_{12} . \mathbf{r}_{12} is the distance between the two dipoles. As it is described in Chapter 6, the magnetic dipole-dipole interaction between two atoms can be used to tune various Hubbard models. Here we derive the matrix elements relation to compute of the magnetic dipole-dipole interaction between atoms with different Zeeman states.

In *spherical tensor* algebra, the dot product is defined as [203]

$$\begin{aligned} \mathbf{a} \cdot \mathbf{b} &= -\sqrt{3} \left[\mathbf{a}^{(1)} \otimes \mathbf{b}^{(1)} \right]_0^{(0)} \\ &= -\sqrt{3} [-a_+ b_- + a_0 b_0 - a_- b_+], \end{aligned} \quad (\text{C.2})$$

where $\left[\mathbf{a}^{(1)} \otimes \mathbf{b}^{(1)} \right]_0^{(0)}$ is the tensor product between vectors \mathbf{a} and \mathbf{b} , and $a_{\pm} = \mp \frac{1}{\sqrt{2}} (a_x \pm ia_y)$ and $a_0 = a_z$. A more general expression of a tensor product is,

$$\left[T^{(1)} \otimes T^{(1)} \right]_q^{(k)} = \sum_m \langle 1m, 1q-m | kq \rangle T(1, m) T(1, q-m), \quad (\text{C.3})$$

where $\langle 1m, 1q - m | kq \rangle$ are the Clebsch-Gordan (CG) coefficients, and $T(k, q)$ are spherical tensors of rank k . Any vector is a first rank spherical tensor, $T^{(1)}$, where $k = 1$. Equations C.2 and C.3 are identical when $k = 0$ and $q = 0$, and we can deduce the value of the CG coefficients.

The second term of \hat{V}_{dd} , Equation C.1, can be rewritten using the following identity,

$$\begin{aligned} (\mathbf{a} \cdot \mathbf{b})(\mathbf{c} \cdot \mathbf{d}) &= 3 \left[\left[\mathbf{a}^{(1)} \otimes \mathbf{b}^{(1)} \right]_0^{(0)} \otimes \left[\mathbf{c}^{(1)} \otimes \mathbf{d}^{(1)} \right]_0^{(0)} \right]_0^{(0)} \\ &= \frac{1}{3} (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) + \frac{1}{2} (\mathbf{a} \times \mathbf{b})(\mathbf{c} \times \mathbf{d}) \\ &\quad + \sum_{q=-2}^2 (-1)^{2-q} \left[a^{(1)} \otimes c^{(1)} \right]_q^{(2)} \left[b^{(1)} \otimes d^{(1)} \right]_q^{(2)}. \end{aligned} \quad (\text{C.4})$$

Combining Equations C.3 and C.4 we can rewrite \hat{V}_{dd} as,

$$\hat{V}_{dd} = \frac{C_d}{4\pi} \sum_{p=-2}^2 (-1)^{2-q} \left[r^{(1)} \otimes r^{(1)} \right]_p^{(2)} \left[d_1^{(1)} \otimes d_2^{(1)} \right]_{-p}^{(2)}, \quad (\text{C.5})$$

where the first term $\left[r^{(1)} \otimes r^{(1)} \right]_p^{(2)}$ is a second rank spherical tensor, $T^{(2)}(\mathbf{r})$, and the second term is the tensor product between the two dipoles.

Let us describe the term $\left[r^{(1)} \otimes r^{(1)} \right]_p^{(2)}$ when $p = 0$ and using Equation C.3 we find that,

$$\left[r^{(1)} \otimes r^{(1)} \right]_0^{(2)} = \sum_{p_1, p_2} \langle 1 p_1, 1 p_2 | 20 \rangle T_{p_1}^1(r) T_{p_2}^1(r) \quad (\text{C.6})$$

$$\begin{aligned} &= \langle 1 1, 1 -1 | 20 \rangle T_1^1(r) T_{-1}^1(r) \\ &\quad + \langle 1 -1, 1 1 | 20 \rangle T_{-1}^1(r) T_1^1(r) + \langle 1 0, 1 0 | 20 \rangle T_0^1(r) T_0^1(r) \\ &= -\frac{1}{\sqrt{6}} (x^2 + y^2) + \sqrt{\frac{2}{3}} z^2, \end{aligned} \quad (\text{C.7})$$

where we use $T_{\pm 1}^1 = a_{\pm 1}$ and $T_0^1 = a_0$, previously defined. We observe that Equation C.7 is proportional to the spherical harmonic $Y_0^2 = \sqrt{\frac{15}{16\pi}} \frac{2z^2 - x^2 - y^2}{r^2}$. Multiplying Equation C.7 by $\sqrt{6}$ we found that $\left[r^{(1)} \otimes r^{(1)} \right]_0^{(2)}$ is the C_0^2 *modified spherical*

harmonic,

$$\left[r^{(1)} \otimes r^{(1)} \right]_0^{(2)} = -\sqrt{6} T_p^2(C) = -\frac{\sqrt{6}}{r^3} C_p^2(\theta, \phi), \quad (\text{C.8})$$

where $C_q^k = \sqrt{\frac{4\pi}{2k+1}} Y_q^k$. Here we list some of the spherical harmonics for $k = 2$,

m	Y_m^2	C_m^2
0	$(5/16\pi)^{1/2}(3\cos^2\theta - 1)$	$\frac{1}{2}(3\cos^2\theta - 1)$
± 1	$\mp(15/8\pi)^{1/2}\cos\theta\sin\theta e^{\pm i\phi}$	$\mp(3/2)^{1/2}\cos\theta\sin\theta e^{\pm i\phi}$
± 2	$(15/32\pi)^{1/2}\sin^2\theta e^{\pm 2i\phi}$	$(3/8)^{1/2}\sin^2\theta e^{\pm 2i\phi}$

Using Equations C.5 and C.8, \hat{V}_{dd} is

$$\begin{aligned} \hat{V}_{dd} &= -\sqrt{6} \frac{C_d}{4\pi} \sum_{p=-2}^2 (-1)^{2-p} T_p^2(C) T_{-p}^2(d_1, d_2) \\ &= -\sqrt{6} \frac{C_d}{4\pi} \sum_{p=-2}^2 (-1)^{2-p} \frac{C_p^2(\theta, \phi)}{r^3} \sum_{p_1, p_2} \langle 1 \ p_1, 1 \ p_2 | 2 - p \rangle T_{p_1}^1(d_1) T_{p_2}^1(d_2) \quad (\text{C.9}) \end{aligned}$$

which fully expanded is,

$$\begin{aligned} \hat{V}_{dd} &= -\frac{C_d}{4\pi r^3} \left\{ \frac{\sqrt{6}}{2} (3\cos^2\theta - 1) \left(\langle 1 \ 1, 1 \ -1 | 2 \ 0 \rangle T_1^1(d_1) T_{-1}^1(d_2) + \langle 1 \ -1, 1 \ 1 | 2 \ 0 \rangle T_{-1}^1(d_1) T_1^1(d_2) \right. \right. \\ &\quad \left. \left. + \langle 1 \ 0, 1 \ 0 | 2 \ 0 \rangle T_0^1(d_1) T_0^1(d_2) \right) \right\}_{p=0} \\ &\quad - \frac{C_d}{4\pi r^3} \left\{ \sqrt{6} (3/2)^{1/2} \cos\theta \sin\theta e^{\pm i\phi} \left(\langle 1 \ -1, 1 \ 0 | 2 \ -1 \rangle T_{-1}^1(d_1) T_0^1(d_2) \right. \right. \\ &\quad \left. \left. + \langle 1 \ 0, 1 \ -1 | 2 \ -1 \rangle T_0^1(d_1) T_{-1}^1(d_2) \right) \right\}_{p=1} \\ &\quad + \frac{C_d}{4\pi r^3} \left\{ \sqrt{6} (3/2)^{1/2} \cos\theta \sin\theta e^{\pm i\phi} \left(\langle 1 \ 1, 1 \ 0 | 2 \ 1 \rangle T_1^1(d_1) T_0^1(d_2) \right. \right. \\ &\quad \left. \left. + \langle 1 \ 0, 1 \ 1 | 2 \ 1 \rangle T_0^1(d_1) T_1^1(d_2) \right) \right\}_{p=-1} \\ &\quad - \frac{C_d}{4\pi r^3} \left\{ \sqrt{6} (3/8)^{1/2} \sin^2\theta e^{\pm 2i\phi} \left(\langle 1 \ -1, 1 \ -1 | 2 \ -2 \rangle T_{-1}^1(d_1) T_{-1}^1(d_2) \right) \right\}_{p=2} \\ &\quad - \frac{C_d}{4\pi r^3} \left\{ \sqrt{6} (3/8)^{1/2} \sin^2\theta e^{\pm 2i\phi} \left(\langle 1 \ 1, 1 \ 1 | 2 \ 2 \rangle T_1^1(d_1) T_1^1(d_2) \right) \right\}_{p=-2}, \quad (\text{C.10}) \end{aligned}$$

when $\theta = 90^\circ$, all the terms except when $p = 0$ vanish leading to a z -axis parallel to the internuclear vector that connects both dipoles. Under this assumption, the *dipole-dipole* equation is,

$$\begin{aligned}
\hat{V}_{dd} &= -\frac{C_d}{4\pi r^3} \left\{ 2\sqrt{6} \left(\langle 1\ 1, 1\ -1 | 2\ 0 \rangle T_1^1(d_1) T_{-1}^1(d_2) + \langle 1\ -1, 1\ 1 | 2\ 0 \rangle T_{-1}^1(d_1) T_1^1(d_2) \right. \right. \\
&\quad \left. \left. + \langle 1\ 0, 1\ 0 | 2\ 0 \rangle T_0^1(d_1) T_0^1(d_2) \right) \right\} \\
&= -\frac{C_d}{4\pi r^3} \left\{ 2\sqrt{6} \left(\sqrt{\frac{2}{3}} T_0^1(d_1) T_0^1(d_2) + \sqrt{\frac{1}{6}} T_{-1}^1(d_1) T_1^1(d_2) + \sqrt{\frac{1}{6}} T_1^1(d_1) T_{-1}^1(d_2) \right) \right\} \\
&= -\frac{C_d}{4\pi r^3} \{ 2T_0^1(d_1) T_0^1(d_2) + T_{-1}^1(d_1) T_1^1(d_2) + T_1^1(d_1) T_{-1}^1(d_2) \}. \tag{C.11}
\end{aligned}$$

C.0.1 Matrix element

The total magnetic dipole operator looks like $\mathbf{d}_J = \mu_0 g_L \mathbf{L} + \mu_0 g_S \mathbf{S}$ where $\mu_0 g_L \mathbf{L}$ is the magnetic dipole moment from the electron orbital angular momentum and $\mu_0 g_S \mathbf{S}$ is the magnetic dipole moment from the electron spin angular momentum. Here we describe how to compute the matrix elements of \hat{V}_{dd} for two magnetic dipoles in the total angular momentum basis set, $|\psi_i\rangle = |\eta_i; J_i M_i\rangle$, where, the total angular momentum is $\mathbf{J} = \mathbf{L} + \mathbf{S}$. η denotes any further quantum number required to characterize the state,

$$\begin{aligned}
\langle \psi'_i | \langle \psi'_j | \hat{V}_{dd} | \psi_i \rangle | \psi_j \rangle &= \langle \eta'_i; J'_i M'_i | \langle \eta'_j; J'_j M'_j | \hat{V}_{dd} | \eta_i; J_i M_i \rangle | \eta_j; J_j M_j \rangle \\
&= -\frac{C_d}{4\pi r^3} \left\{ \langle \eta'_i; J'_i M'_i | \langle \eta'_j; J'_j M'_j | T_1^1(d_1) T_{-1}^1(d_2) | \eta_i; J_i M_i \rangle | \eta_j; J_j M_j \rangle \right. \\
&\quad \left. + \langle \eta'_i; J'_i M'_i | \langle \eta'_j; J'_j M'_j | T_{-1}^1(d_1) T_1^1(d_2) | \eta_i; J_i M_i \rangle | \eta_j; J_j M_j \rangle \right\} \\
&\quad - 2\frac{C_d}{4\pi r^3} \langle \eta'_i; J'_i M'_i | \langle \eta'_j; J'_j M'_j | T_0^1(d_1) T_0^1(d_2) | \eta_i; J_i M_i \rangle | \eta_j; J_j M_j \rangle \\
&\tag{C.12}
\end{aligned}$$

The computation of $\langle \psi'_i | \langle \psi'_j | \hat{V}_{dd} | \psi_i \rangle | \psi_j \rangle$ only depends on the values of $\langle J'_i M'_i | T_{p_1}^1(d) | J_i M_i \rangle$ because,

$$\begin{aligned} \langle \eta'_i; J'_i M'_i | \langle \eta'_j; J'_j M'_j | T_0^1(d_1) T_0^1(d_2) | \eta_i; J_i M_i \rangle | \eta_j; J_j M_j \rangle &= \langle \eta'_i; J'_i M'_i | T_0^1(d_1) | \eta_i; J_i M_i \rangle \\ &\times \langle \eta'_j; J'_j M'_j | T_0^1(d_2) | \eta_j; J_j M_j \rangle \quad (\text{C.13}) \end{aligned}$$

where $\langle \eta'_j; J'_j M'_j | T_0^1(d_2) | \eta_j; J_j M_j \rangle$ can be computed using the Wigner-Eckart theorem [203],

$$\langle \eta'; J' M' | T_p^k(\mathbf{A}) | \eta; J M \rangle = (-1)^{J'-M'} \begin{pmatrix} J' & k & J \\ -M' & p & M \end{pmatrix} \langle \eta'; J' || T^k(\mathbf{A}) || \eta; J \rangle \quad (\text{C.14})$$

In the case of magnetic dipoles, $T^k(\mathbf{d}_i)$ has no physical dependence on the quantum number η , and the reduced density matrix $\langle \eta'; J' || T^k(\mathbf{A}) || \eta; J \rangle$ is equal to $\delta_{\eta', \eta} \delta_{J', J} [J(J+1)(2J+1)]^{1/2}$. The matrix dipole element is,

$$\begin{aligned} \langle J' M' | T_{p_1}^1(d) | J M \rangle &= (-1)^{J'-M'} \begin{pmatrix} J' & 1 & J \\ -M' & p_1 & M \end{pmatrix} \langle J' || T^k(\mathbf{d}) || J \rangle \\ &= (-1)^{J'-M'} \begin{pmatrix} J' & 1 & J \\ -M' & p_1 & M \end{pmatrix} [J(J+1)(2J+1)]^{1/2} \delta_{J', J} \\ &= (-1)^{-(M'+M+1)} \delta_{J', J} (2J+1)^{-1/2} \langle J' - M'; 1 p_1 | J - M \rangle [J(J+1)(2J+1)]^{1/2} \end{aligned} \quad (\text{C.15})$$

from the above equation, we can deduce the first selection rule, **the magnetic dipole-dipole interaction only couples states with the same J** . To derive the rest of the selection rules we first present the value of the $3j$ symbols of the CG coefficients,

$$\begin{pmatrix} J & 1 & J \\ -M & 0 & M \end{pmatrix} = (-1)^{J-M} \frac{M}{[J(J+1)(2J+1)]^{1/2}}. \quad (\text{C.16})$$

Using the above equation we can fully compute $\langle J' M' | T_p^1(d) | J M \rangle = (-1)^{J'-M'}$.

The first case we consider is for $p = 0$,

$$\begin{aligned}
\langle \eta'_i; J'_i M'_i | T_0^1(d_1) | \eta_i; J_i M_i \rangle &= (-1)^{J'_i - M'_i} \begin{pmatrix} J'_i & 1 & J_i \\ -M'_i & 0 & M_i \end{pmatrix} \langle J_i || T^1(d_1) || J_i \rangle \delta_{\eta'_i, \eta_i} \\
&= (-1)^{J_i - M_i} \begin{pmatrix} J_i & 1 & J_i \\ -M_i & 0 & M_i \end{pmatrix} [J_i(J_i + 1)(2J_i + 1)]^{1/2} \\
&= (-1)^{2J'_i - 2M'_i} \frac{M}{[J_i(J_i + 1)(2J_i + 1)]^{1/2}} [J_i(J_i + 1)(2J_i + 1)]^{1/2} \\
&= M \delta_{\eta'_i, \eta_i} \tag{C.17}
\end{aligned}$$

therefore,

$$\langle \eta'_i; J'_i M'_i | \langle \eta'_j; J'_j M'_j | T_0^1(d_1) T_0^1(d_2) | \eta_i; J_i M_i \rangle | \eta_j; J_j M_j \rangle = M_i \times M_j \delta_{\eta'_i, \eta_i} \delta_{\eta'_j, \eta_j} \tag{C.18}$$

The other two terms left in equation C.12 have the same form,

$$\begin{aligned}
\langle \eta'_i; J'_i M'_i | \langle \eta'_j; J'_j M'_j | T_{\pm 1}^1(d_1) T_{\mp 1}^1(d_2) | \eta_i; J_i M_i \rangle | \eta_j; J_j M_j \rangle &= \langle \eta'_i; J'_i M'_i | T_{\pm 1}^1(d_1) | \eta_i; J_i M_i \rangle \\
&\quad \times \langle \eta'_j; J'_j M'_j | T_{\mp 1}^1(d_2) | \eta_j; J_j M_j \rangle \tag{C.19}
\end{aligned}$$

and its value depends on the 3j symbol when $j_2 = 1$ and $m_2 = \pm 1$,

$$\begin{aligned}
\begin{pmatrix} J & 1 & J \\ -M \mp 1 & \pm 1 & M \end{pmatrix} &= \pm (-1)^{J-M} \left[\frac{(J \mp M)(J \pm M + 1)}{2J(J+1)(2J+1)} \right]^{1/2} \\
&= \pm (-1)^{J-M} \left[\frac{J(J+1) - M(M \pm 1)}{2J(J+1)(2J+1)} \right]^{1/2} \tag{C.20}
\end{aligned}$$

Inserting Equation C.20 into any of the rand-hand side terms in Equation C.19

we obtain,

$$\begin{aligned}
\langle \eta'_i; J'_i M'_i | T_{\pm 1}^1(d_1) | \eta_i; J_i M_i \rangle &= (-1)^{J'_i - M'_i} \begin{pmatrix} J'_i & 1 & J_i \\ -M'_i & \pm 1 & M_i \end{pmatrix} \langle J_i \| T^1(d_1) \| J_i \rangle \delta_{\eta'_i, \eta_i} \\
&= (-1)^{J_i - M'_i} \begin{pmatrix} J_i & 1 & J_i \\ -M'_i & \pm 1 & M_i \end{pmatrix} [J_i(J_i + 1)(2J_i + 1)]^{1/2} \\
&= \pm (-1)^{J_i - M_i + 1} (-1)^{J_i - M_i} \left[\frac{J_i(J_i + 1) - M_i(M_i \pm 1)}{2J_i(J_i + 1)(2J_i + 1)} \right]^{1/2} [J_i(J_i + 1)(2J_i + 1)]^{1/2} \\
&= \mp \frac{1}{\sqrt{2}} [J_i(J_i + 1) - M_i(M_i \pm 1)]^{1/2} \delta_{M'_i, M_i \pm 1}
\end{aligned} \tag{C.21}$$

$$\begin{aligned}
\langle \eta'_i; J'_i M'_i | \langle \eta'_j; J'_j M'_j | T_1^1(d_1) T_{-1}^1(d_2) | \eta_i; J_i M_i \rangle | \eta_j; J_j M_j \rangle &= -\frac{1}{2} [J_i(J_i + 1) - M_i(M_i \pm 1)]^{1/2} \\
&\quad \times [J_j(J_j + 1) - M_j(M_j \mp 1)]^{1/2}
\end{aligned} \tag{C.22}$$

where the second selection rule appears, **the magnetic dipole-dipole interaction couples states with $|M - M'| = 1$.**

Inserting Equations C.18 and C.22 into Equation C.12, we obtaine,

$$\begin{aligned}
\langle \hat{V}_{dd} \rangle &= -\frac{C_d}{4\pi r^3} \left\{ -\frac{1}{2} [J_i(J_i + 1) - M_i(M_i + 1)]^{1/2} [J_j(J_j + 1) - M_j(M_j - 1)]^{1/2} \right\} \delta_{M'_i, M_i + 1} \delta_{M'_j, M_j - 1} \\
&\quad -\frac{C_d}{4\pi r^3} \left\{ -\frac{1}{2} [J_i(J_i + 1) - M_i(M_i - 1)]^{1/2} [J_j(J_j + 1) - M_j(M_j + 1)]^{1/2} \right\} \delta_{M'_i, M_i - 1} \delta_{M'_j, M_j + 1} \\
&\quad -2\frac{C_d}{4\pi r^3} \{ M_i \times M_j \} \delta_{M'_i, M_i} \delta_{M'_j, M_j}
\end{aligned} \tag{C.23}$$

Equation C.23 can be rewritten into a more compact form by remembering that $[J(J + 1) - M(M \pm 1)]^{1/2}$ looks like the eigenvalue of the *raising* and *lowering* operator \hat{J}_{\pm} , and M is the eigenvalue of the \hat{J}_{z_i} ,

$$\hat{V}_{dd} = \frac{C_d}{4\pi r^3} \left\{ \frac{1}{2} [\hat{J}_{i,-} \hat{J}_{j,+} + \hat{J}_{i,+} \hat{J}_{j,-}] - 2\hat{J}_{i,z} \hat{J}_{j,z} \right\} \tag{C.24}$$