

Dynamical processes in the condensed phase: methods and models

Matthew R. Carbone

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2021

© 2021

Matthew R. Carbone

All Rights Reserved

Abstract

Dynamical processes in the condensed phase: methods and models

Matthew R. Carbone

In this thesis, we study a broad range of physical phenomena from the perspectives of theory-driven, and machine learning models.

We begin by introducing a generalization of the Momentum Average method for finding numerically exact Green's functions of arbitrary polaron systems at zero and finite temperature. This method utilizes the physical ansatz that phonons are produced largely in clouds, and systematically constructs a closure of auxiliary Green's functions to ultimately solve for the spectrum. We seamlessly apply this method to a variety of problems, including the Holstein, Peierls, and mixed-boson mode models. Next, we leverage fundamental quantum mechanics to develop a microscopic model of exciton and trion scattering in monolayer transition metal dichalcogenides. We conclude that elastic scattering mechanisms are largely the dominant contributor, and confirm that our calculated doping-dependent linewidths qualitatively agree with experiment. In addition, we use Monte Carlo dynamics to examine entropically activated dynamics in continuous phase space models, and show that global and local dynamics both exhibit entropy-driven activation.

The second type of work discussed in this thesis pertains to data-driven machine learning models. These approaches offer the utility of instantaneous inference, which has tremendous potential application in applied science in areas such as surrogate modeling and creating digital twins of expensive experiments. First, we demonstrate that x-ray absorption spectra can be used to classify absorbing sites' local atomic information, specifically its coordination number. Next, we show that graph-based neural networks can to quantitative accuracy, predict the x-ray absorption spectrum of small molecules in the QM9 database. We highlight the various ways in which these types of methodologies can be applied to e.g. closing the design loop and surrogate modeling in general.

Table of Contents

List of Tables	vi
List of Figures	viii
Acknowledgements	xiv
Chapter 1: Introduction	1
1.1 Theory- vs. data-driven models	2
1.2 Quantum theory	3
1.3 Machine learning	4
1.4 Outline	5
I Quantum dynamics in the condensed phase	7
Chapter 2: Numerically Exact Generalized Green's Function Cluster Expansions for Electron-Phonon Problems	8
2.1 Introduction	9
2.2 Methodology and General Considerations	11
2.2.1 A Generalized Equation of Motion	13
2.2.2 Implementation	17
2.2.3 Comparison to Other Methods	19

2.3	Results	22
2.3.1	Holstein Model	22
2.3.2	Peierls Model	26
2.3.3	Mixed-Boson Mode Holstein+Peierls Model	26
2.4	Conclusions	28
2.A	Reduction Rules for AGFs	29
2.B	Examples of the Generalized Notation	30
2.C	Additional Notation for Mixed-Boson Mode HP Models	31
2.D	Finite-temperature Thermofield Dynamics	32
2.D.1	The Thermofield Double	34
2.D.2	The separable density matrix initial condition	35
2.D.3	Showing a unitary form is possible	37
2.E	The thermal Green's function	41
2.E.1	Managing the time-ordering operator	43
2.E.2	The transformed Hamiltonian	44
2.E.3	The free particle propagator	46
2.F	Outlook and future plans	46
Chapter 3:	Microscopic model of the doping dependence of line widths in monolayer transition metal dichalcogenides	48
3.1	Introduction	49
3.2	Methodology	51
3.2.1	Electron-exciton elastic scattering	51
3.2.2	Electron-trion scattering	53

3.2.3	Dielectric Function	55
3.3	Results and Discussion	56
3.4	Conclusion	60
3.A	RPA Polarizability	61
3.B	Exciton-electron elastic scattering	62
3.B.1	General form of the matrix elements	62
3.B.2	Spin contributions	64
3.B.3	1s-1s scattering	65
3.B.4	2s-2s scattering	66
3.C	Trion-electron elastic scattering	66
3.D	Computational Details	71
3.E	Comparison to experiment	72
II	Machine learning for chem-informatics application	74
Chapter 4:	Classification of local chemical environments from x-ray absorption spectra using supervised machine mearning	75
4.1	Introduction	76
4.2	Methods	79
4.2.1	Data acquisition	79
4.2.2	Local chemical environment class labeling	82
4.2.3	Training machine learning models	84
4.3	Results and discussion	85
4.3.1	Visual inspection of the spectral database	85
4.3.2	Principal component analysis of the spectral database	87

4.3.3	Machine learning classifier performance	88
4.3.4	Importance of features beyond the pre-edge	90
4.3.5	Discussion	93
4.4	Conclusion	97
Chapter 5: Machine-learning x-ray absorption spectra to quantitative accuracy		99
5.1	Introduction	100
5.2	Dataset construction	102
5.2.1	Preliminary analysis	103
5.3	Results and discussion	107
5.4	Concluding remarks	111
5.A	Machine learning details	112
5.A.1	Graph feature generation	112
5.A.2	Technical details of the Message Passing Neural Network	112
5.A.3	Learning curves	117
III Monte Carlo dynamics for high-dimensional, continuous phase space models		118
Chapter 6: Effective trap-like activated dynamics in a continuous landscape		119
6.1	Introduction	120
6.2	Dimensionality and entropic effects	122
6.3	Physical model	124
6.4	Off-equilibrium dynamics	125
6.4.1	Direct Sampling Monte Carlo	125

6.4.2	Markov Chain Monte Carlo	129
6.5	Trap-like behavior	131
6.6	Discussion & conclusions	133
6.A	Simulation details	137
6.A.1	Monte Carlo procedure	137
6.A.2	Calculation of the trapping time distributions	138
6.A.3	Calculation of the aging functions	138
6.B	MCMC calculation of the threshold	140
6.B.1	Threshold Radius for MCMC steps	142
IV	Back matter	145
Bibliography		146
List of publications & preprints as a graduate student		166

List of Tables

4.1	The distribution of classes in the structure database used for training the machine learning models.	83
4.2	\bar{F}_1 scores for different absorbing species, as the averages of the class-wise F_1 scores (red, green and blue bars) presented in Figure 4.3, both with (CNN) and without (MLP) the convolutional layer. Comparisons are made between models trained from the full feature space and reduced feature space corresponding to only the pre-edge region of the spectra.	89
4.3	The class-wise difference between the F_1 score evaluated over the entire feature space and over only the pre-edge region (Δ). Results are averaged over the early (Ti, V, Cr and Mn) and late (Fe, Co, Ni and Cu) transition metal elements.	92
5.1	Performance metrics based on the MAE of the spectra, $\overline{\Delta E}$ and $\overline{\Delta \mu}$	110
5.2	Summary of the input features. The first six denote node (atom) features \mathbf{V}_i , the next two are edge (bond) features \mathbf{E}_i and last is the adjacency matrix Q_i , which encodes all atomic connectivity. *Hydrogen atoms are initialized with a one-hot vector of all zeros. This serves two purposes: both indicating a lack of hybridization and forcing the MPNN to weight hydrogen atoms slightly less compared to the heavy atoms in the molecule. **Each bond type can be a single, double, triple or aromatic. ***For a molecule with z atoms, the adjacency matrix will be $z \times z$	113

6.1 For each model and dynamics, we specify whether we have energy- or entropy-driven activation. Trap model (TM) and Step model (SM) live in a fully connected phase space, so they cannot have local dynamics. The funnel model studied in this paper consists of a single well, so it cannot have energy-driven trap-like behavior. It does have entropy-driven activation for both local and global dynamics. The number partitioning problem (NPP) and the random energy model (REM) have energy-driven trap-like activation when using single spin flip dynamics. With global dynamics, the NPP has (as well as the exponential version of the REM) entropy-driven activation. This makes the NPP and REM models good candidates for having both kinds of trap-like behaviors simultaneously (i.e. with the same dynamics), since we can expect the same phenomenology of the funnel model.
^aThe trap and step models cannot have local dynamics. ^bThe subject of this work.
^cModels like NPP and REM could have entropically activated trap-like behavior also with local dynamics. ^dAn exponential REM with global dynamics is a SM.
The usual, Gaussian, REM has not been examined.

136

List of Figures

- | | |
|---|----|
| 2.1 Cartoon of an $L = 4$ -site boson cloud, $\mathbf{n} = [2, 0, 0, 3]$ (blue), embedded within a variational space specified by a maximum cloud extent $M = 5$ and maximum number of bosons $N = 6$. In this example, the constraint $M = 5$ spans sites $i - 1$ to $i + 4$ so that bosons can be created only on these sites (e.g., green circles), and are not allowed outside of the M -site cloud (e.g., red circle). $N = 6$ implies that states with one more boson than those depicted in the figure (blue circles) are omitted from the variational space. Note that the carrier (not shown) is allowed to be anywhere on the chain. | 16 |
| 2.2 Ground state energy E_{GS}/t as a function of coupling strength $\lambda = \alpha^2/2\Omega t$ for the Holstein model in adiabaticity limits extending from anti- ($\Lambda \gg 1$) to extreme-adiabatic ($\Lambda \ll 1$). Values for Ω/t are shown in grey, and GGCE results for $\Omega/t = 0.1$ and 0.5 are compared to DMC data (symbols) [70]. Ground state peak locations are converged with respect to N , and generally require ≈ 10 bosons at small couplings, but up to ≈ 30 at large couplings. | 23 |
| 2.3 Spectral function $A(k, \omega)$ at $k = 0, \pi/2$ and π for the Holstein model in (a) the mildly adiabatic ($\Omega/t = 0.5$) and (b) adiabatic limits ($\Omega/t = 0.1$), for dimensionless couplings $\lambda = 0.2, 0.5$ and 0.8 . In this figure we use $\eta = 0.005$. Various values of (M, N) are shown in the legend. Specifically, the gradient from red to black shows convergence with respect to M for fixed N . We demonstrate convergence with respect to N via the blue line, which shows results that use the largest used M , but with one less boson than the largest-used N . The onset of the continuum is shown in shaded gray, and is defined as $E_{\text{GS}} + \Omega$, where E_{GS} is the polaron ground-state energy. | 24 |
| 2.4 Spectral function $A(k, \omega)$ (scaled to a maximum of 1) of the Peierls model for $\Omega/t = 1$, $\eta = 0.05$ and various values of the dimensionless coupling strength, λ . $M = 5$ and $N = 10$ used here are sufficient for convergence of the bands on the scale of the plot. It is worth noting that fine structure in states above the lowest energy band can be resolved on a finer grid and smaller value of η , although the intensity of these states is at most roughly an order of magnitude smaller than that of the lowest energy band. | 25 |

2.5 Spectral function $A(k, w)$ (scaled to a maximum of 1) of the mixed-boson mode Holstein+Peierls model for various values of Ω_H and Ω_P , $\lambda_H = \lambda_P = 1$, $t = 1$ and $\eta = 0.05$. For these calculations, we use $M_H = M_P = 3$, and a maximum total cloud length, or absolute extent, $A = 3$ (see Appendix 2.C), and $N_H = N_P = 5$., for which semi-quantitative convergence is achieved.	28
2.6 Convergence of the energy of the ground-state polaron band, $E_P(k)$, for parameters shown in Fig. 2.5 against combinations of M_H , M_P , A , N_H and N_P	29
2.7 Example of a configuration of HP bosons corresponding to $n_H = [1, 0, 2, 1, 0]$ and $n_P = [0, 0, 1, 3, 1]$. Similar to the single boson models, we require that $\sum_j n_H^{(j)} \leq N_H$, $\sum_j n_P^{(j)} \leq N_P$, $L_H \leq M_H$, $L_P \leq M_P$ and $L \leq A$	32
3.1 Line width broadening of monolayer MoSe ₂ as a function of electron doping density for BN-encapsulated ($\varepsilon_0 = 4.5$) [146] monolayers. The following parameters were used: in the exciton calculation, the effective Bohr radii $\lambda_0 = 10.3$ and in the case of the trion, $\lambda_1 = \lambda_0$ and $\lambda_2 = 25.2$ Å [103]. In the exciton 2s elastic scattering, $a = 7.79$ Å and $b = 6.20$ Å (see Appendix 3.B.4). The electron (hole) effective masses employed were 0.49 (0.61) (in units of m_0) [147], and the polarizability $\chi_{2D} = 8.23$ Å [103]. In the case of exciton elastic scattering, the singlet and triplet contributions are identical as the exchange contribution to the potential dominates; trion triplet states are not considered. Additionally, screening using the effective frequency-dependent dielectric function (see Eq. 3.17) are presented for the trion, as the effective screening does not appear to affect the exciton line width.	57
3.2 Doping dependence of the 1s exciton line width at temperatures $T = 5, 6, \dots, 15$ K (left). Parameters describe monolayer MoSe ₂ , as seen in Fig 3.1. Lower line widths correspond to lower temperatures. The horizontal dashed lines show the plateau location. The value of the maximum (plateau) is also plotted as a function of temperature, showing clear linear behavior at low T (right).	59
3.3 Comparison of our theoretical results with the experimental results [157]. To compare qualitatively, we performed the following transformations to the experimental data: First, the intrinsic line widths are subtracted off of the experimental results. Second, while the experiments were conducted at 5 K, we compare the experimental results to our theoretical calculations at 25 K so as to better visualize the qualitative dependence on n	72

4.1 Workflow of the spectrum-based local chemical environment classification framework using supervised machine learning, which contains three modules: (I) data acquisition supplemented by high throughput computing (HTC) calculations, (II) labeling and (III) training of machine learning models. The machine learning architecture (set of hyperparameters) used in this work is shown in Module III. Notably, the model consists of an optional convolutional layer (shown in orange) followed by three hidden layers l_1, l_2 and l_3 consisting of 90, 60 and 20 neurons, respectively, ending with a softmax output. Further details of the network are described in Subsection 4.2.3.	80
4.2 FEFF K-edge XANES database of eight 3d transition metal families. Spectra were spline-interpolated onto discretized grids of 100 points, and scaled on the vertical axis such that the maximum value is 1 (prior to shifting the mean to 0 and scaling to unit variance). The average spectrum for each class, with its scaled spread $\sigma/5$ is shown. Principal component analysis of two regions is shown in the two insets: the full feature space (lower right), and only the pre-edge region (lower center, discussed in Section 4.3), with an equal sampling of the three classes displayed to aid visual clarity. The x and y-axes correspond to the first and second principal axes. Classes are color coded as follows: blue for tetrahedral (T4), purple for square pyramidal (S5) and red for octahedral (O6).	86
4.3 Class-wise F_1 scores calculated using different machine learning models (CNN/MLP) for T4 (blue/cyan), S5 (purple/light purple) and O6 (red/pink) local coordination environments in eight 3d transition metal elements. While the full height of each bar represents the results trained on the full feature space, gray bars overlaid on top with lower F_1 values represent the results for the models trained only on the pre-edge region. For example, for the Co S5 CNN, the F_1 score reported for training on the full spectral space is about 0.7, but decreases sharply to about 0.2 when trained on the pre-edge region only. Error bars correspond to standard deviation; the ones with wider caps correspond to the full feature space.	91
4.4 Comparison between representative experimental (solid) and FEFF (dashed) XANES spectra. FEFF calculations are shifted on the energy axis to maximize such that the Pearson correlation coefficient (PCC) in order to find the best match between experimental and theoretical XANES. The experimental spectra of $K_6Ti_2O_7$ and rutile were extracted from Farges <i>et al.</i> [178], and the experimental spectra of both pairs of Mn and Co oxides from Manceau <i>et al.</i> [225].	93
4.5 \overline{F}_1 score as a function of the amount of energy shift (ΔE) applied to the test set. The two graphs on the left illustrate the results of MLCs trained on augmented data (as presented thus far: ± 1 and 2 eV, generating 5 times the base amount of training data), while those on the right illustrate MLCs trained without augmenting the training set.	95

4.6	\overline{F}_1 score as a function of the standard deviation (σ) used to generate Gaussian random noise, introduced into every point E in the spectra $\mu(E)$. The two graphs on the left illustrate the results of MLCs trained on a training set augmented with Gaussian random noise of $\sigma = 0.03$ (also generating 5 times the base amount of training data), and those on the right illustrate MLCs trained without data augmentation.	96
5.1	The average oxygen spectrum for each considered functional group in S'_O , with the scaled standard deviation $\sigma/2$ presented to highlight regions of high variance. The approximate percent distribution of functional groups throughout the presented dataset is 30.0/42.9/10.5/8.1/8.6 (using the order from the legend) with a total of 47k in S'_O .	104
5.2	The average nitrogen spectrum for each considered functional group in S'_N , with the scaled standard deviation $\sigma/2$ presented to highlight regions of high variance. The approximate percent distribution of functional groups throughout the presented dataset is 19.3/3.3/6.0/32.4/37.3/1.7 (using the order from the legend) with a total of 35k in S'_N .	104
5.3	PCA plots for both the TCC and spectra proxies for the molecules in \mathcal{D}'_A labeled by NB, NA and FG. The total number of non-hydrogenic bonds (NB, top) range from 1 (violet) to 13 (red). The total number of atoms bonded to the absorbing atom (NA, center) takes on one of three values: 1, 2 or 3 (black, red and blue, respectively). The color legends for the functional group of the absorbing atom (FG, bottom) are the same as in Figs. 5.1 and 5.2.	105
5.4	Performance metrics for the MPNN evaluated on the \mathcal{D}_A testing sets. Top: waterfall plots of sample spectra (labeled by their SMILES codes) of ground truth (black) and predictions (dashed red), where prominent peaks (see text) are indicated by triangles. One randomly selected sample from every decile is sorted by MAE (first: best; last: worst).	108
5.5	Comparison between the prediction and ground truth in peak locations. Contains a total of 824 and 868 peak predictions and ground truths for the oxygen and nitrogen datasets, respectively. The lightest squares represent one data point, with the darkest representing roughly 40. The R^2 values for the presented data are 0.96 and 0.98, respectively.	109
5.6	Distribution of the absolute error of predicted peak heights, $\Delta\mu$, for both the oxygen and nitrogen datasets.	110

5.7 Molecular features of an exemplary molecule as processed by the MPNN. The connectivity matrix (zero-indexed) of the molecule is shown in addition to the features of atom 6 and bond 6-7. Specific connectivity is highlighted with purple lines on the molecule, and the corresponding entries circled in the connectivity matrix. . . .	114
5.8 Flowchart of the operation of the MPNN. Multi-layer perceptrons (MLPs) are denoted by gears, operations by purple blocks, and internal hidden variables in grey. Initial features are contained in the larger red square, and the target in light blue. Corresponding color-coded cartoons of the networks are shown to the right of the flowchart (note these are not representative of the actual architectures). Finally, above the cartoons is an example graph/molecule during the message passing (black arrows) and update (blue absorbing nitrogen atom) phases. The red squares indicate that edge and nearest neighbor information are used in MLP1 to construct the message.	115
5.9 L1 loss of the oxygen and nitrogen data sets as a function of training set size (see legend) and epoch. The largest training set size corresponds to the full training set available for that database. Solid lines: training losses; dashed lines: validation losses.	117
6.1 Ratio of the probability of the tracer moving towards the well singularity, P_{\downarrow} , and the tracer moving either away from it or staying in the same location, $1 - P_{\downarrow}$. Each curve represents a different starting radius, r_0	123
6.2 Threshold radius (green) and equilibrium value of the energy (black) plotted as a function of β/β_c for $N = 10$. The equilibrium phase (EP) regime, where the system enjoys both a well-defined threshold radius and equilibrium energy is shown in blue ($0 < \beta < \beta_c$). The thermal activation (TA) regime is shown in red. Finally, the entropic aging (EA) regime is shown in grey, where the tracer is relentlessly attracted towards the center of the well ($\beta > 2\beta_c$).	127
6.3 Main set: (a) Probability of the tracer moving up ($P_{\uparrow}(r; \beta)$), down ($P_{\downarrow}(r)$) or not moving ($P_0(r; \beta)$) for $N = 10$ and $\beta = 1.5\beta_c$ (TA regime). The threshold radius is identified by $P_{\uparrow} = P_{\downarrow}$. For $r > r_{\text{th}}$, we have $P_{\downarrow} > P_{\uparrow}$. For $r < r_{\text{th}}$, we have $P_{\downarrow} < P_{\uparrow}$ (see the inset (b) for a closeup of the difference between the two). The slowdown of the dynamics with small r is encoded in P_0 going to 1.	128

6.4 Top: We show the energy of three trajectories as a function of the Monte Carlo timestep t , for $N = 100$ and $\delta = 0.1$, each in a different dynamical regime. On the left we show DSMC dynamics (a), and on the right we show MCMC dynamics (b). The EP curve ($\beta = 0.2$) converges quickly to its equilibrium value $\langle E(\beta) \rangle = -1.25$ (horizontal blue dashed line), while TA and EA curves ($\beta = 1.5$ and 2.8 , respectively) diverge logarithmically to $-\infty$. Bottom: Configuration and basin trapping time distributions, $\psi_C(\tau_C)$ (dashed) and $\psi_B(\tau_B)$ (solid), for DSMC (c) and MCMC (d) dynamics. Results are averaged over 200 identical simulations, each with 500 tracers.	129
6.5 Main set: (a) Threshold radius as a function of β in a system of size $N = 50$, for DSMC [Eq. (6.10)] and MCMC [Eq. (6.16)] dynamics. The y axis is truncated at 0.5 to improve the figure's clarity. Points are shown in the curves in order to make overlapping curves visible. Insets: Ratio between the MCMC and the DSMC threshold radii for (b) $N = 50$ and (c) $N = 200$.	131
6.6 Top: (a) Aging functions $\Pi_B(t_w, 1.5 t_w)$ for DSMC dynamics, with $N = 100$, $\beta = 1.15, 1.25, 1.45, 1.75$. Bottom: (b) Aging functions $\Pi_B(t_w, 1.5 t_w)$ for MCMC dynamics for $N = 10$ and $\beta = 1.05, 1.15, 1.25$. The dashed horizontal lines correspond to the trap value $H_{2-\beta}(0.5)$. Results are averaged over 200 identical simulations, each with 500 tracers.	132
6.7 Top: (a) Aging functions $\Pi_B(t_w, 1.5 t_w)$ for DSMC dynamics with $\beta = 1.2$ and $N = 3, 6, 10, 30, 60, 100$. Bottom: (b) Aging functions $\Pi_B(t_w, 1.5 t_w)$ for MCMC dynamics with $\beta = 1.05$ and the same values for N . The dashed horizontal lines correspond to the trap values $H_{2-\beta}(0.5)$. Results are averaged over 100 identical simulations, each with 200 tracers. Inset: (c) The time at which the average $\Pi_B^{\text{MCMC}}(t_w, 1.5 t_w)$ function crosses 0.4 as a function of N (obtained through linear interpolation). Note the semi-log scale on the y-axis implies exponential scaling in N .	133

Acknowledgements

One thing I've realized about science: it's a team game, and nobody does it alone. Ever since high school, I have had a great deal of supporters, both academic and otherwise. This document and what it represents would not have been possible without them, and while it's difficult to put into words how much their support has meant to me, but I will try in the paragraphs to follow.

To my mentors, teachers, coworkers and friends. To begin, I owe a great deal to my PhD advisor, David Reichman. Dave is possibly best known amongst his students and post-docs for his outrageous, near-encyclopedic knowledge of all things quantum mechanical and ability intuitively crack difficult analytical problems, and while this has been incredibly helpful to me as a researcher, I want to focus more on his skills as an advisor. Having started from almost no coding/software experience, and barely any knowledge of condensed matter theory (at least in the context of research), the first few years were a struggle, but Dave really made the experience about the learning process, not the results, and that refocusing of my attention was invaluable. It is true that Dave generally takes a more hands-off approach, but he knows how to manage it, and ultimately this helped me to begin to think independently early on. Always careful to give me exactly the amount of help I needed to make progress, Dave enabled my development as a practitioner.

There is also the matter of my somewhat unorthodox path through my PhD. Throughout my time at Columbia, I have had quite a few opportunities to collaborate with scientists on many different projects. From two of Dave's post-docs, to scientists at Brookhaven National Laboratory, Dave has also been nothing but supportive of my extra-Columbia research endeavors, which is not

something a PhD student should take for granted. It is hard to overstate the incredibly positive impact these opportunities have had on my experience as a PhD student. Not every advisor would have green-lit every collaboration I ended up having, but Dave did because he knew it would benefit me, even if he wasn't involved. His willingness to let me pursue my own scientific questions not only helped me develop new skills, it has changed the course of my career for the better, and for that I will always be grateful.

I would also like to thank the Reichman Group, past and present, for being a simply excellent bunch of people, not to mention the incredible scientific talent present at any given time in the office. While at the time of writing these acknowledgements I have not seen many of you [in person] for over a year now due to the COVID crisis, it does not diminish the memories I have of us all hanging out in the "before times". While I'm not always the most social, I really did enjoy the lunches we all shared (including the legendary bagel lunches¹) and of course whisky Fridays, where the conversations were always political in nature (American politics, obviously) and the whisky always top-tier! Specifically, I want to thank Matt for his mentorship when I first joined the group, and James for the always-positive chats about science and just about anything else (and of course, the Soylent). Ian and Benedikt, it was a lot of fun hanging out in the back of the office for a few years with you both; I will forever be looking over my shoulder for wild bouncy balls on collision courses with my coffee mug! PJ and Jack, in a few months you'll be the senior-most graduate students; it's crazy how fast time flies, and I wish you both the best of luck in the remainder of your PhD studies, which I have no doubt will be successful. I have also gotten to know quite a few post-docs throughout my time in the group. Roel, Martina, Chitrak, Joonho, Dmytro and Lachlan, having your perspectives and talent available on such a wide variety of topics was always helpful. Marco and John, I have the privilege of calling you not just my friends, but my mentors as well. You are both excellent teachers, and I am happy to have had the opportunity to work together, and hope we can continue our research projects long into the future.

Finally, I want to also thank some other Columbia faculty for their instrumental role in my

¹Best Bagel and Coffee on 35th street is better than Absolute Bagels though. I am 100% objectively correct on this matter.

professional development. Rich Friesner, Chris Marianetti, and Tim Berkelbach, I am happy to have had the chance to learn from you all in various capacities. Additionally, to Socky, Daisy, Sheila, Dani and Alison, you are the backbone of this department, thank you for always making sure the gears keep turning.

Through the Department of Energy Computational Science Graduate Fellowship (CSGF), which I was awarded in 2017, I have had the opportunity to collaborate with numerous outstanding scientists at Brookhaven National Laboratory (BNL). In a lot of ways, I have had the privilege of learning under not just my advisor at Columbia, but another group of advisors and colleagues at BNL. Before I give individual thanks, I would like to acknowledge the incredible opportunity that the CSGF has afforded me. Lindsey and the entire team at the Krell Institution run a tight ship, and the Annual Program Review conference every year in Washington D.C. is an event I will look forward to attending as an alum indefinitely. Moreover, I would like to thank Shinjae Yoo and Deyu Lu, who together have been my advisors in all things machine learning and how it can be applied to outstanding physical and chemical problems. They have not only given me the opportunity to collaborate with them on research projects, but have also allowed me to begin to get my feet wet writing real scientific proposals, an opportunity not many people have. It has been a pleasure working with them both, and I am looking forward to continuing to do so in the future. I also cannot imagine my time at BNL having been the same without my practicum points-of-contact Nick D'Imperio and Mike McGuigan, as well as the Center for Functional Nanomaterials theory group leader Mark Hybertsen (who also happens to be a close collaborator of the Reichman Group), who have served for a while now not only as professional advisors, but as trusted mentors. Finally, I have had the great pleasure of collaborating with the incredible teams of Shinjae and Mark, as well as others at the Condensed Matter and Materials Science division, including (but not limited to) Mehmet, Huub, Wei, Xiaohui, Ray, Ai, Qin, Erica, Andreas and Robert. Thank you to everyone on those teams, and all my other collaborators, for being a truly fantastic group of people to work alongside. With any luck, we will have many opportunities to continue to work together for quite a while.

Prior to coming to Columbia University, I studied chemistry and physics at the University of Rochester (U of R) in upstate New York, where I had the opportunity to learn from and work with a group of people who were instrumental in shaping my future. First, I thank Joshua Goodman, whose organic chemistry classes (and opportunities to be a teaching assistant for his organic chemistry classes) helped me decide on chemistry as a career, and Rudi Fasan for giving me my first opportunity to work in a research lab. Next, I am happy to have had the opportunity to work in both Lewis Rothberg's and Pengfei (Frank) Huo's groups in my last year at the U of R. I will always remember Lewis as a wonderful mentor who cares tremendously about his students, working with him both in lab, and as a TA for his classes was a real joy. I owe a special thanks to Frank for welcoming me, as his first student, into his lab, and teaching me the basics of theoretical chemistry. This opportunity was invaluable, and paved the way for me to ultimately join the Reichman Group at Columbia.

I would also like to take a moment and thank the teachers I've had even before college. The problem solving skills I learned at Watchung Hills Regional High School served me exceptionally well in college, and helped set me up for success later on in life. First, I would like to thank my old science and mathematics teachers, Michael Amendola, Michael Gangluff, Dr. Sophia Gershman, Brian Brown, Salvatore Fazzino, Daryn Mount, and Christopher Gibson. From chemistry (and vuvuzelas) to physics and even independent research, you all showed me how rewarding science can be, despite the occasional frustration.

Before I give thanks to my incredibly supportive family and friends, there are two teachers I would like to thank specifically, since perhaps even without them noticing, their teaching has had a profound effect on my character and confidence, both in and out of the classroom.

I would not be where I am today, in the field of work I am in today, without a truly exceptional person, Ignacio Franco. Ignacio pushed me to pursue a career as a theorist, something prior to our conversations, I simply lacked the confidence to do, despite my obvious love of the subject matter. It is not an overstatement to say that if not for him, I may very well be in an entirely different field of work, possibly one I am not meant for. His rigor and passion for physics and chemistry have

benefited me in so many ways, and will continue to benefit countless students who come after me.

Last but certainly not least, the most influential person in my educational career has, perhaps surprisingly, not been a science teacher. My freshman and junior year English teacher (and now rightfully the director of curriculum of the entire high school), Mary Ellen Phelan, has been a profound influence in my life ever since my first year at Watchung Hills. She taught me that being a student is more than learning in a classroom, it is about becoming an informed, thoughtful citizen of the world. Since then, we have brainstormed on countless topics, many of which one day we might try to write down, and have collaborated on a somewhat annual hackathon at my old high school. I am and will always be grateful for her trusted advice on just about any topic I can think of, and for the excellent conversation (when we can find overlapping time in our schedules), which I'm sure will continue long into the future.

Having been already luck enough to have myriad exceptional teachers, mentors and collaborators join me in this journey, I have also been blessed with an incredible group of family and friends. Given the nature of my work, there is often overlap between these two groups, so I will henceforth focus on those not yet mentioned. There are far too many people to thank in this group, and despite highlighting only a subset, know that I am thankful to you all.

To start, I would like to thank my good friend of almost a decade, Steven Torrisi. Starting from the days of freshman physics at the U of R, onwards to the "experience" that is classical and quantum mechanics homework in the POA, through the early and now later days of research and the CSGF we both were awarded, to our collaboration and now looking towards our future careers, it's nice to have taken this journey in parallel with you. And as always, I am looking forward to the doubtless many conference coffee breaks and (hopefully) many more opportunities to work together. To Abigail and Sam, and Natalia, thank you for being there to chat about life, science and whatever else I felt like complaining about (and for endless Chipotle past, present and future). To Alyssa, it has been fantastic getting back in touch with you, and given we're both going to stay in Manhattan, I'm looking forward to many more lunches at Franchia and the seemingly infinite number of other excellent places here in the city. To everyone else, including Malte, John,

Kevin, CJ, Matt, Randy and many others, thank you for your friendship; hopefully, the darkness of COVID has almost passed and we can get back to life as usual!

To my family. Throughout my life, including the difficulties of my childhood all the way to the writing of this document, there has been the constant of family. Their love and support made all of this possible.

First, I am lucky to have a small army of aunts, uncles and cousins for whom family comes above all else. To Paul and Farah, thank you for being there any time, for always being so welcoming, and for the absolutely top-tier, restaurant-quality, unmatched home cooking. To Mike and my more-newly acquired uncle, Alan, you are both a joy, here's to many more lunches in Manhattan! To Kathy, you are the only other person in the family who truly understands the "process" of getting a PhD-level degree, it's been nice to share that with you. Finally, to Ed, who helped me become computer-fluent early on in my life (and for all the video games, obviously). To everyone else, Jennifer, Adam, Amina, Keith, Jake, Aiden, John, Lorraine, Robin and Adrianna, you make me happy to be a part of the family. To many in my now-extended family, including Edie, Mel, Al, and Paul, thank you for always being so welcoming, kind, and thoughtful to me.

Next, to my grandmother and late grandfather, who helped raise me: you imbued in me the principles of hard work and discipline. While we may sometimes not see eye-to-eye, know that you helped to create a free-thinker, who cares about others and the world. When times were most difficult, we banded together as a family and got through it. When you were called upon in your golden years to take on raising children, *again*, you did without question. You taught me how valuable family is, and for that I thank you.

In the midst of the pain of loss, which never truly disappears even now, I am lucky to have two incredible women in my life who have become somewhat like mothers to me. To Helene, while I admit I can certainly appear to be somewhat... boring in comparison to you and Dad, I greatly appreciate your sound insight and advice, our conversations in general, and as my Dad will attest to, your hilarious, spontaneous sense of humor. To Ivy, when you start dating someone, it is

undeniable that you roll the dice regarding the family you acquire through that person, but clearly, I hit the jackpot. I always find our chats heartwarming, and thank you for your endless support.

To my stepbrothers, Matt and Spencer, when my father first met your mother, I had no idea who I'd be getting in terms of my newly acquired siblings. Needless to say, I couldn't be happier to have gotten you both. Spencer, you light up a room and are always excellent company. Matt, even as I stare, seemingly irritated, at the times you crack a joke and bring the house down, I'm laughing inside, I promise.

To Dad, it goes without saying that you enabled my success. Throughout your life, you navigated impossible waters: your childhood, West Point, the company you built from nothing and ran for over twenty years, raising your children without their mother, and beyond. Learning about, and seeing firsthand, your difficulties, and the way you overcame them, taught me how to handle the challenges life throws at us. From the moment you bet on my success and sent me to U of R debt free, I was determined to make you proud, I hope I have. Thank you for being there for me, I am proud to be your son.

To my sister, Jillian, wow the time flies, right? At the time of writing this, you have just gotten your first job after graduating from college. It seems like just yesterday we were arguing with each other as teenagers over inane things I cannot even remember. At that time, it was hard to imagine we'd even be friends (sibling rivalry and all that), but obviously that cannot be further from the truth. I am so proud of you, and so happy you are my sister. Never let anyone stand in the way of your dreams, never let anyone tell you that you cannot do something for any reason. Go be the person you are, unapologetically, and know you will always have my unconditional support.

Last but certainly not least, to my partner, Sarah. First, I'm sorry, I know I can be a massive pain. You have had the difficult job of helping me manage, in a very concentrated environment (studio apartment during COVID), the immense pressures that come with doing a PhD. For a long while, I didn't quite believe I would find someone like you, but low and behold, here we are five and a half years later! Thank you for being a relentlessly supportive force in my life and for being there for me when times were the most challenging. You are not only my partner but my

most-trusted confidant and best friend. I love you and am looking forward to the years to come!

Mom, this is for you.

Your strength and courage will forever be an inspiration.

Chapter 1

Introduction

At its core, the goal of science is to develop theories and models that can explain real-world phenomena. The set of all models is vast, ranging from theories of the unimaginably small (e.g. quantum mechanics) to the equally unfathomably large (e.g. general relativity). Perhaps disturbingly, no model is and will ever be "correct," no matter how accurate it's predictions. Such claims fall squarely outside the realm of science (we leave this to the philosophers), and instead, models will only ever be more or less wrong than others in one situation or another. The theme of this thesis is presenting an array of work which attempts to model a variety of phenomena in quite a few different ways, including using "microscopic" quantum mechanical methods (Chapters 2 and 3), machine learning models of various types (Chapters 4 and 5) and toy model simulations of activated dynamics (Chapter 6).

In this chapter, we briefly discuss some of the models that exist in the literature, and the various ways they are applied. We will also highlight their limitations, and compare and contrast two fundamentally different modeling paradigms: that of the theory- vs. data-driven models. Finally, we outline the chapters in this thesis, and provide an overview of the important topics and conclusions reached.

1.1 Theory- vs. data-driven models

A large portion of this thesis focuses on implementing machine learning (ML) techniques, and thus it is important to highlight the critical distinction between what we call theory- and data-driven. The first two chapters of this thesis focus on theory-driven models, and the third and fourth on data-driven models.

Consider for instance the time-independent Schrödinger equation,

$$H\psi_n(\mathbf{x}) = E_n\psi_n(\mathbf{x}). \quad (1.1)$$

Under the Born-Oppenheimer approximation [1], the ground state energy of some assembly of atoms is a function of the nuclear coordinates (and atom types), $E_0 \equiv E = E(\mathbf{R})$. Formally, Eq. (1.1) is what we would call a theory-driven model, because the model itself is based on physical principles, and it does not require conditioning on data to make predictions. In other words, one cannot *trivially* improve a theory-based model in a conditional sense, without completely revising the model (improving the theory itself).

A data-driven model takes the opposite approach. For instance, let's say you had a bag of different molecules, and for each one you had measured the ground state energy using some experimental method, such that you had pairs of data points, $\mathcal{D} = \{(\mathbf{R}_1, E_1), (\mathbf{R}_2, E_2), \dots\}$. With this data in hand, one can parameterize a model f_θ such that θ corresponds to the minimum value of some average loss, e.g.,

$$L(\theta) \propto \sum_i (f_\theta(\mathbf{R}_i) - E_i)^2. \quad (1.2)$$

This is effectively the ML paradigm,¹ where f_θ is a data-driven model. The Born-Oppenheimer approximation would be "improved" if higher-level theory were incorporated into it, but the data-driven model would only be improved if it received access to more data. In this way, a data-driven

¹While the general strategy is similar to this regardless, the outlined protocol is explicitly the approach take in *supervised* learning.

model can be improved in a conditional sense.

Generally speaking, these two model paradigms come with various strengths and weaknesses. For example, in terms of speed, a theory-driven model might be extremely expensive to evaluate, and will be every time. Density Functional Theory does not, for example, get any less expensive the more data it sees. A data-driven model on the other hand must be trained on such ground truth data, which can be expensive to generate, but then can make accurate predictions instantaneously (relatively speaking). Another example is that of interpretation. Generally, theory-based models are easily interpretable (though perhaps not easy to evaluate) by simple observation of the algebraic structure of the equations. However, data-based models are usually *not* easily interpretable. Deep neural networks can contain billions of trainable weights, used at different levels of the network, passing data which is non-linearly transformed. Interpretation of these trained network architectures is an active and open research question. These and other tradeoffs are largely summarized in Chapters 4 and 5.

1.2 Quantum theory

Quantum mechanics is bar none the most well-validated scientific theory ever conceived [2]. It makes almost magically accurate predictions that completely break from classical scientific theory, and at the atomic scale, it is the current state of the art for predicting the properties and behavior of matter. However, there is trouble in paradise: real systems consist of many particles, and the equations that govern the time-evolution and static properties in the quantum mechanics framework become exponentially expensive to solve.² The time-dependent Schrödinger equation governs all dynamics,

$$i\hbar \frac{\partial}{\partial t} \psi(\mathbf{x}, t) = H\psi(\mathbf{x}, t), \quad (1.3)$$

and the time-independent Schrödinger equation [Eq. (1.1)] provides access to the wave functions which completely characterize the system state, but the high dimension of the many-body wave

²Formally, the computational complexity of the full, exact solution in quantum mechanics scales exponentially in the number of electrons.

function makes the problem unsolvable in practice. Specifically, the electron-electron coupling in H is by far the most difficult part of the problem.

To circumvent these challenges, a historied suite of methods have been introduced to surgically tune the tradeoff between accuracy and computational tractability. Generally, all theories build in some way on the Hartree-Fock approximation, which is an independent-electron treatment³ of the many-body problem, where electron-electron interactions are treated in a mean-field sense. On the computational chemistry side of the field, methods such as Coupled Cluster Theory [3], which provides the most accurate results at a high computational cost, and Density Functional Theory [4], which offers more of a balance between accuracy and computational cost, have emerged as frontrunners. On the different side of the same coin, we have Green's function-based methods. Through the Lehman representation [5], e.g. [6],

$$G(\mathbf{k}, \omega) = \sum_n \frac{|\langle \psi_n | c_{\mathbf{k}} | \psi_0 \rangle|^2}{\omega - E_n + i\eta}, \quad (1.4)$$

Green's functions are deeply connected to the aforementioned wave function-based methods, but offer a different utility [7], which is discussed in especially Chapter 2.

1.3 Machine learning

Broadly speaking, ML is a protocol for training a program to make predictions without being explicitly programmed to do so. In many ways, ML is no different than being presented with points on a 2-dimensional plot, and being asked to draw an interpolating function between them. Except, of course, that in practice it is an arbitrarily high-dimensional plot, and the interpolating function (the model itself) must strike a careful balance between capturing the usually highly nonlinear trends and not over-fitting. Fine-tuning models for optimal performance is the challenge of the ML practitioner.

With the strengths and weaknesses of ML [c.f. Subsection 1.1] in mind, we highlight that while

³Independent in the sense that only the Pauli Exclusion Principle is obeyed.

interpolating procedures on their own can never truly *discover* new physics, they can be utilized in resourceful ways to speed up calculations and access new domains of study. For example, Neural Network Potentials [8] such as AENET [9, 10] have been demonstrated to predict materials properties with high accuracy, and offer enormous speed up over conventional methods. Augmenting existing simulation techniques with ML could offer access to new time-scales without incurring extra computational cost. In a tangential field, Google Deepmind has seemingly solved the protein folding problem using a combination of supervised and reinforcement learning techniques [11], which if broadly generalizable is a major breakthrough.⁴ While there is no doubt that ML offers huge utility in applied problems, finding new and clever ways to apply ML in the domain of scientific *discovery* is an extremely active field of research.

1.4 Outline

We henceforth outline the work described in this thesis. Work in the first two chapters was performed under the supervision of primarily my advisor: Professor David R. Reichman,⁵ and also John Sous⁶ (for Chapter 2 only). Work in the second two chapters was performed under the supervision of Shinjae Yoo⁷ and Deyu Lu⁸ in collaboration with Brookhaven National Laboratory. Finally, work in the last chapter was performed under the supervision of Marco Baity-Jesi.⁹

In Chapter 2, we leverage the Dyson equation and zero- T Green's function formalism to develop a generalized boson cluster expansion method for solving arbitrary lattice phonon model Hamiltonians. Specifically, we build off of the established family of momentum average methods [12], and generalize it to the numerically exact limit, including providing a computationally feasible approach to further generalizing the developed code to finite temperatures and higher di-

⁴See also <https://www.nature.com/articles/d41586-020-03348-4>, which outlines the impact of that work.

⁵Professor of Chemistry, Columbia University

⁶Joint Post-doc in the Reichman and Millis groups, Columbia University

⁷Machine Learning Group Leader, Computational Science Initiative, Brookhaven National Laboratory

⁸Staff Scientist, Center for Functional Nanomaterials, Brookhaven National Laboratory

⁹Post-doc in the Reichman Group, Columbia University during the time of this work, now the Machine Learning and Data Science Group Leader, Eawag, Switzerland

mensions.

In Chapter 3, to study the broadening mechanisms of excitons and trions in monolayer transition metal dichalcogenides, we develop a fully microscopic theory based on a Fermi Golden Rule approach to explain the dominant contributions to the lifetimes of these quasi-particles. Our microscopic approach assumes reasonable models for encapsulating dielectric materials and quasi-particle wave functions, with scattering matrix elements and all other terms calculated in a numerically exact fashion. After the publishing of this work, we later validate the accuracy our approach against experimental data.

Chapters 4 and 5 share the common theme of leveraging ML algorithms to make accurate predictions of the physical properties of materials and molecules. Specifically, in Chapter 4 we train feed-forward neural networks to classify local absorbing atom information across a broad range of transition metal oxides, from their simulated x-ray absorption near edge structure (XANES) spectra. In Chapter 5, we do the reverse, and train graph-based neural networks to predict the XANES spectra directly from molecular structures to quantitative accuracy.

In Chapter 6, we explore entropically activated dynamics in the funnel model. Specifically, we both generalize to a continuous phase space, and showcase the curious result that non-equilibrium behavior of a tracer in this model is seemingly largely independent of the choice of dynamical method, be it direct sampling or Markov Chain Monte Carlo.

Part I

Quantum dynamics in the condensed phase

Chapter 2

Numerically Exact Generalized Green's Function Cluster Expansions for Electron-Phonon Problems

The content presented in this chapter is based on the following work posted on the arXiv.

- Numerically Exact Generalized Green's Function Cluster Expansions for Electron-Phonon Problems [[arXiv](#)]

M. R. Carbone,^{*} D. R. Reichman & J. Sous.^{*} (under review, Phys. Rev. B.)

^{*} Indicates corresponding authors.

2.1 Introduction

The interaction of a particle with its environment is central to the study of many physical systems. One classic problem of this type is that of the polaron, which describes a mobile carrier dressed by bosonic fluctuations [5]. Originally predicted by Landau [13], expanded upon by Pekar [14, 15] and cemented into condensed matter canon by Lee, Low and Pines [16], Fröhlich, Pelzer and Zienau [17, 18], Feynman [19], and Holstein [20, 21], a polaron forms when a particle such as an electron or hole moves in a deformable medium. The motion of the particle induces a local polarization cloud, which is dragged along with the particle as it moves, renormalizing its effective mass and yielding a non-zero quasiparticle weight. Polarons arise in a variety of physical contexts beyond that of electron-phonon systems [22], such as excitons in photoexcited molecular crystals [23–26], hole-doped magnets [27], light-matter systems [28–30], impurities embedded in ultracold gases [31–34] and in other more exotic physical settings [35–37].

Over the last two and a half decades, many (in principle) exact numerical methods have been devised to study polaronic problems. One can broadly classify these approaches into two main categories: real- and imaginary-frequency methods. Approaches in the former class include Variational Exact Diagonalization [38], and its variants [39, 40], Limited Phonon Basis Exact Diagonalization [41] and Matrix-Product-State techniques [42–45]. Methods in the latter class are most prominently Monte Carlo methods, such as Diagrammatic [46–48], Path-integral [49] and Continuous-time [50] Monte Carlo. While Monte Carlo techniques are well suited for the study of finite-temperature systems over the complete range of polaronic model parameters, they require ill-conditioned analytic continuation to the real-frequency axis in order to study dynamics. In contrast, direct real-time methods face a daunting challenge in several parameter regimes, including the so-called adiabatic limit where the lattice response is slow, as well as the strong-coupling limit, where a large number of bosons is excited in the system and the size of basis states becomes too large to efficiently manage.

In this work, we introduce the Generalized Green’s function Cluster Expansion (GGCE), a

non-perturbative approach that enables an exact, efficient numerical computation of real-frequency Green's functions of polaronic models even in regimes challenging for related real-frequency approaches. In particular, we show that the GGCE provides access to exact spectra in the portions of the adiabatic and strong-coupling limits inaccessible to more standard Variational Exact Diagonalization approaches, while converging more rapidly in accessible regimes. Our method builds on the Momentum Average (MA) Approximation,¹ proposed by Berciu in 2006 [12] which has since been adapted to describe realistic materials [61, 62]. Our procedure is applicable to any form of particle-boson coupling, and proceeds via efficient generation of an equation of motion (EOM) in orders of the spatial extent of bosonic clusters that arise in the dynamics. We show that this approach variationally recovers the exact infinite boson Hilbert space, provided that one converges the computation with respect to the cluster size, and we find that this is achieved with a high level of efficiency when compared against standard numerical approaches, even in the adiabatic limit. In addition to providing access to quasiparticle spectra over a wide frequency range, the GGCE comes with several strengths. In particular, it is formulated in the *infinite* system size limit, and thus provides access to exact spectra in the thermodynamic regime. It affords sufficient flexibility that permits extensions to finite-ranged models at finite temperatures and in higher dimensions, as well as to studies of bipolarons, and systems with different boundary conditions. Additionally, it allows the study of dynamics of non-equilibrium initial states. Lastly, since existing linear algebra solvers represent the only computational bottleneck in the approach, the GGCE serves as an easy-to-implement, methodologically unconstrained technique whose performance is limited only by access to computational resources such as large-scale parallel computing or GPU technology.

Our manuscript is organized as follows. In Section 2.2, we review the foundations of the MA methods and devise a generalized formalism we use in the GCCE approach (Subsection 2.2.1). We briefly discuss our computational implementation of the method (Subsection 2.2.2) and highlight the relationship to and differences between our and other methods (Subsection 2.2.3). In Sec-

¹The MA approach has been validated for a large number of systems, including, but not limited to, Holstein [6, 12, 51–54], Peierls [55], Edwards [56], and dual-coupled polarons [57], Holstein [58] and Peierls bipolarons [59], and has been applied to model experimental systems such as graphene [60] and cuprates [61]

tion 2.3, we demonstrate the power and scope of this implementation and present a combination of numerically exact and quasi-converged results on the Holstein, Peierls and mixed-boson mode Holstein+Peierls models. Finally, in Section 2.4, we conclude and discuss possible future work.

2.2 Methodology and General Considerations

Consider a mobile particle (e.g. electron, hole, etc.) coupled to a bosonic field²

$$\hat{H} = \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} \hat{c}_{\mathbf{k}}^\dagger \hat{c}_{\mathbf{k}} + \sum_{\mathbf{q}} \hbar \Omega_{\mathbf{q}} \hat{b}_{\mathbf{q}}^\dagger \hat{b}_{\mathbf{q}} + \sum_{\mathbf{k}, \mathbf{q}} g(\mathbf{k}, \mathbf{q}) \hat{c}_{\mathbf{k}+\mathbf{q}}^\dagger \hat{c}_{\mathbf{k}} (\hat{b}_{-\mathbf{q}}^\dagger + \hat{b}_{\mathbf{q}}). \quad (2.1)$$

Here, the carrier (boson) has dispersion $\varepsilon_{\mathbf{k}}$ ($\hbar \Omega_{\mathbf{q}}$), and the interaction \hat{V} contains a vertex $g(\mathbf{k}, \mathbf{q})$ that in general depends on both \mathbf{k} and \mathbf{q} . We use a compact notation $\sum_{\mathbf{k}}$ to imply a discrete sum for a problem formulated on the lattice or a d -dimensional integral $\frac{L^d}{(2\pi)^d} \int d^d k$ with L^d the system volume for a problem in the continuum.

The goal of our approach is to derive the EOM of the one-electron Green's function at zero temperature [5],

$$G(\mathbf{k}, \omega) = \langle 0 | \hat{c}_{\mathbf{k}} \hat{G}(\omega) \hat{c}_{\mathbf{k}}^\dagger | 0 \rangle. \quad (2.2)$$

For Hamiltonians of the form in Eq. (2.1), only the retarded component of $G(\mathbf{k}, t)$ contributes [12], and the propagator, in real frequency, takes the form

$$\hat{G}(\omega) = [\omega - \hat{H} + i\eta]^{-1}, \quad (2.3)$$

where $\eta = 0^+$ is an artificial broadening parameter. Repeated application of Dyson's equation,

$$\hat{G}(\omega) = \hat{G}_0(\omega) + \hat{G}(\omega) \hat{V} \hat{G}_0(\omega), \quad (2.4)$$

²In this chapter, we label all quantum operators with hats to alleviate possibly confusing notation between e.g. the propagator, $\hat{G}(\omega)$, and the Green's function, $G(k, \omega)$.

with $\hat{H}_0 = \hat{H} - \hat{V}$ yields an infinite hierarchy of equations,³ which we compute in the basis states $|k, n\rangle$ labeling a delocalized state of a carrier and n bosons with total momentum momentum k . We start by deriving the EOM for $G(\mathbf{k}, \omega) \equiv \langle k, 0 | \hat{G}(\omega) | k, 0 \rangle$. The first application of Dyson's equation yields

$$G(\mathbf{k}, \omega) = G_0(\mathbf{k}, \omega) \left[1 + \langle 0 | \hat{c}_{\mathbf{k}} \hat{G}(\omega) \hat{V} \hat{c}_{\mathbf{k}}^\dagger | 0 \rangle \right], \quad (2.5)$$

and the second gives

$$\langle 0 | \hat{c}_{\mathbf{k}} \hat{G}(\omega) \hat{V} \hat{c}_{\mathbf{k}}^\dagger | 0 \rangle = \langle 0 | \hat{c}_{\mathbf{k}} \hat{G}(\omega) \hat{V} \hat{G}_0(\omega) \hat{V} \hat{c}_{\mathbf{k}}^\dagger | 0 \rangle, \quad (2.6)$$

where $\hat{G}_0(\omega)$ is the free particle propagator, and

$$\hat{G}_0(\omega) |k, n\rangle = G_0(k, \omega - n\hbar\Omega) |k, n\rangle. \quad (2.7)$$

Note this expansion can be indexed by the number of bosons contained in the created states. A coupling \hat{V} that is linear in boson operators either creates or annihilates a boson, thus coupling states with n bosons to states with $n \pm 1$ bosons. A key development made by Berciu [12, 51] is to recast the EOM as a hierarchical "expansion" in orders of the spatial extent of the bosonic cloud, M , rather than treating it as a direct expansion in the number of bosons. Making use of the spatial structure of the Green's functions generated in the EOM allows one to derive a scheme in which states corresponding to clouds larger than a certain spatial extent M are excluded. To illustrate the idea, consider the example of $M = 2$. At this level of approximation, only states with bosons localized on single and first-neighbor sites are retained in the hierarchy. Note that this imposes no restriction on the distance between the carrier and the boson cloud. We can view this approximation as a variational ansatz in the space of Green's functions in which one allows the carrier anywhere in the system, but with bosons clustered in a cloud of a maximum length M .

Below, we detail the approach we use to construct and solve the linear system of equations

³The connection to other approaches that utilize exact hierarchies for dynamics in polaron models (for example, the HEOM approach [63, 64]) remains to be explored.

in the EOM. Specifically, in Subsection 2.2.1 we derive a generalized expression for $G(k, \omega)$ for arbitrary models. Then, in Subsection 2.2.2, we explain how to systematically generate and solve the system of equations in computer simulations. Finally, in Subsection 2.2.3, we discuss the relation of the GGCE to other methods.

2.2.1 A Generalized Equation of Motion

We now specialize our construction to the case of one-dimensional (1D) lattice models described by Hamiltonians of the form

$$\hat{H} = -t \sum_{\langle ij \rangle} \hat{c}_i^\dagger \hat{c}_j + \Omega \sum_i \hat{b}_i^\dagger \hat{b}_i + \hat{V}, \quad (2.8)$$

where $\langle ij \rangle$ denotes nearest neighbors, for which numerical results are available. This allows us to both benchmark GGCE against exact numerics and to tackle regimes that are typically difficult to study or inaccessible by related techniques even in the well-studied 1D limit. In what follows we set $\hbar = 1$ and the lattice constant $a = 1$.

Beginning with Eq. (2.5), we derive a generalized EOM (GEOM). Here the free particle Green's function is given by

$$G_0(k, \omega) = [\omega - \varepsilon_k + i\eta]^{-1}, \quad (2.9)$$

with free particle dispersion $\varepsilon_k = -2t \cos k$.

Consider a generalized representation of \hat{V} for models that describe coupling between a carrier and a single bosonic mode,

$$\hat{V} = \sum_{(g, \psi, \phi, \xi)} g \sum_i \hat{c}_i^\dagger \hat{c}_{i+\psi} \hat{b}_{i+\phi}^\xi. \quad (2.10)$$

Here g is the coupling constant, $\psi, \phi \in \mathbb{Z}$ encode the spatial dependence of the coupling, and $\xi = \{-, +\}$ labels bosonic operators as either annihilation ($b^- \equiv b$) or creation ($b^+ \equiv b^\dagger$). This generalized notation completely specifies \hat{V} for a given arbitrary finite-ranged model. We present examples of such models in Appendix 2.B. For clarity, let us specialize to the Holstein model as

an example:

$$\hat{V}_H = \alpha \sum_i \hat{c}_i^\dagger \hat{c}_i (\hat{b}_i^\dagger + \hat{b}_i) \quad (2.11)$$

can be represented in this notation as follows

$$\hat{V}_H = \alpha \sum_i \hat{c}_i^\dagger \hat{c}_i \hat{b}_i^\dagger + \alpha \sum_i \hat{c}_i^\dagger \hat{c}_i \hat{b}_i \leftrightarrow \{(\alpha, 0, 0, +), (\alpha, 0, 0, -)\}. \quad (2.12)$$

We allow for an arbitrary but finite number of interaction terms, which need not be equal and can thus be used to model, for example, a long-ranged coupling of a carrier to a bosonic mode.

Using Eq. (2.5), we arrive at the GEOM for $G(k, \omega)$,

$$f_0(0) = G_0(k, \omega) \left[1 + \sum_{(g, \psi, \phi, \xi)} g e^{\pm k R_{\psi-\phi}} f_1(\phi) \right]. \quad (2.13)$$

Here, we have defined an *auxiliary* Green's function (AGF) [12, 56] given by

$$f_n(\delta) = N^{-1/2} \sum_i e^{\pm k R_i} \langle 0 | \hat{c}_k \hat{G}(\omega) \hat{c}_{i-\delta}^\dagger \hat{b}_i^\dagger | 0 \rangle, \quad (2.14)$$

where N is the number of lattice sites, $R_m \equiv m$ and $f_n(\delta) \equiv f_n(k, \delta, \omega)$. The AGFs can be thought of as higher-order propagators of an electron in a spatial cloud composed of multiple bosonic excitations. Further, we note the identity $f_0(\delta) = e^{\pm k R_\delta} G(k, \omega)$, c.f. Eq. (2.14).

It is now necessary to introduce additional notation for describing how AGFs with > 0 bosons couple. Since bosons can in general be created anywhere on the lattice, we define an occupation number vector \mathbf{n} , which labels the number of boson excitations starting from site i on a cloud embedded within the infinite lattice,

$$\mathbf{n} \equiv [n^{(i)}, n^{(i+1)}, \dots, n^{(i+L-1)}], \quad (2.15)$$

where $L \leq M$ is the length of \mathbf{n} . This vector serves as a device for labeling the bosonic Hilbert space in the following way: $\mathbf{n} \leftrightarrow B_{i,\mathbf{n}}^\dagger |0\rangle$, where $B_{i,\mathbf{n}}^\dagger \equiv \hat{b}_i^{\dagger n_0} \hat{b}_{i+1}^{\dagger n_1} \dots \hat{b}_{i+L-1}^{\dagger n_{L-1}}$. This allows us to write a generalized version of Eq. (2.14), where the n becomes a vector,

$$f_{\mathbf{n}}(\delta) = \mathcal{N}^{-1/2} \sum_i e^{\pm k R_i} \langle 0 | \hat{c}_k \hat{G}(\omega) \hat{c}_{i-\delta}^\dagger \hat{B}_{i,\mathbf{n}}^\dagger | 0 \rangle. \quad (2.16)$$

Upon Fourier transforming to reciprocal space and substituting Dyson's equation, we obtain

$$f_{\mathbf{n}}(\delta) = \mathcal{N}^{-1} \sum_i e^{\pm k R_i} \sum_q e^{-\pm q R_{i-\delta}} G_0(q, \omega - n_T \Omega) \langle 0 | \hat{c}_k \hat{G}(\omega) \hat{V} \hat{c}_q^\dagger \hat{B}_{i,\mathbf{n}}^\dagger | 0 \rangle, \quad (2.17)$$

where n_T is the total number of bosons in the configuration labeled by \mathbf{n} . Here we used the fact that when $n_T > 0$, $\langle 0 | \hat{c}_k \hat{G}_0(\omega) \hat{c}_q^\dagger \hat{B}_{i,\mathbf{n}}^\dagger | 0 \rangle = 0$. Defining $\tilde{\omega} \equiv \omega - n_T \Omega$ and adopting a combined real/momentum-space representation, we have

$$f_{\mathbf{n}}(\delta) = \mathcal{N}^{-3/2} \sum_i e^{\pm k R_i} \sum_q e^{-\pm q R_{i-\delta}} G_0(q, \tilde{\omega}) \sum_m e^{\pm q R_m} \langle 0 | \hat{c}_k \hat{G}(\omega) \hat{V} \hat{c}_m^\dagger \hat{B}_{i,\mathbf{n}}^\dagger | 0 \rangle. \quad (2.18)$$

The goal of the procedure is to extract a relationship between AGFs with n_T and $n_T \pm 1$ bosons. This depends on the specific form of \hat{V} . It is thus advantageous to express \hat{V} as defined in Eq. (2.10) to obtain

$$\hat{V} \hat{c}_m^\dagger \hat{B}_{i,\mathbf{n}}^\dagger | 0 \rangle = \sum_{(g,\psi,\phi,\xi)} g \sum_j \hat{c}_j^\dagger \hat{b}_{j+\phi}^\xi \hat{B}_{i,\mathbf{n}}^\dagger | 0 \rangle \delta_{m,j+\psi}. \quad (2.19)$$

Consider the case when $\xi = -$, implying the boson operator removes a boson from site $j+\phi$. Such a process can only have a non-zero contribution when a boson is removed from an occupied site, and the domain of sites where $\hat{b}_{j+\phi}$ can act in general is $j + \phi - i \in \Gamma_L^- = \{0, 1, \dots, L-1\}$. In this case, an extra prefactor appears due to the boson commutation relations, $\hat{b}_j \hat{b}_i^{\dagger m} = \hat{b}_i^{\dagger m} \hat{b}_j + m \delta_{ij} \hat{b}_i^{\dagger m-1}$.

Up until now, this derivation has been exact. We now impose a limit on the maximum cloud extent, M , restricting the cluster of sites where bosons can be created to at most M connected

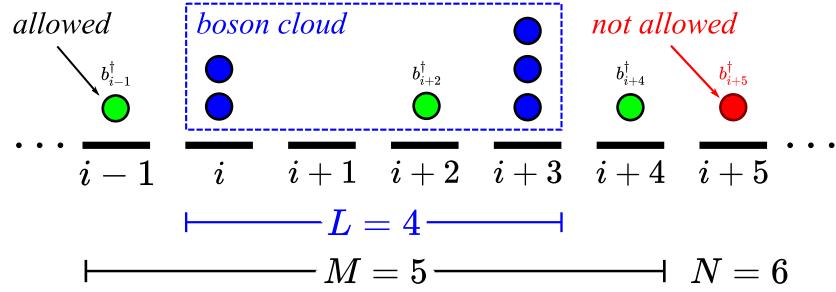


Figure 2.1: Cartoon of an $L = 4$ -site boson cloud, $\mathbf{n} = [2, 0, 0, 3]$ (blue), embedded within a variational space specified by a maximum cloud extent $M = 5$ and maximum number of bosons $N = 6$. In this example, the constraint $M = 5$ spans sites $i - 1$ to $i + 4$ so that bosons can be created only on these sites (e.g., green circles), and are not allowed outside of the M -site cloud (e.g., red circle). $N = 6$ implies that states with one more boson than those depicted in the figure (blue circles) are omitted from the variational space. Note that the carrier (not shown) is allowed to be anywhere on the chain.

sites, which are occupied with up to N bosons.⁴ Thus, when $\xi = +$, we have $j + \phi - i \in \Gamma_L^+ = \{L - M, L - M + 1, \dots, M - 1\}$. This restriction requires that we replace the sum over j with a sum over the elements of the aforementioned set: $\sum_j \rightarrow \sum_{\gamma \in \Gamma_L^\xi}$.

To continue the derivation of the EOM, we introduce the notation: $\hat{B}_{i,\mathbf{n}}^{(\xi,\gamma)\dagger} |0\rangle$ as the state $\hat{B}_{i,\mathbf{n}}^\dagger |0\rangle$ with an extra boson created ($\xi = +$) or destroyed ($\xi = -$) on site $i + \gamma$ within the permitted variational space specified by the above restriction. We omit states indexed by \mathbf{n} whose $n_T > N$ from the space of AGFs. Fig. 2.1 demonstrates the variational space encoded in our notation.

Summing over m and q in Eq. (2.18) produces the following general form,

$$f_{\mathbf{n}}(\delta) = \sum_{(g,\psi,\phi,\xi)} g \sum_{\gamma \in \Gamma_L^\xi} n^{(\xi,\gamma)} g_0(\delta + \gamma - \phi + \psi, \tilde{\omega}) \times \mathcal{N}^{-1/2} \sum_i e^{ikR_i} \langle 0 | \hat{c}_k \hat{G}(\omega) \hat{c}_{i+\gamma-\phi}^\dagger \hat{B}_{i,\mathbf{n}}^{(\xi,\gamma)\dagger} | 0 \rangle, \quad (2.20)$$

where $n^{(\xi,\gamma)}$ is a prefactor associated with applying a boson creation or annihilation operator: it is equal to 1 if $\xi = +$, and is equal to the number of bosons on site $i + \gamma$ (before a boson is annihilated)

⁴Note that the quasi-analytical formulation of the *approximate* MA methods represent a specific case of this general formalism in which $M = 1, 2$ or 3

if $\xi = -$.

Here, the free particle propagator in real space is given by [65]

$$g_0(\delta, \omega) = \frac{1}{N} \sum_q e^{iqR_\delta} G_0(q, \omega) = -\frac{i \left[-\omega_\eta/2t + i\sqrt{1 - (\omega_\eta/2t)^2} \right]^{|\delta|}}{\sqrt{4t^2 - \omega_\eta^2}}, \quad \omega_\eta \equiv \omega + i\eta. \quad (2.21)$$

Observe that the second line in Eq. (2.20) is precisely an AGF with different arguments and with $n_T \rightarrow n_T \pm 1$ bosons. Indexing a new AGF in the same manner as before we have

$$f_{\mathbf{n}}(\delta) = \sum_{(g, \psi, \phi, \xi)} g \sum_{\gamma \in \Gamma_L^\xi} n^{(\xi, \gamma)} g_0(\delta + \gamma - \phi + \psi, \tilde{\omega}) f_{\mathbf{n}}^{(\xi, \gamma)}(\phi - \gamma). \quad (2.22)$$

Finally, we note that in order to abide by our labeling convention, certain "reduction rules" for the AGFs must be followed in order to produce a valid closure. When removing or adding bosons, as in the case $f_{\mathbf{n}} \rightarrow f_{\mathbf{n}}^{(\xi, \gamma)}$, additional phase prefactors may appear. The details of these rules are summarized in Appendix 2.A (see also Ref. 56 for a specific example).

2.2.2 Implementation

Together, Eqs. (2.13) and (2.22), along with the rules in Appendix 2.A, contain all information necessary to solve for $G(k, \omega)$ for some chosen values of M, N . In this section, we describe the *computational* approach for representing these equations and solving them numerically.

Every possible combination of $1 \leq n \leq N$ bosons on $1 \leq L \leq M$ sites will contribute to the calculation of $G(k, \omega)$. In the first step, we systematically generate all combinations, noting the only requirement that the first and last sites for some cloud extent L must be at least singly occupied. This amounts to symbolically constructing and storing representations of these objects, e.g.

$$\mathcal{G} = \{f_{[0]}(\delta), f_{[1]}(\delta), f_{[1,1]}(\delta), f_{[1,0,2]}(\delta), \dots\}, \quad (2.23)$$

such that all possible AGFs corresponding to a given configuration are generated. This can be

thought of precisely as the classic combinatorics problem of N indistinguishable balls in M distinguishable bins, with the added constraint of requiring at least one boson on each end of the cloud. In this way, the total number of equations generated at this step (the total number of elements in \mathcal{G} , defined as $|\mathcal{G}|$) has a straightforward representation,

$$|\mathcal{G}| = 1 + \sum_{L=1}^M \sum_{n=1}^N \begin{cases} 1 & \text{if } L = 1 \text{ or } n = 2 \\ \binom{L+n-3}{n-2} & \text{otherwise} \end{cases}, \quad (2.24)$$

where the one extra term reflects the first equation in the set of equations (for $G(k, \omega)$).

The second step consists of finding the values for δ each function $f_{\mathbf{n}}$ requires. Observing that the only δ -dependence on the RHS of Eq. (2.22) is contained in g_0 (and importantly not in $f_{\mathbf{n}}$), we obtain the full closure of equations by finding, for every $f_{\mathbf{n}}$, the values of δ prescribed by the indices $\phi - \gamma$ on the RHS. This set is informally denoted as \mathcal{S} , e.g.,

$$\mathcal{S} = \{f_{[0]}(-1), f_{[0]}(0), f_{[1]}(-1), \dots\}. \quad (2.25)$$

The terms contained in \mathcal{S} are determined by a nontrivial function of M, N and depend on the model type. Every term in \mathcal{S} is simply a specific case of the LHS of Eq. (2.22).

In the final step, we formulate this as an inhomogeneous linear system of equations and aim to find the solution for all $f_{\mathbf{n}}(\delta)$ for some values of k, ω, M, N ,

$$A\mathbf{f} = \mathbf{b}. \quad (2.26)$$

Above, A is a matrix of coefficients which can be read from the aforementioned equations, and \mathbf{b} is proportional to the unit vector and inherits the inhomogeneity of Eq. (2.13). This matrix equation can be solved in one of two ways. The solution for \mathbf{f} can be obtained in a single step, which amounts to applying some direct solver to the $|\mathcal{S}| \times |\mathcal{S}|$ matrix A . However, this approach is either inefficient (using a sparse solver) or intractable using a dense solver due to the large size

of A in cases such as the extreme adiabatic limit. Alternatively, we find that a continued fraction approach using dense linear algebra provides the optimal middle ground. Formally, the continued fractions $V_n = \mathcal{A}_n(\mathbf{k}, \omega)V_{n-1} + \mathcal{B}_n(\mathbf{k}, \omega)V_{n+1}$, where $\mathcal{A}_n(\mathbf{k}, \omega)$ and $\mathcal{B}_n(\mathbf{k}, \omega)$, are sparse matrices read off directly from the EOM, and V_n is a vector of AGF's with $n \leq N$ bosons [51, 56]. The matrix inversions required are much smaller in this approach, although there are $\Theta(N)$ of them. We note that using this more efficient approach, the calculations become challenging in our current implementation only around $(M, N) \sim (10, 7)$, which produces $\sim 60k$ equations. Adding one more boson balloons the calculation to $\sim 150k$ equations, which are in principle within reach on large supercomputer architectures with sufficient memory capacity.

To approach the infinite phonon Hilbert space limit using the continued fraction approach, we set $V_{N+1} = 0$, solving the set of equations until we obtain $G(\mathbf{k}, \omega)$, which corresponds to V_0 . In the $N \rightarrow \infty$ limit, this represents a sensible boundary condition because it becomes energetically expensive to generate clouds with larger than N bosons. In practice we treat N as a convergence parameter. All results shown in this work appear to be converged with respect to N to desirable accuracy, unless otherwise stated.

2.2.3 Comparison to Other Methods

Comparison to related methods: Momentum Average (MA) and Limited Phonon Basis Exact Diagonalization (LPBED) methods.

The GGCE method combines advantages from the MA and Limited Phonon Basis Exact Diagonalization [41] (LPBED) methods. In the MA approach, one makes an educated guess of the value of M needed to obtain accurate results, in essence employing a variational ansatz to the EOM. One then derives the EOM in $MA(M)$ analytically "by hand" and solves for $G(k, \omega)$ numerically. LPBED is a more general ED analog of MA, and in principle also relies on a variational ansatz, albeit one different from that of MA. Another successful version of LPBED [66] discussed in the literature included clouds of size $M = 5$ whilst allowing for two extra bosons anywhere on the

lattice even away from the cloud, but with a more restricted total number of bosons.⁵

We can roughly view MA and LPBED methods as specific variational cases of the GGCE, which benefits from allowing an arbitrary systematic choice of maximal cloud extent, M , in the $N \rightarrow \infty$ limit. The GGCE thus serves as a systematically exact method which allows one to tailor resources based on the underlying physics of the problem, and is limited only by computational resources. This provides the potential to access regimes that are difficult to quantitatively describe by other approaches, as we show below.

Comparison to Variational Exact Diagonalization (VED) methods.

Variational Exact Diagonalization (VED) [38] represents another class of successful approaches to the polaron problem. In VED, a variational Hilbert space is iteratively generated by applying the off-diagonal parts of the Hamiltonian to a reference state taken to be a Bloch state of an electron and zero bosons in an infinite system. After N_h iterations, one diagonalizes the Hamiltonian in the generated basis using standard Lanczos techniques. Convergence with respect to N_h , when possible, guarantees access to the exact ground state and a small manifold of low-lying excited states [39]. There are at least two main differences between GGCE and VED.

First, VED naturally imposes a restriction on the distance between the electron and phonon configurations, which can be at most $\sim N_h$ sites (the precise value depends on the coupling), while GGCE (and MA [56]) includes states with the electron arbitrarily far away from the phonon clouds with no restriction (this can be seen from the application of $\hat{G}_0(\omega)$ in the EOM on states in AGFs with both an electron and phonons, which moves the electron arbitrarily in the system without regard to the location of the bosonic cloud, c.f. Eq. (2.20)). We note that VED is capable of describing the ground and low-lying excited states in the weak- and lower-intermediate regimes of coupling in the adiabatic limit [38]. We suspect that the restriction on the distance between the electron and the phonons in VED prohibits access to very strong couplings in the adiabatic regime and to continuum states since these are generally delocalized states (see discussion below). In

⁵We note that it appears the largest values for M used in the MA method is 3, while the largest values for M used in the LPBED method is 5.

contrast, as we show below, GGCE can tackle strong coupling in the adiabatic regime.

Second, GGCE is formulated as an expansion in terms of cloud sizes, and the computation must be converged with respect to the cloud size, while VED imposes no restriction on cloud sizes (a cloud in VED can extend over, at most, $\sim N_h$ sites). For example, $N_h = 11$ implies clouds extended over $\sim 10\text{-}11$ sites (the exact number depends on the specific model of the electron-boson coupling). Such a value of N_h represents a rough lower bound within what is typically used in VED in the intermediate adiabatic limit. These values imply clouds with sizes that are much larger than those used in GGCE in the current work. This suggests that GGCE may benefit in terms of efficiency by employing a smaller number of states resulting from smaller clouds without compromising accuracy. We believe this is a direct result of using an EOM formulation of propagators, which ensures we keep only those states generated in the dynamics and nothing further. Comparing, empirically, to Ref. 38, we note that the number of states needed in GGCE appears to be two orders of magnitude smaller than those in VED in order to achieve convergence in similar parameter regimes.

Finally, we note that other variants of VED with extra restrictions on the variational space have been used with great success [39, 67, 68]. These, however, are either not formulated in a general enough manner to be applied to a generic form of electron-boson coupling [67] or involve further constraints that, while variational, are not completely motivated physically especially at strong couplings. In contrast, GGCE in its current form follows naturally from the EOM and has no restrictions beyond the cloud size, which is taken to the infinite limit sequentially and in an efficient manner. In principle further restrictions of this type can be imposed in our GGCE, but we do not explore this direction in the current manuscript.

The preliminary analysis presented here suggests that GGCE may perform more favorably than related approaches, at least in some parameter regimes and for some quantities. Future work must be devoted to address these issues and compare the range of variational restricted-basis approaches over the full range of parameter space for both ground-state energies and spectral functions to fully access the utility and efficiency of each approach.

2.3 Results

In this section, we show results for a variety of 1D lattice models described by the Hamiltonian defined by Eqs. (2.8) and (2.10). This allows us to both benchmark GGCE against exact numerics, and to tackle regimes typically inaccessible even in the well-studied 1D limit. In what follows, we characterize the interaction strength via the dimensionless coupling constant

$$\lambda = E_{\text{GS}}(t=0)/E_{\text{GS}}(\alpha=0), \quad (2.27)$$

which is the ratio of the ground state (GS) energy in the atomic limit to that in the free particle limit, and the adiabaticity ratio

$$\Lambda = \Omega/W, \quad (2.28)$$

where $W = 4t$ is the carrier's bandwidth.

While DMC and other quantum Monte Carlo methods may access the GS in the adiabatic limit, dynamics are generally difficult to obtain due to uncertainties associated with analytical continuation to the real-frequency axis. We showcase the ability of the GGCE to simulate dynamics in the low-energy regime for the Holstein [20, 21] (H) and Peierls (P) (also known as the Su-Schrieffer-Heeger [69]) models. Finally, we study an experimentally motivated mixed Holstein+Peierls (HP) model in which the carrier couples to two different boson modes, one describes a Holstein coupling and the other a Peierls coupling.

2.3.1 Holstein Model

We first consider the prototypical Holstein model [20, 21] for which

$$\hat{V} = \alpha \sum_i \hat{c}_i^\dagger \hat{c}_i (\hat{b}_i^\dagger + \hat{b}_i), \quad \lambda_H \equiv \alpha^2/2\Omega t. \quad (2.29)$$

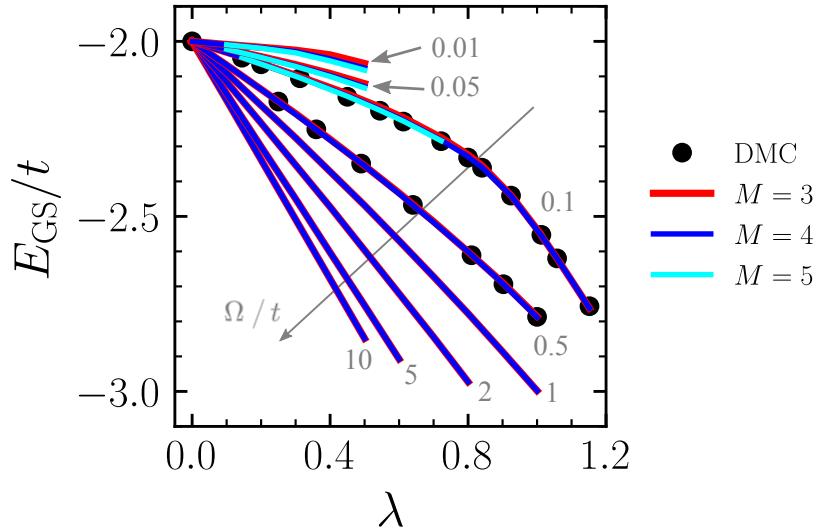


Figure 2.2: Ground state energy E_{GS}/t as a function of coupling strength $\lambda = \alpha^2/2\Omega t$ for the Holstein model in adiabaticity limits extending from anti- ($\Lambda \gg 1$) to extreme-adiabatic ($\Lambda \ll 1$). Values for Ω/t are shown in grey, and GGCE results for $\Omega/t = 0.1$ and 0.5 are compared to DMC data (symbols) [70]. Ground state peak locations are converged with respect to N , and generally require ≈ 10 bosons at small couplings, but up to ≈ 30 at large couplings.

In Fig. 2.2, we compute the GS energy of a Holstein polaron for $\Lambda \in [0.0025, 2.5]$. For $\Omega/t = 0.1$ and 0.5 , we compare our results to those obtained via Diagrammatic Monte Carlo (DMC) [70]. Not only does GGCE converge to the exact result for $\lambda \in [0, 1.2]$, but it also yields slightly lower GS energies than DMC in the strong-coupling regime $\lambda \gtrsim 1$, although the differences are likely due to statistical errors in DMC.⁶ Importantly, we are able to converge our results to the exact limit even at extremely small $\Omega/t \in [0.01, 0.1]$ for intermediate coupling strengths $\lambda \gtrsim 0.5$, overcoming previous limitations of momentum average methods. Beyond demonstrating GGCE’s ability to simulate the adiabatic limit of massive bosons, our results show a trend at intermediate couplings of the polaron binding energy $|E_{\text{GS}}(\lambda) - E_{\text{GS}}(0)|$ that monotonically decreases with Ω .

To demonstrate the ability of the GGCE method to converge spectral functions and probe a broad range of physical regimes, we present an array of spectral functions in Fig. 2.3. These results cover all combinations of $\Omega/t \in \{0.1, 0.5\}$, $k \in \{0, \pi/2, \pi\}$ and $\lambda \in \{0.2, 0.5, 0.8\}$ and highlight

⁶We do not have access to statistical error bars in the DMC calculations.

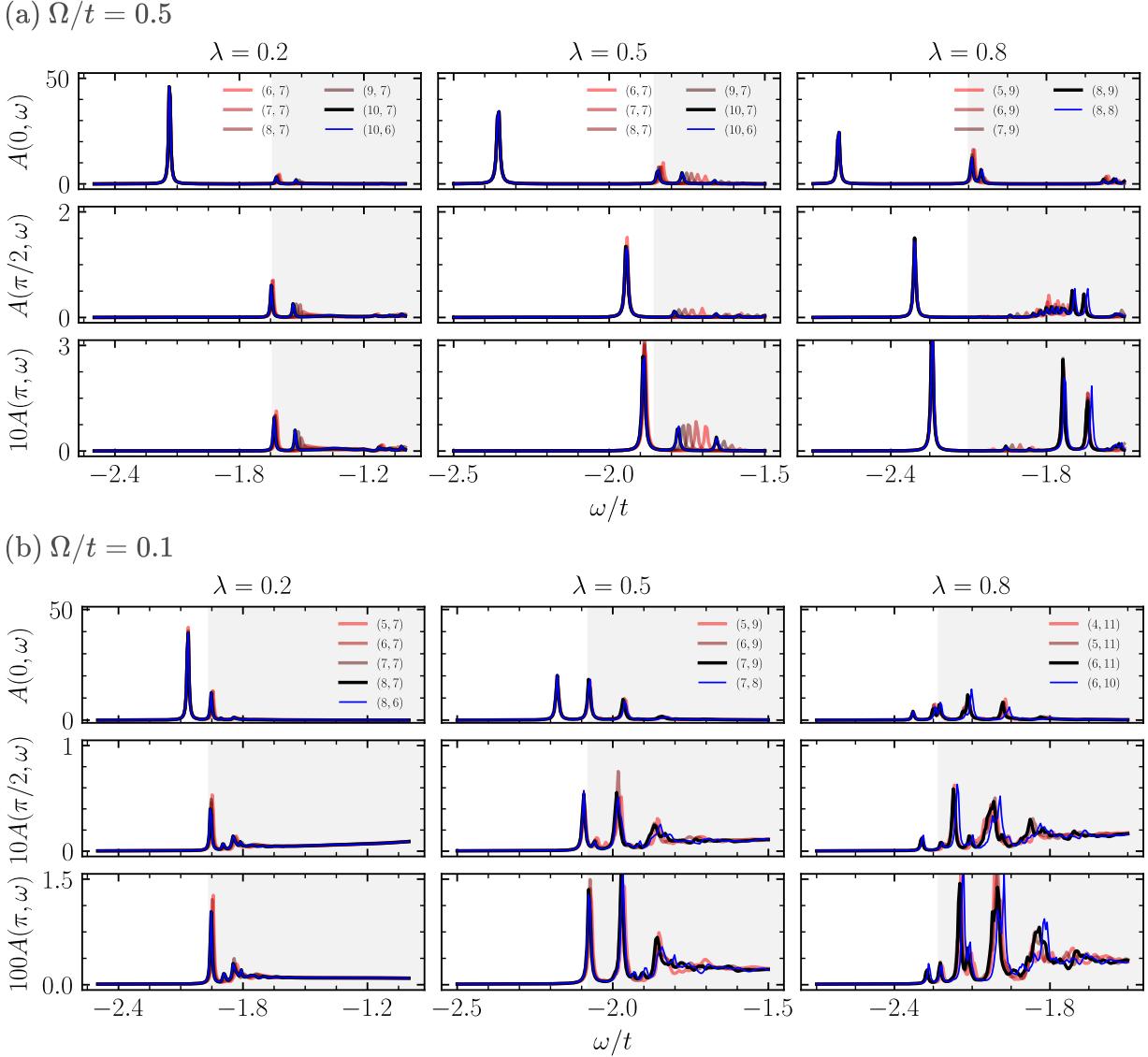


Figure 2.3: Spectral function $A(k, \omega)$ at $k = 0, \pi/2$ and π for the Holstein model in (a) the mildly adiabatic ($\Omega/t = 0.5$) and (b) adiabatic limits ($\Omega/t = 0.1$), for dimensionless couplings $\lambda = 0.2, 0.5$ and 0.8 . In this figure we use $\eta = 0.005$. Various values of (M, N) are shown in the legend. Specifically, the gradient from red to black shows convergence with respect to M for fixed N . We demonstrate convergence with respect to N via the blue line, which shows results that use the largest used M , but with one less boson than the largest-used N . The onset of the continuum is shown in shaded gray, and is defined as $E_{\text{GS}} + \Omega$, where E_{GS} is the polaron ground-state energy.

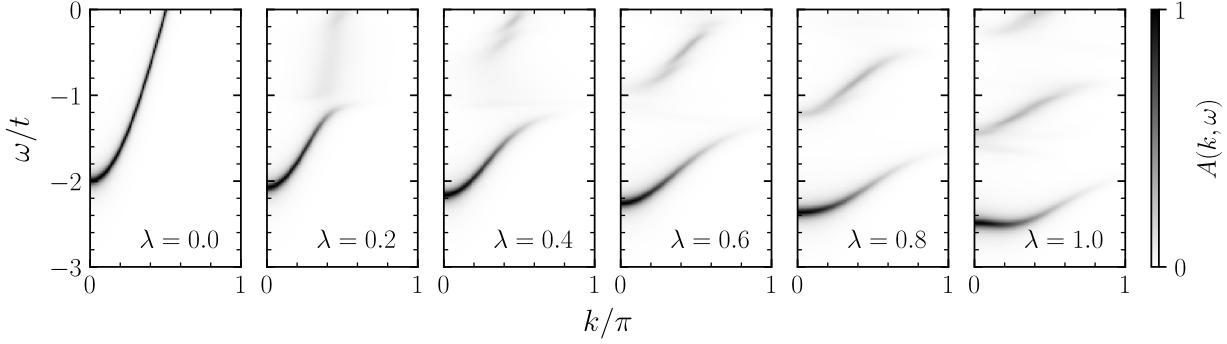


Figure 2.4: Spectral function $A(k, \omega)$ (scaled to a maximum of 1) of the Peierls model for $\Omega/t = 1$, $\eta = 0.05$ and various values of the dimensionless coupling strength, λ . $M = 5$ and $N = 10$ used here are sufficient for convergence of the bands on the scale of the plot. It is worth noting that fine structure in states above the lowest energy band can be resolved on a finer grid and smaller value of η , although the intensity of these states is at most roughly an order of magnitude smaller than that of the lowest energy band.

the potential of the method. For example, in both cases treated in Fig. 2.3, we find excellent convergence of the ground state peak location and structure. The first excited state, which for the values of λ considered, lies in the polaron + one boson continuum, proves more difficult to converge. Nonetheless, we show reasonable convergence of this second peak for a wide range of parameters. However, convergence becomes more challenging for $\Omega/t = 0.5$ at $\lambda = 0.5$, as seen in the second column of Fig. 2.3(a), even when using extremely large cloud sizes ($M = 10$), and as a result this peak is not sufficiently converged. Difficulty in resolving excitations above $E_{\text{GS}} + \Omega$ is not surprising, since the nature of these continuum states involves scattering between a delocalized electronic state and an extended cloud of phonons that is generally not small. As such, a sufficiently large cloud and therefore a bigger variational space is needed for convergence. Thus, with increasing computational resources, convergence of the spectral function proceeds naturally from low to high energy. This implies that one can readily achieve convergence of lower-energy states with much ease.

2.3.2 Peierls Model

In Fig. 2.4, we present exact spectral functions of a polaron in the Peierls model [69, 71–73] defined by

$$\hat{V} = \alpha \sum_{\langle ij \rangle} (\hat{c}_i^\dagger \hat{c}_j + \text{h.c.}) (\hat{b}_i^\dagger + \hat{b}_i - \hat{b}_j^\dagger - \hat{b}_j), \quad \lambda_P = 2\alpha^2/\Omega t, \quad (2.30)$$

for a variety of different dimensionless couplings. Although in principle no more difficult than for the case of the Holstein model, we reserve exploring the extreme-adiabatic limit ($\Omega/t \ll 1$) to future work and show results only for $\Omega/t = 1$.

The Peierls model exhibits distinct polaron physics when compared with the Holstein model. A Peierls polaron exhibits a sharp transition from a state with $k_{GS} = 0$ to one with $k_{GS} \neq 0$ for $\lambda > \lambda_c(\Omega/t)$ [55], while a Peierls bipolaron exhibits a significantly smaller mass than its Holstein counterpart [59] and can exhibit transitions under certain conditions [74]. We observe the transition to a band minimum at a finite wave vector in Fig. 2.4 as λ changes from $\lambda = 0.8$ to $\lambda = 1$, consistent with Ref. 55. Importantly, we are able to resolve the spectrum above the ground state within sufficient accuracy. The excited states of this model play an important role in presence of other perturbations, as will become apparent next.

2.3.3 Mixed-Boson Mode Holstein+Peierls Model

We now consider a realistic model applicable to organic crystals, molecular complexes, etc., in which the charge carrier couples to both Holstein and Peierls phonon modes, each with its own frequency [75–78]. The Hamiltonian is given by

$$\begin{aligned} \hat{H} = & -t \sum_{\langle ij \rangle} \hat{c}_i^\dagger \hat{c}_j + \Omega_H \sum_i \hat{h}_i^\dagger \hat{h}_i + \Omega_P \sum_i \hat{p}_i^\dagger \hat{p}_i \\ & + \alpha_H \sum_i \hat{c}_i^\dagger \hat{c}_i (\hat{h}_i^\dagger + \hat{h}_i) + \alpha_P \sum_{\langle ij \rangle} (\hat{c}_i^\dagger \hat{c}_j + \text{h.c.}) (\hat{p}_i^\dagger + \hat{p}_i - \hat{p}_j^\dagger - \hat{p}_j), \end{aligned} \quad (2.31)$$

where $\hat{h}_i \equiv \hat{b}_i^H$, $\hat{p}_i \equiv \hat{b}_i^P$ and the Holstein and Peierls boson operators act on *different* boson Hilbert spaces. We note that the combinatorics of multi-phonon models require vastly more resources than single mode cases. Here, $\lambda_H = \alpha_H^2/2\Omega_H t$ and $\lambda_P = 2\alpha_P^2/\Omega_P t$, as before.

First, we detail the differences between this HP model and that presented in Ref. 57. The latter model represents a toy model of a carrier coupled to *one boson type*, with two coupling contributions: diagonal (Holstein) and off-diagonal (Peierls). Computations for this type of model possess the same scaling complexity as that for H or P models, making it much easier to converge. However, a realistic calculation requires modeling couplings to multiple phonon modes, typically of vastly different energies, characteristic of experimental systems. A straightforward generalization of our implementation allows us to treat the boson modes as explicitly distinguishable, even when $\Omega_P = \Omega_H$. We simply introduce two types of bosonic clouds, one for Holstein bosons with M_H and one for Peierls bosons with M_P . These can overlap, and we thus need an extra variational parameter to constraint the absolute extent, A , over which the combined clouds extend. We detail this construction in Appendix 2.C. This approach allows us to explore this more experimentally relevant model. As previously mentioned, this comes with the downside of increased computational complexity. However, as we show below, we are able to semi-quantitatively converge the lowest-energy band for reasonably large couplings, and, with modest computational resources, we resolve the spectrum in the experimentally relevant regime (see Fig. 2.5), $\Omega_P < \Omega_H$, as well as in a hypothetical scenario with the frequencies reversed. This requires modest choices of M_H , M_P and A . In Fig. 2.5), in the more experimentally relevant case, with $\Omega_H/t = 2.5$ and $\Omega_P/t = 0.5$, we see a non-negligible bandwidth and thus significant P-like character, an important observation for experiment. In our simulations of this model, we also find interesting behavior in the second peak in the spectrum involving a minimum away from $k = 0$ (not shown), which we leave to a future detailed analysis.

We quantify the ground state convergence as a function of the individual maximum cloud extents M_H , M_P , the absolute cloud extent A , and maximum number of bosons in the variational space, N_H, N_P in Fig. 2.6. This analysis suggests that an increase of computational resources,

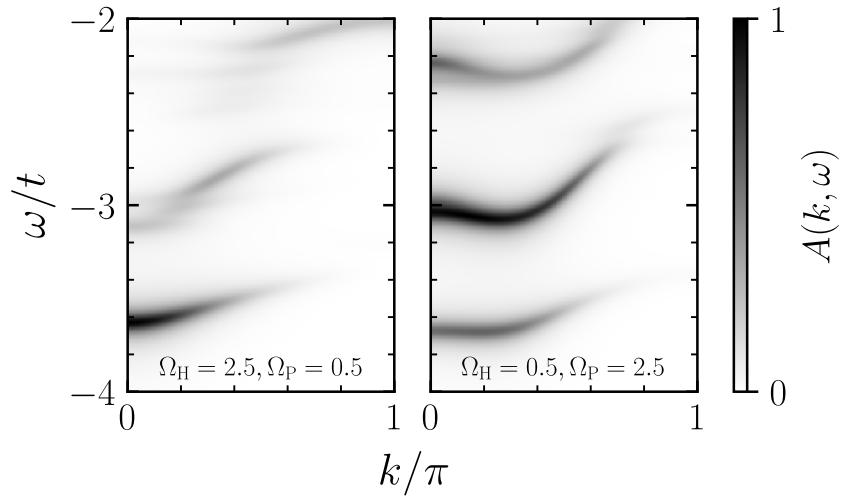


Figure 2.5: Spectral function $A(k, \omega)$ (scaled to a maximum of 1) of the mixed-boson mode Holstein+Peierls model for various values of Ω_H and Ω_P , $\lambda_H = \lambda_P = 1$, $t = 1$ and $\eta = 0.05$. For these calculations, we use $M_H = M_P = 3$, and a maximum total cloud length, or absolute extent, $A = 3$ (see Appendix 2.C), and $N_H = N_P = 5$., for which semi-quantitative convergence is achieved.

within reach on large computers, will permit complete convergence.

2.4 Conclusions

We have presented an exact, general approach to solving the EOM of a Green's function of a particle dressed by bosons, suitable for treating difficult regimes such as the adiabatic limit, and have demonstrated the power of the approach by calculating the polaron ground state and spectral functions in coupling regimes ranging from weak to strong, and adiabaticity limits ranging from extreme anti-adiabatic to extreme adiabatic. We note that at large couplings, the GGCE achieves ground state energies in agreement with DMC (Fig. 2.2), without the introduction of stochastic error. Exact simulated spectra for $\Lambda \ll 1$ are, in general, difficult to achieve with Monte Carlo methods due to the reliance on analytic continuation, and inaccessible to most Exact Diagonalization methods due to the large basis size needed for convergence.

We emphasize the success achieved by the MA method in characterizing polarons and bi-

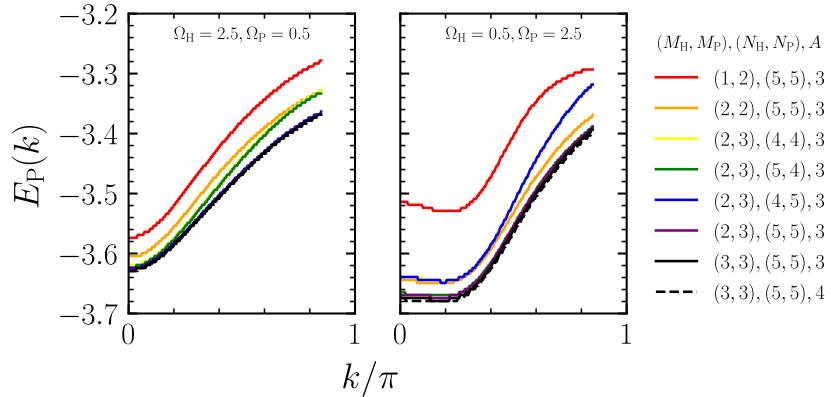


Figure 2.6: Convergence of the energy of the ground-state polaron band, $E_P(k)$, for parameters shown in Fig. 2.5 against combinations of M_H , M_P , A , N_H and N_P .

larons in various systems under different physical conditions of experimental relevance. In most of these cases, verification of the accuracy of the method against an exact approach was needed to justify *a posteriori* its utility and potential in limits where exact numerics are difficult to obtain, e.g., in higher dimensional systems. The GGCE method systematically makes use of the MA hierarchy, resulting in an exact yet efficient approach, and a new physically motivated expansion in orders of the boson cluster size, thus expanding the horizon of possibilities in characterizing dressed quasiparticles in previously challenging regimes. Finally, the GGCE computational framework is well-suited for future practical extensions, including higher-dimensional systems, finite-temperature studies, the computation of observables connected to higher-order Green's functions, such as optical spectra and polaron mobilities [79], as well as studies of the dynamics of bipolarons [80], and in other contexts we plan to address in future work.

2.A Reduction Rules for AGFs

In this Appendix, we detail the reduction rules the AGFs follow in order to produce a valid EOM.

Annihilating or creating a boson to the right of the last occupied site does not come with any

additional rule for re-indexing:

$$f_{[n,n',\dots,n'',0,\dots,0,1]}(\delta) \xrightarrow{\hat{b}} f_{[n,n',\dots,n'',0,\dots,0]}(\delta) = f_{[n,n',\dots,n'']}(\delta), \quad (2.32)$$

$$f_{[n,n',\dots,n'']}(\delta) \xrightarrow{\hat{b}^\dagger} f_{[n,n',\dots,n'',0,\dots,0,1]}(\delta) = f_{[n,n',\dots,n'',0,\dots,0,1]}(\delta), \quad (2.33)$$

where here $n, n'' > 0$.

However, when creating or annihilating a boson to the left of the first occupied site on the chain, we must re-index the state such that the label i always references the first occupied site:

$$f_{[1,0,\dots,0,n,n',\dots,n'']}(\delta) \xrightarrow{\hat{b}} f_{[0,\dots,0,n,n',\dots,n'']}(\delta) \rightarrow e^{-ikR_z} f_{[n,n',\dots,n'']}(\delta + z), \quad (2.34)$$

$$f_{[n,n',\dots,n'']}(\delta) \xrightarrow{\hat{b}^\dagger} f_{[1,0,\dots,0,n,n',\dots,n'']}(\delta) \rightarrow e^{ikR_z} f_{[1,0,\dots,0,n,n',\dots,n'']}(\delta - z), \quad (2.35)$$

where z is the number of shifted sites $i \rightarrow i \pm 1 \rightarrow i \pm 2, \dots$ in the phase incurred.

2.B Examples of the Generalized Notation

In this work, we considered H, P and HP models, each of which have different carrier-boson couplings, \hat{V} . Within the framework of the GGCE, these differences amount to a simple change in input parameters. The fully expanded coupling terms \hat{V} , and their representation in terms of the notation defined in Eq. (2.10), are shown here. We present the three models used and reference the derivation as performed in Section 2.2. First, recall that the vectors which represent the coupling are notated as (g, ψ, ϕ, ξ) .

In the H model, this notation translates to

$$\hat{V}_H = \underbrace{\alpha \sum_i \hat{c}_i^\dagger \hat{c}_i \hat{b}_i^\dagger}_{(\alpha, 0, 0, +)} + \underbrace{\alpha \sum_i \hat{c}_i^\dagger \hat{c}_i \hat{b}_i}_{(\alpha, 0, 0, -)}. \quad (2.36)$$

In the P model, we have

$$\begin{aligned} \hat{V}_P = & \underbrace{\alpha \sum_i \hat{c}_i^\dagger \hat{c}_{i+1} \hat{b}_i^\dagger}_{(\alpha, 1, 0, +)} + \underbrace{\alpha \sum_i \hat{c}_i^\dagger \hat{c}_{i+1} \hat{b}_i}_{(\alpha, 1, 0, -)} - \underbrace{\alpha \sum_i \hat{c}_i^\dagger \hat{c}_{i+1} \hat{b}_{i+1}^\dagger}_{(-\alpha, 1, 1, +)} - \underbrace{\alpha \sum_i \hat{c}_i^\dagger \hat{c}_{i+1} \hat{b}_{i+1}}_{(-\alpha, 1, 1, -)} \\ & + \underbrace{\alpha \sum_i \hat{c}_i^\dagger \hat{c}_{i-1} \hat{b}_{i-1}^\dagger}_{(\alpha, -1, -1, +)} + \underbrace{\alpha \sum_i \hat{c}_i^\dagger \hat{c}_{i-1} \hat{b}_{i-1}}_{(\alpha, -1, -1, -)} - \underbrace{\alpha \sum_i \hat{c}_i^\dagger \hat{c}_{i-1} \hat{b}_i^\dagger}_{(-\alpha, -1, 0, +)} - \underbrace{\alpha \sum_i \hat{c}_i^\dagger \hat{c}_{i-1} \hat{b}_i}_{(-\alpha, -1, 0, -)}. \end{aligned} \quad (2.37)$$

The case of the HP model is a bit more elaborate, since the model involves different boson operators: $\hat{h}_i \equiv \hat{b}_i^{(\Omega_H)}$ and $\hat{p}_i \equiv \hat{b}_i^{(\Omega_P)}$. Thus, we have

$$\begin{aligned} \hat{V}_{HP} = & \underbrace{\alpha_H \sum_i \hat{c}_i^\dagger \hat{c}_i \hat{h}_i^\dagger}_{(\alpha_H, 0, 0, +)} + \underbrace{\alpha_H \sum_i \hat{c}_i^\dagger \hat{c}_i \hat{h}_i}_{(\alpha_H, 0, 0, -)} + \underbrace{\alpha_P \sum_i \hat{c}_i^\dagger \hat{c}_{i+1} \hat{p}_i^\dagger}_{(\alpha_P, 1, 0, +)} + \underbrace{\alpha_P \sum_i \hat{c}_i^\dagger \hat{c}_{i+1} \hat{p}_i}_{(\alpha_P, 1, 0, -)} \\ & - \underbrace{\alpha_P \sum_i \hat{c}_i^\dagger \hat{c}_{i+1} \hat{p}_{i+1}^\dagger}_{(-\alpha_P, 1, 1, +)} - \underbrace{\alpha_P \sum_i \hat{c}_i^\dagger \hat{c}_{i+1} \hat{p}_{i+1}}_{(-\alpha_P, 1, 1, -)} + \underbrace{\alpha_P \sum_i \hat{c}_i^\dagger \hat{c}_{i-1} \hat{p}_{i-1}^\dagger}_{(\alpha_P, -1, -1, +)} + \underbrace{\alpha_P \sum_i \hat{c}_i^\dagger \hat{c}_{i-1} \hat{p}_{i-1}}_{(\alpha_P, -1, -1, -)} \\ & - \underbrace{\alpha_P \sum_i \hat{c}_i^\dagger \hat{c}_{i-1} \hat{p}_i^\dagger}_{(-\alpha_P, -1, 0, +)} - \underbrace{\alpha_P \sum_i \hat{c}_i^\dagger \hat{c}_{i-1} \hat{p}_i}_{(-\alpha_P, -1, 0, -)}. \end{aligned} \quad (2.38)$$

2.C Additional Notation for Mixed-Boson Mode HP Models

In Subsection 2.3.3, and specifically Fig. 2.5, we introduced new notation required to define the configuration space of the HP model. First, the occupation number vector \mathbf{n} is now a two-row matrix, \bar{n} , where as usual the columns index the site index starting with i , and the two rows correspond to the occupation numbers of the Holstein and Peierls bosons. For clarity, we label the first row n_H and the second n_P . The logic presented in Section 2.2 still applies in for the HP model: \hat{V} can still create or destroy only a single boson at a time, $\hat{B}_{i,\bar{n}}$ and corresponding objects now reference both sets of boson occupation numbers (and boson operators now carry a boson-type index), $\tilde{\omega} \equiv \omega - \Omega_H \sum_j n_H^{(j)} - \Omega_P \sum_j n_P^{(j)}$, etc. As before, the left-most occupied site is still the anchor for

the entire cloud, and thus the same reduction rules in Appendix 2.A apply.

In terms of the configuration space, we now limit the maximum number of Holstein and Peierls bosons individually, using N_H and N_P , respectively, and the extent of the clouds individually, using M_H and M_P , respectively. Given there are now two "overlapping" clouds of bosons which live in different Hilbert spaces, we must define yet another configuration space parameter, which we call the absolute extent, A . This is the maximum extent of the cloud measured from the site index of the left-most boson to the site index of the right-most boson, regardless of boson type. Note that we converged results in Fig. 2.5 with respect to A as well as the other four convergence parameters. We present an exemplary configuration space in Fig. 2.7 to further highlight the aforementioned definitions.

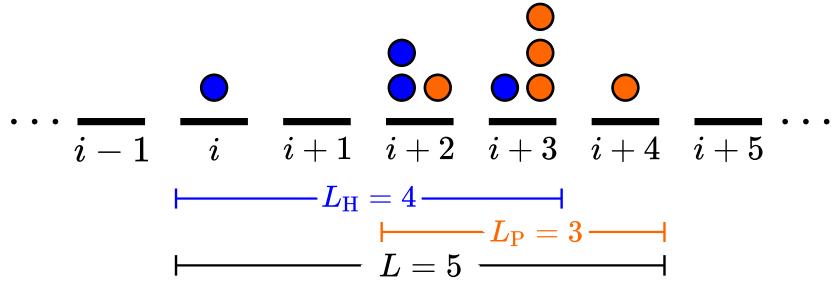


Figure 2.7: Example of a configuration of HP bosons corresponding to $n_H = [1, 0, 2, 1, 0]$ and $n_P = [0, 0, 1, 3, 1]$. Similar to the single boson models, we require that $\sum_j n_H^{(j)} \leq N_H$, $\sum_j n_P^{(j)} \leq N_P$, $L_H \leq M_H$, $L_P \leq M_P$ and $L \leq A$.

2.D Finite-temperature Thermofield Dynamics

The GGCE methodology outlined so far in this chapter allows for access to exact Green's functions in lattice models, at least in principle. By leveraging an expansion in boson cloud size, we take advantage of the physically motivated ansatz that boson clouds are created in clusters. However, the GGCE method comes with the drawback that it is apparently limited to zero-temperature Green's functions. In this Chapter, we show an *exact* finite-temperature formalism for Green's functions compatible with the GGCE using a framework called Thermofield Dynamics (TFD) [81, 82].

Consider the usual problem of finding some observable at a finite temperature. Quantum-mechanically, this is formulated rigorously as the thermal trace of some general time-dependent operator,⁷

$$\langle A(t) \rangle_\beta = \text{Tr}\{A(t)\rho(\beta)\} = \text{Tr}\{e^{\pm iHt} A e^{-\pm iHt} \rho(\beta)\} \equiv \sum_n \langle \psi_n | e^{\pm iHt} A e^{-\pm iHt} \rho(\beta) | \psi_n \rangle \quad (2.39)$$

where $A \equiv A(0)$ is the operator of interest at $t = 0$, and $\rho(\beta)$ is the initial equilibrium density matrix of the full system (corresponding to Hamiltonian H),

$$\rho(\beta) = \frac{1}{Z(\beta)} e^{-\beta H}. \quad (2.40)$$

The denominator is known as the canonical partition function,

$$Z(\beta) = \sum_n \langle \psi_n | e^{-\beta H} | \psi_n \rangle. \quad (2.41)$$

The trace runs over the full set of states of the system, $|\psi_n\rangle$, in a complete basis of choice. Often computing this sum is expensive and/or cumbersome; TFD attempts to write the trace *like* an expectation value over a pure state. In other words, the first question is, can the trace be rewritten as a single expectation value over thermal states,

$$\langle A(t) \rangle_\beta \stackrel{?}{=} \langle \bar{0}(\beta) | A(t) | \bar{0}(\beta) \rangle. \quad (2.42)$$

The states $|\bar{0}(\beta)\rangle$ are known as the "thermal vacuum" for reasons that will be explained later. The second question, which will ultimately showcase the majority of the utility of the TFD method, is can we write this thermal state as a unitary rotation from some temperature-independent vacuum,

$$|\bar{0}(\beta)\rangle \stackrel{?}{=} e^{-\pm iG(\beta)} |\bar{0}\rangle. \quad (2.43)$$

⁷Note that we no longer label operators with hats from this point on.

The states $|\bar{0}\rangle$ are known as the "doubled vacuum," again for reasons that will be explained later.

In the following subsections and Appendix 2.E, we will show how the representations in Eqs. (2.42) and (2.43) can be obtained and used to great effect in the framework of the GGCE method.

2.D.1 The Thermofield Double

We first set out to prove the thermal expectation can be written in the form of Eq. (2.42). First, due to the cyclic property of the trace, we can symmetrize the expectation value (and drop time-dependence for now, A can always be time-dependent),

$$\langle A \rangle_\beta = \frac{1}{Z(\beta)} \sum_n \langle \psi_n | e^{-\beta H/2} A e^{-\beta H/2} | \psi_n \rangle. \quad (2.44)$$

Consider that the states $|\psi_n\rangle$ live in a Hilbert space \mathcal{H} . We define a second, "doubled" space $|\tilde{\psi}_n\rangle$ which lives in a second Hilbert space $\tilde{\mathcal{H}}$, such that operators that act on states in \mathcal{H} have no effect on states in $\tilde{\mathcal{H}}$, and visa-versa. Moreover, we can then write $|\bar{0}(\beta)\rangle$ in terms of these "real" and doubled space kets,

$$|\bar{0}(\beta)\rangle = Z^{-1/2}(\beta) \sum_n e^{-\beta H/2} |\psi_n\rangle \otimes |\tilde{\psi}_n\rangle. \quad (2.45)$$

There are two critical observations here:

1. The introduction of the fictitious tilde space has no effect on the trace since the tilde states live in a different Hilbert space than the actual system in which H and A are defined.
2. Perhaps uncomfortably, the doubled states identically mirror their "parent" real states. In other words, it is the case that $|\tilde{\psi}_n\rangle$ is a function of $|\psi_n\rangle$ such that their values are equal.

One can show by inspection that this definition does precisely reproduce the trace (we now write $|\psi_n\rangle = |n\rangle$ for brevity),

$$\begin{aligned}
\langle A \rangle_\beta &= \langle \bar{0}(\beta) | A | \bar{0}(\beta) \rangle \\
&= Z^{-1}(\beta) \sum_{nn'} \langle \tilde{n}' | \otimes \langle n' | e^{-\beta H/2} A e^{-\beta H/2} | n \rangle \otimes | \tilde{n} \rangle \\
&= Z^{-1}(\beta) \sum_{nn'} \underbrace{\langle \tilde{n}' | \tilde{n} \rangle}_{=\delta_{nn'}} \langle n' | e^{-\beta H/2} A e^{-\beta H/2} | n \rangle \\
&= Z^{-1}(\beta) \sum_n \langle n | e^{-\beta H/2} A e^{-\beta H/2} | n \rangle = \text{Tr}\{A \rho(\beta)\}.
\end{aligned} \tag{2.46}$$

2.D.2 The separable density matrix initial condition

The utility of the derived form of $|\bar{0}(\beta)\rangle$ is not yet clear, since the complexity of the problem is retained: each ket still requires a sum over all states. In order to show that $|\bar{0}(\beta)\rangle$ can be written as a unitary rotation as in Eq. (2.43), an approximation is required, which happens to be *exact* for the problems considered in this chapter, such as the Holstein, Peierls, or e.g. H+P models.

Consider the physical process captured by the zero-temperature Green's function [c.f. Eq. (2.2)]: a carrier is inserted at momentum k onto an otherwise empty lattice, propagated according to the full propagator, $G(\omega)$, and then annihilated at momentum k , with the expectation taken in the zero-carrier, zero-phonon vacuum. The requirement of an initially empty lattice is akin to the conduction band in an insulator, and requires a separable initial density matrix condition in the carrier and phonon manifolds. In other words, the initial state of the system is characterized by density matrix

$$\rho(\beta) = |0\rangle\langle 0| \otimes \frac{1}{Z_{\text{vib}}(\beta)} e^{-\beta H_{\text{vib}}}, \tag{2.47}$$

where $H_{\text{vib}} = \Omega \sum_i b_i^\dagger b_i$. This is *exact* for the model problems of interest (see e.g. relevant discussions in Ref. 5).

Immediately, this allows for simplification of the derived form of $|\bar{0}(\beta)\rangle$. Re-defining the sum

over all carrier-phonon states n as just a sum over phonon states, we may write

$$\langle A \rangle_\beta = Z_{\text{vib}}^{-1}(\beta) \langle 0 | \otimes \sum_n \langle n | e^{-\beta H_{\text{vib}}/2} A e^{-\beta H_{\text{vib}}/2} | n \rangle \otimes | 0 \rangle, \quad (2.48)$$

where $|0\rangle$ is the zero-carrier state, and $|n\rangle$ indexes some phonon configuration on the lattice. This then implies that

$$|\bar{0}(\beta)\rangle = Z_{\text{vib}}^{-1/2}(\beta) \sum_n e^{-\beta\Omega/2} |n\rangle \otimes |\tilde{n}\rangle. \quad (2.49)$$

Finally, we consider the single lattice site case for simplicity, collect the summation over n into a single exponential by the definition of the creation operator,

$$|n\rangle = \frac{1}{\sqrt{n!}} (b^\dagger)^n |0\rangle, \quad (2.50)$$

and utilize the definition of the vibrational partition function (for a single site),

$$Z_{\text{vib}}(\beta) = (1 - e^{-\beta\Omega})^{-1/2} \quad (2.51)$$

to write the thermal vacuum as

$$|\bar{0}(\beta)\rangle = (1 - e^{-\beta\Omega})^{1/2} \exp\left\{e^{-\beta\Omega/2} b^\dagger \tilde{b}^\dagger\right\} |0\rangle \otimes |\tilde{0}\rangle. \quad (2.52)$$

The meaning of the so-called doubled vacuum state is now apparent: it is the real and doubled (often also referred to as "fictitious") vacuum, $|\bar{0}\rangle \equiv |0\rangle \otimes |\tilde{0}\rangle$. The single-site to infinite-site generalization is trivial, but the next step in showing that Eq. (2.43) is possible is not, and is now discussed.

2.D.3 Showing a unitary form is possible

We ask show that the doubled vacuum state can be written as a unitary transformation, i.e.,

$$e^{-\frac{1}{2}G} = e^P = (1 - e^{-\beta\Omega})^{1/2} \exp\left\{e^{-\beta\Omega/2} b^\dagger \tilde{b}^\dagger\right\}. \quad (2.53)$$

To begin to address this claim, we rewrite the unitary transform as

$$U \equiv e^P = e^{\alpha k_+ - \alpha^* k_-}, \quad (2.54)$$

where the condition for this to be of unitary form is that $k_+ = k_-^\dagger$. The proof of this is simple: one asks the question, when is $P^\dagger = -P$? This is easily solved: $\alpha^* k_+^\dagger - \alpha k_-^\dagger = \alpha^* k_- - \alpha k_+$, where the result follows by inspection. Taking $k_+ = b^\dagger \tilde{b}^\dagger$ directly constrains $k_- = \tilde{b} b$. This has the apparent form of $SU(1, 1)$, which requires two constraints,

$$[k_0, k_\pm] = \pm k_\pm, \quad [k_+, k_-] = -2k_0. \quad (2.55)$$

By the second constraint above,

$$k_0 = \frac{1}{2}(\tilde{b}^\dagger \tilde{b} + b^\dagger b + 1). \quad (2.56)$$

It is easily checked that the first constraint is also satisfied by using the fundamental commutation relations $[b, b^\dagger] = [\tilde{b}, \tilde{b}^\dagger] = 1$ and $[b, \tilde{b}] = [b^\dagger, \tilde{b}^\dagger] = 0$ (and noting that real operators commute with fictitious ones).

We may rewrite our unitary transformation as

$$e^{\alpha k_+ - \alpha^* k_-} = e^{\alpha k_+} e^{b k_0} e^{c k_-}, \quad (2.57)$$

which is a result of the $SU(1, 1)$ Lie algebra [83, 84]. The disentanglement of exponentials, i.e., representing $e^{A+B} = e^A e^B \prod_i e^{C_i(A,B)}$ is generally a highly non-trivial task. However, if A and B

are part of a Lie algebra, the infinite product sum is finite [83].

Consider the example in question re-written slightly differently,

$$\exp\{\theta(\alpha_+k_+ + \alpha_0k_0 + \alpha_-k_-)\} = \exp\{f_+(\theta)k_+\}\exp\{f_0(\theta)k_0\}\exp\{f_-(\theta)k_-\}, \quad (2.58)$$

where later on we will set $\alpha_0 = 0$, $\alpha_- = \alpha_+^*$ and $\theta = 1$. To solve this, we first differentiate with respect to θ ,

$$\begin{aligned} Ke^{\theta K} &= f'_+(\theta)k_+\exp\{f_+(\theta)k_+\}\exp\{f_0(\theta)k_0\}\exp\{f_-(\theta)k_-\} \\ &\quad + \exp\{f_+(\theta)k_+\}f'_0(\theta)k_0\exp\{f_0(\theta)k_0\}\exp\{f_-(\theta)k_-\} \\ &\quad + \exp\{f_+(\theta)k_+\}\exp\{f_0(\theta)k_0\}f'_-(\theta)k_-\exp\{f_-(\theta)k_-\}, \end{aligned} \quad (2.59)$$

where we define $K \equiv \alpha_+k_+ + \alpha_0k_0 + \alpha_-k_-$ for brevity. Leveraging the identity

$$[e^{A_1} \cdots e^{A_m}]^{-1} = e^{-A_m} \cdots e^{-A_1}, \quad (2.60)$$

which holds if A_i are non-singular (which we assume), we multiply from the right with $e^{-\theta K}$,

$$\begin{aligned} K &= f'_+(\theta)k_+ + f'_0(\theta)\exp\{f_+(\theta)k_+\}k_0\exp\{-f_+(\theta)k_+\} \\ &\quad + f'_-(\theta)\exp\{f_+(\theta)k_+\}\exp\{f_0(\theta)k_0\}k_-\exp\{-f_0(\theta)k_0\}\exp\{-f_+(\theta)k_+\}, \end{aligned} \quad (2.61)$$

where $f'(\theta) \equiv df(\theta)/d\theta$. From here, we require a few similarity transforms,

$$e^{-i\phi k_0}k_{\pm}e^{i\phi k_0} = e^{\mp i\phi}k_{\pm}, \quad (2.62a)$$

$$e^{-i\phi k_+}k_0e^{i\phi k_+} = k_0 + i\phi k_+, \quad (2.62b)$$

$$e^{-i\phi k_+}k_-e^{i\phi k_+} = k_- + 2i\phi k_0 - \phi^2 k_+. \quad (2.62c)$$

Using these relations, we make the substitution $f_m(\theta) = -\dot{\phi} \phi_m(\theta)$, for $m \in \{+, -, 0\}$, and find

$$e^{f_+(\theta)k_+} k_0 e^{-f_+(\theta)k_+} = k_0 - f_+(\theta)k_+, \quad (2.63a)$$

$$e^{-f_0(\theta)} e^{-\dot{\phi}k_+} e^{f_0(\theta)k_0} k_- e^{-f_0(\theta)k_0} e^{\dot{\phi}k_+} = e^{-f_0(\theta)} [k_- - 2f_+(\theta)k_0 + f_+^2(\theta)k_+]. \quad (2.63b)$$

This leads to,

$$\alpha_+ k_+ + \alpha_0 k_0 + \alpha_- k_- = f'_+(\theta)k_+ + f'_0(\theta)[k_0 - f_+(\theta)k_+] + e^{-f_0(\theta)} f'_-(\theta)[k_- - 2f_+(\theta)k_0 + f_+^2(\theta)k_+], \quad (2.64)$$

which is a nonlinear system of equations after matching the coefficients of the operators (we drop θ -dependence for brevity),

$$\alpha_+ = f'_+ - f'_0 f_+ + e^{-f_0} f_+^2 f'_-, \quad (2.65a)$$

$$\alpha_0 = f'_0 - 2e^{-f_0} f'_- f_+, \quad (2.65b)$$

$$\alpha_- = e^{-f_0} f'_-. \quad (2.65c)$$

It follows that $f_- = \alpha_- e^{f_0}$, which we substitute into the other two equations,

$$\alpha_+ = f'_+ - f'_0 f_+ + \alpha_- f_+^2, \quad (2.66a)$$

$$\alpha_0 = f'_0 - 2\alpha_- f_+. \quad (2.66b)$$

The solution to this system is not straightforward to work out but is given by a form related to the Riccati equations. We do not present that here (see pg. 52, Ref. 83). Instead, we present the core result, and note that when $\theta = 1$ and $\alpha_0 = 0$, we recover $f_+ = a$ and $f_- = c$. In this situation, we have,

$$f_+ = a = -e^{\dot{\delta}} \tanh(-|\alpha|), \quad (2.67a)$$

$$f_- = c = e^{-\dot{\delta}} \tanh(-|\alpha|), \quad (2.67b)$$

$$f_0 = b = -2 \ln \cosh(-\alpha), \quad (2.67c)$$

where δ is a real phase such that $\alpha = |\alpha| e^{\pm i\delta}$. It will end up being that α is real, so δ will ultimately not matter.

Returning to the beginning of the calculation, let's simplify the form of $e^{ak_+} e^{bk_0} e^{ck_-} |\bar{0}\rangle$. Noting that

$$e^{ck_-} |\bar{0}\rangle = [1 + O(\tilde{a}, a)] |\bar{0}\rangle = |\bar{0}\rangle \quad (2.68)$$

and that $|\bar{0}\rangle$ is actually an eigenvector of k_0 such that

$$e^{bk_0} |\bar{0}\rangle = e^{b/2} |\bar{0}\rangle \quad (2.69)$$

(that is easily shown via direct inspection), we have

$$e^{ak_+} e^{bk_0} e^{ck_-} |\bar{0}\rangle = e^{b/2} e^{ak_+} |\bar{0}\rangle. \quad (2.70)$$

Now it's just term matching. First, matching to that of rightmost part of Eq. (2.53), we have

$$e^{b/2} = (1 - e^{-\beta\Omega})^{1/2}, \quad a = e^{-\beta\Omega/2}. \quad (2.71)$$

Second, we can actually solve for G via the form $e^{-iG} = e^{\alpha k_+ - \alpha^* k_-}$, which of course requires solving for α . Combining previous results, we have the relation

$$\cosh^2(-|\alpha|) = (1 - e^{-\beta\Omega})^{-1}, \quad (2.72)$$

the solution of which is

$$|\alpha| = \operatorname{arctanh} \left(e^{-\beta\Omega/2} \right). \quad (2.73)$$

At this point, we simply need to find the phase of α , which is easily done by using the relation for

a ,

$$a = e^{-\beta\Omega/2} = -e^{i\delta} \tanh\left(-\operatorname{arctanh}\left(e^{-\beta\Omega/2}\right)\right), \quad (2.74)$$

which can be solved to immediately constrain $\delta = 0$ which leads to $\alpha \in \mathbb{R}$. This finally constrains the final form of α to be

$$\alpha = \alpha^* = \operatorname{arctanh}\left(e^{-\beta\Omega/2}\right). \quad (2.75)$$

With this result, we substitute back into the unitary form to find G ,

$$-\mathbb{i}G = \operatorname{arctanh}\left(e^{-\beta\Omega/2}\right) [b^\dagger \tilde{b}^\dagger - b \tilde{b}], \quad (2.76)$$

which after some rearrangement is precisely the form found in the literature.

In summary, the core result of this Appendix is to show that

$$\operatorname{Tr}\{A\rho(\beta)\} = \langle 0|_{\text{elec}} \otimes \langle \bar{0}| e^{\mathbb{i}G(\beta)} A e^{-\mathbb{i}G(\beta)} |\bar{0}\rangle \otimes |0\rangle_{\text{elec}} \quad (2.77)$$

In Appendix 2.E, we will show rigorously that when A is the exact time-ordered Green's function, that the core result of TFD will be seamlessly compatible with the apparatus of GGCE, straightforwardly allowing finite-temperature spectra calculations at the cost of a doubled Hilbert space size.

2.E The thermal Green's function

The next step is to connect the previous derivations to the exact thermal Green's function.⁸ The true thermal Green's function, \mathcal{G} , is a thermal trace over all states in the system,⁹

$$\mathbb{i}\mathcal{G}(k, t; \beta) = \left\langle T c_k(t) c_k^\dagger \right\rangle_\beta. \quad (2.78)$$

⁸We remind the reader that this derivation is only exact for systems in which the initial density matrix is separable [c.f. Eq. (2.47)].

⁹Note that in this section, $G(\beta)$ is part of the thermal transformation operator $e^{\mathbb{i}G(\beta)}$ and $\mathcal{G}(k, \cdot; \beta)$ is the thermal Green's function.

Using the separable density matrix initial condition, we may write the thermal Green's function as

$$i\mathcal{G}(k, t; \beta) = \langle \bar{0} | e^{iG(\beta)} T c_k(t) c_k^\dagger e^{-iG(\beta)} | \bar{0} \rangle, \quad (2.79)$$

where we have defined a composite doubled vacuum as

$$|\bar{0}\rangle \equiv |\bar{0}\rangle \otimes |0\rangle_{\text{elec}}. \quad (2.80)$$

So far, we have only used the definition of the thermal trace in the TFD formalism. The time ordering operator commutes with any time-independent operators, including $G(\beta)$, which combined with expanding $c_k(t)$ into its Schrödinger picture representation, we have

$$i\mathcal{G}(k, t; \beta) = \langle \bar{0} | T e^{iG(\beta)} e^{iHt} c_k e^{-iHt} c_k^\dagger e^{-iG(\beta)} | \bar{0} \rangle. \quad (2.81)$$

Next, we require a slight modification to H . In order to ensure that the final Hamiltonian does not contain terms which directly couple the real and fictitious phonon spaces, we insert the identity operator, $e^{-i\tilde{H}_{\text{vib}}t} e^{i\tilde{H}_{\text{vib}}t}$, where

$$\tilde{H}_{\text{vib}} = \Omega \sum_i \tilde{b}_i^\dagger \tilde{b}_i. \quad (2.82)$$

This leads to

$$i\mathcal{G}(k, t; \beta) = \langle \bar{0} | T e^{iG(\beta)} e^{i\tilde{H}t} c_k e^{-i\tilde{H}t} c_k^\dagger e^{-iG(\beta)} | \bar{0} \rangle, \quad (2.83)$$

where

$$\tilde{H} \equiv H - \tilde{H}_{\text{vib}}. \quad (2.84)$$

This choice for \tilde{H}_{vib} is *not* unique,¹⁰ it could have been any fictitious space operator. Since this is an insertion of identity, the expectation value of the thermal Green's function does not change.

¹⁰It is an instructive exercise to carry out the derivations in this section without making the modification to follow. The reader will then see how the final vibrational Hamiltonian *cannot* be written with the real and fictitious phonon operators decoupled. In other words, there will be terms like $\tilde{b}_i^\dagger b_i$, which render the vibrational Hamiltonian much more difficult to resolve.

Insertion of a second identity $e^{-iG(\beta)}e^{iG(\beta)}$ allows us to straightforwardly transform the Hamiltonians in \mathcal{G} into thermal Hamiltonians,¹¹

$$i\mathcal{G}(k, t; \beta) = \langle \bar{0} | T e^{i\bar{H}(\beta)t} c_k e^{-i\bar{H}(\beta)t} c_k^\dagger | \bar{0} \rangle, \quad (2.85)$$

where

$$\bar{H}(\beta) \equiv e^{iG(\beta)} \bar{H} e^{-iG(\beta)}. \quad (2.86)$$

We note the similarity to the time-ordered zero-temperature Green's function, which is an expectation value over the vacuum,

$$i\mathcal{G}(k, t) = \langle 0 |_{\text{elec}} \otimes \langle 0 | T e^{iHt} c_k e^{-iHt} c_k^\dagger | 0 \rangle \otimes | 0 \rangle_{\text{elec}}, \quad (2.87)$$

where $|0\rangle \otimes |0\rangle_{\text{elec}}$ contains no phonons or carriers.

2.E.1 Managing the time-ordering operator

Ultimately, we search for a form of $\mathcal{G}(k, t; \beta)$ without the time-ordering operator present. To do this, we expand the definition of the time ordering operator,

$$i\mathcal{G}(k, t; \beta) = \langle \bar{0} | [\Theta(t) c_k(t; \beta) c_k^\dagger - \Theta(-t) c_k^\dagger c_k(t; \beta)] | \bar{0} \rangle, \quad (2.88)$$

where $c_k(t; \beta) \equiv e^{i\bar{H}(\beta)t} c_k e^{-i\bar{H}(\beta)t}$ and $\Theta(t)$ is the Heaviside step function. The advanced contribution is 0, since $\langle \bar{0} | c_k^\dagger$ involves only the electronic, time and temperature-independent components, and directly annihilates the vacuum. This means $\mathcal{G}(k, t; \beta)$ takes the form

$$i\mathcal{G}(k, t; \beta) = \Theta(t) \langle \bar{0} | e^{i\bar{H}(\beta)t} c_k e^{-i\bar{H}(\beta)t} c_k^\dagger | \bar{0} \rangle. \quad (2.89)$$

¹¹We utilize the fact that the carrier operators commute with $G(\beta)$.

By the invariance property [81],

$$e^{\pm G(\beta)}(H_{\text{vib}} - \tilde{H}_{\text{vib}})e^{-\pm G(\beta)} = H_{\text{vib}} - \tilde{H}_{\text{vib}} \quad (2.90)$$

and due to the electronic operators in the Hamiltonians of the form we're considering (e.g. Holstein), the first exponentiated Hamiltonian is given by $\langle \bar{0} | e^{\pm \tilde{H}(\beta)t} = \langle \bar{0} |$, which leaves us at a familiar form [12]. We can now write the TFD Green's function in frequency space:

$$\mathcal{G}(k, \omega; \beta) = \langle \bar{0} | c_k \frac{1}{\omega - \tilde{H}(\beta) + i\eta} c_k^\dagger | \bar{0} \rangle. \quad (2.91)$$

At this point, the only remaining task is to evaluate the precise form of $\tilde{H}(\beta)$.

2.E.2 The transformed Hamiltonian

To make progress we now need to know the form of $\tilde{H}(\beta)$. Noting that carrier operators do not transform,

$$e^{\pm G(\beta)} c_k e^{-\pm G(\beta)} = c_k, \quad (2.92)$$

and that phonon operators transform as,

$$e^{\pm G(\beta)} b_i e^{-\pm G(\beta)} = \cosh(\theta[\beta]) b_i + \sinh(\theta[\beta]) \tilde{b}_i^\dagger, \quad (2.93)$$

where

$$\theta[\beta] = \operatorname{arctanh}(e^{-\beta\Omega/2}), \quad (2.94)$$

and

$$G(\beta) = -i\theta[\beta] \sum_i (b_i \tilde{b}_i - b_i^\dagger \tilde{b}_i^\dagger), \quad (2.95)$$

we can construct $\tilde{H}(\beta)$ explicitly for, e.g., the Holstein model. Note that at $\beta \rightarrow \infty \iff T \rightarrow 0$, we have $\theta[\beta] \rightarrow \operatorname{arctanh}(0) = 0$, which causes all fictitious operators to have coefficients of zero,

recovering the zero-temperature solution.

The Holstein Hamiltonian¹² is

$$H = K + \Omega \sum_i b_i^\dagger b_i + \alpha \sum_i c_i^\dagger c_i (b_i^\dagger + b_i), \quad (2.96)$$

where K is the nearest-neighbor hopping term

$$K \equiv -t \sum_{\langle ij \rangle} c_i^\dagger c_j \quad (2.97)$$

which does not transform under the thermal transformation operator. The thermal Holstein model under the \tilde{H}_{vib} shift is

$$\bar{H}(\beta) = K + \Omega \sum_i (b_i^\dagger b_i - \tilde{b}_i^\dagger \tilde{b}_i) + \alpha \sum_i c_i^\dagger c_i (e^{\imath G} b_i^\dagger e^{-\imath G} + e^{\imath G} b_i e^{-\imath G}) \quad (2.98)$$

(again by recalling that $\bar{H}_{\text{vib}} \equiv H_{\text{vib}} - \tilde{H}_{\text{vib}}$ does not transform). After executing all of the unitary transforms for the interaction term V , we have,

$$V(\beta) \equiv e^{\imath G} V e^{-\imath G} = \alpha \sum_i c_i^\dagger c_i [\cosh(\theta[\beta])(b_i^\dagger + b_i) + \sinh(\theta[\beta])(\tilde{b}_i^\dagger + \tilde{b}_i)]. \quad (2.99)$$

This representation has incredible utility, because it is *precisely* in the correct form to represent it as, essentially, two *separate* Holstein couplings,

$$V(\beta) = \cosh(\theta[\beta])V + \sinh(\theta[\beta])\tilde{V}, \quad (2.100)$$

with

$$\tilde{V} \equiv \alpha \sum_i c_i^\dagger c_i (\tilde{b}_i^\dagger + \tilde{b}_i), \quad (2.101)$$

which is rigorously no different than the current GGCE implementation for different boson modes.

¹²See e.g. Ref. 85 for similar derivations on the molecular Holstein Hamiltonian using TFD.

The only other difference is in the vibrational energy in the free particle propagator, which we will now address.

2.E.3 The free particle propagator

The exactly solvable parts of the model are given by K and \bar{H}_{vib} . We already know that the free particle Green's function is

$$G_0(k, \omega) = \frac{1}{\omega - \varepsilon_k + i\eta}, \quad (2.102)$$

where $\varepsilon_k = -2t \cos ka$. For some state $|k, n, \tilde{n}\rangle$ (we now write the composite state as a single ket for clarity, with the ordering being the carrier momentum index, followed by the number of phonon excitations in the real, then fictitious space), the Green's function is given by

$$G_0(\omega) |k, n, \tilde{n}\rangle = G_0 [k, \omega - \Omega(n - \tilde{n})] |k, n, \tilde{n}\rangle. \quad (2.103)$$

Note that the real and fictitious spaces "compete" in terms of their energies. For equal number of real and fictitious excitations, there is no shift in the free particle energy due to the phonons.

Physically, creating fictitious phonons now *stabilizes* the energy of the system. The process of creating a fictitious phonon can be thought of as the system ejecting a phonon quanta into the bath. The possibility for stabilizing the system through creating fictitious phonons likely renders finite-temperature GGCE non-variational, but in principle, this approach should still be much more efficient than computing the full thermal trace.

2.F Outlook and future plans

In the chapter so far, we have outlined a procedure for acquiring numerically exact Green's functions at zero and finite temperature. Our method allows the user to tune the "level of theory" from approximate (e.g. MA, which is a subset of the methods available within GGCE) to numerically exact benchmarks, where $M, N \rightarrow \infty$. In this Appendix, we briefly outline planned future work,

and highlight some strengths and limitations of the method.

First, we describe the planned implementations which we believe will be straightforward, either due to simple inspection or because literature precedent exists.

1. *Two and three dimensions.* The general formalism of the GGCE allows for trivial extensions to greater than one dimension. Rigorously, there are only two aspects of the implementation that will change. First, the variable M now indexes the length of a d -dimensional cube, such that for instance, $M = 3$ now has $3^3 = 27$ sites in 3 dimensions. Second, the form of the free-particle Green's function changes from the analytic expression in Eq. (2.21) to one which must be evaluated numerically [65].
2. *Multi-carrier models.* Implementing multiple carriers, as in the case of bipolarons has already been accomplished within the MA [59]. Thus we know that while this complicates the analytics of the method, it is possible to solve electron-phonon problems in the bipolaron limit, and thus it should not be an issue to generalize the implementation to that which is compatible with the GGCE.
3. *Multi carrier-band models.* Treating a single carrier which has access to multiple bands through some hopping mechanism should be possible, and allow our methods to be more useful to a broader group of people within the condensed matter and computational chemistry communities.

Next, we summarize the challenges and limitations of the GGCE method. Most prominently, the GGCE is currently a single-particle response function method. The apparatus of GGCE is not setup to handle e.g. two-particle correlation functions, the likes of which will be necessary to compute mobilities or (neural) absorption spectra. This challenge has to some degree been addressed within the MA [79], but it is unclear how seamlessly the method will generalize to $M > 1$. Moreover, finite-temperature two-particle quantities are likely out of reach using GGCE+TFD, since the squeezing transformation only produces tractable results in the insulating limit, a condition not satisfied in general for two-particle response functions.

Chapter 3

Microscopic model of the doping dependence of line widths in monolayer transition metal dichalcogenides

The content presented in this chapter is based on the following published work.

- Microscopic model of the doping dependence of line widths in monolayer transition metal dichalcogenides [arXiv]

M. R. Carbone,* M. Z. Mayers & D. R. Reichman, J. Chem. Phys. **152**, 194705 (2020).

Part of a special issue on 2D materials

Copyright 2020, American Institute of Physics

* Indicates corresponding authors.

3.1 Introduction

Monolayer transition metal dichalcogenides (TMDCs) are quasi-two-dimensional (2D) materials known to exhibit extraordinary physical phenomena [86, 87]. These materials may be viewed as semiconducting analogs of graphene [88–90] and present with non-trivial optical, electronic, and, under some circumstances, topological and superconducting properties [91, 92]. Due to their unique characteristics, monolayer TMDCs have been proposed for myriad practical applications [93] such as opto-electronics [94–97], field-effect transistors [98] and digital logic gates [99, 100]. Of particular fundamental interest is the nature of electron-hole complexes such as excitons [101, 102] and trions [103, 104] in TMDCs. Due to the reduced screening in 2D systems, such stable carrier complexes may have anomalously large binding energies, with that of the exciton reaching ~ 0.5 eV [101, 105–107], and that of the trion reported to be in the range of 20–35 meV [103, 106, 108–110], implying that trions are bound even at room temperature. These observations indicate that monolayer TMDCs are unique systems for investigating the properties of strongly interacting quasiparticles. In addition, they may provide unprecedented experimental clarity concerning the nature of interactions between these electron-hole complexes and phonons [111, 112], as well as with charge carriers and other quasiparticles.

A standard means of probing the nature of the interactions of excitons and trions with other excitations is via the broadening of line widths in clean samples with respect to control parameters such as the temperature or carrier density. Intrinsic homogeneous quasiparticle (QP) line widths [111] are generally obfuscated by inhomogeneous broadening due to the high level of static defects in processing. However, recent work has led to the observation of very narrow QP line widths via the preparation of ultra-clean samples by both dry transfer methods and chemical vapor deposition [113–115], and by the usage of non-linear spectroscopy to extract the homogeneous line width from inhomogeneously-broadened spectra [111]. The optical interrogation of the exciton and trion line widths in these less defective samples offers a unique opportunity to understand the mechanisms of the 2D exciton and trion scattering processes in quasi-2D systems.

There are many factors that affect line broadening in monolayer TMDCs, most notably interactions with phonons (as controlled by temperature) and interactions with other charge carriers (as controlled by doping). At very low temperatures and near the charge-neutrality point [112], it is expected that the intrinsic homogeneous line width due to lifetime broadening may be observed if the sample is clean enough. As temperature increases, phonons begin to play a significant role and will eventually dominate the line broadening process. The interaction of excitons with phonons has been studied in some detail in TMDCs [112, 116], and a variety of coupling motifs have been elucidated experimentally and theoretically.

Additionally, the concentration of electrons as controlled by gating can alter lines widths and line shapes in a non-trivial fashion [101, 117–119]. Studies which have investigated the electron density dependence of optical line shapes in monolayer TMDCs from the standpoint of the Fermi-polaron picture provide a means of describing optical line broadening as a function of doping [120–126]. Such many-body multiple scattering theories are essential for properly describing the full range of doping-dependent behavior, as the Fermi Golden Rule breaks down at sizable doping levels. However, the use of graphene gating and clean samples renders the investigation of the doping regime close to the charge-neutrality point possible [127]. Here, detailed microscopic Golden Rule-based calculations may be performed which can provide new insights into the line broadening mechanisms. Motivated by the aforementioned recent experimental works, we follow this latter path to assess how the elastic scattering of excitons and trions with free charge carriers may alter line widths of both ground and excited state excitonic complexes in the low doping regime. In particular, we investigate the circumstances for which doping-related broadening may compete with phonon-induced broadening, and we discuss the breakdown of the perturbative approach as a function of temperature and carrier density. The importance of our work extends beyond the description of linewidths and is of relevance for describing scattering processes such as Auger recombination and impact ionization in TMDCs.

Our paper is organized as follows: We first present an outline of the microscopic theory in Section 3.2, focusing on the electron-exciton scattering calculation, which is discussed in Sub-

section 3.2.1. Calculations for the electron-trion scattering are similar to that of the exciton and discussed (briefly) in Subsection 3.2.2. The low-temperature results for the trion and exciton line widths, in addition to the details of the model and limitations of the Golden Rule approach, are presented and discussed in Section 3.3. Finally, in Section 3.4, we summarize our conclusions and discuss outlook and potential future work. Details not contained in the main text are located in several appendices.

3.2 Methodology

In this section, the elastic (energy-conserving) scattering of electrons from both excitons and trions described within the Fermi Golden Rule approximation. Additionally, because we work at the Golden Rule level of theory, bound states in scattering are not considered. While such a treatment can only be valid at extremely low doping densities, recent synthetic work using encapsulated samples points to a route to experimentally controlled access to this regime. Furthermore, the use of the Golden Rule allows for a very detailed microscopic description [128, 129], the limitations which will be discussed in the following sections.

3.2.1 Electron-exciton elastic scattering

In order to facilitate the computation, we use a simple variational guess for the exciton wave function

$$\phi(r) = \sqrt{\frac{\pi}{2\lambda^2}} e^{-r/\lambda}, \quad (3.1)$$

where r is the relative coordinate of the two-body system. The optimal effective Bohr radius λ is chosen to best match the functional form of Eq. (3.1) to the ground state of a Wannier exciton in a Rytova-Keldysh potential [130, 131] found using exact diagonalization.

The second-quantized form of the exciton-free electron scattering state is

$$|\mathbf{k}_x, \mathbf{k}_e\rangle = \sum_{\mathbf{k}'} \phi_{\alpha_x \mathbf{k}_x + \mathbf{k}'}^* \psi_{\mathbf{k}_e}^* c_{-\mathbf{k}'}^\dagger d_{\mathbf{k}_x + \mathbf{k}'}^\dagger c_{\mathbf{k}_e}^\dagger |0\rangle, \quad (3.2)$$

which is a direct product state of the free exciton and electron states, $|\mathbf{k}_x\rangle \otimes |\mathbf{k}_e\rangle$. The wave function

$$\phi_{\mathbf{k}} = \sqrt{\frac{8\pi\lambda^2}{A}} g(\lambda k) \quad (3.3)$$

satisfies normalization $\sum_{\mathbf{k}} \phi_{\mathbf{k}}^2 \rightarrow \frac{A}{(2\pi)^2} \int d^2k \phi_{\mathbf{k}}^2 = 1$ and is derived by performing an in-plane Fourier transform of Eq. (3.1), where $g(x) = [1 + x^2]^{-3/2}$, $c_{\mathbf{k}}$ ($d_{\mathbf{k}}$) are electron (hole) annihilation operators for momentum index \mathbf{k} , A is the in-plane area of the 2D material, and $\alpha_x = m_e/M_x$ is the ratio of the electron and exciton effective masses (which manifests during the coordinate transform to relative/center of mass coordinates). The free-electron wave function $\psi_{\mathbf{k}} \propto e^{-i\mathbf{k}\cdot\mathbf{R}}$ characterizes an electron which may exhibit free in-plane motion, and together with the center of mass coordinate of the exciton, contributes only a global phase factor which may be ignored in subsequent calculations, as it does not contribute to the determination of the scattering rate.

Scattering matrix elements are computed by evaluating the coupling between an initial QP-free electron state, $|\mathbf{k}_x, \mathbf{k}_e\rangle$, and a final QP-free electron state in which momentum \mathbf{q} is transferred, $\langle \mathbf{k}_x + \mathbf{q}, \mathbf{k}_e - \mathbf{q} |$. The second-quantized, momentum-conserving potential energy operator $V = V_{eh} + V_{ee}$ that mediates this coupling may be split into electron-hole and electron-electron components,

$$V_{eh} = - \sum_{\substack{\mathbf{k}_1, \mathbf{k}_2, \mathbf{q} \\ s=\uparrow, \downarrow}} v_{\mathbf{q}} c_{\mathbf{k}_1+\mathbf{q}}^{s\dagger} d_{\mathbf{k}_2-\mathbf{q}}^\dagger d_{\mathbf{k}_2} c_{\mathbf{k}_1}^s, \quad (3.4a)$$

and

$$V_{ee} = \frac{1}{2} \sum_{\substack{\mathbf{k}_1, \mathbf{k}_2, \mathbf{q} \\ s_1, s_2 = \uparrow, \downarrow}} v_{\mathbf{q}} c_{\mathbf{k}_1+\mathbf{q}}^{s_1\dagger} c_{\mathbf{k}_2-\mathbf{q}}^{s_2\dagger} c_{\mathbf{k}_2}^{s_2} c_{\mathbf{k}_1}^{s_1}, \quad (3.4b)$$

where $v_{\mathbf{q}} = \frac{2\pi e^2}{A q \epsilon(q)}$ is the magnitude of the two-body interactions and $\epsilon(q)$ is a static dielectric function discussed in Section 3.2.3. The exciton-free electron elastic scattering matrix elements are henceforth defined as

$$V(\mathbf{q}, \mathbf{k}_e, \mathbf{k}_x) = \langle \mathbf{k}_x + \mathbf{q}, \mathbf{k}_e - \mathbf{q} | V | \mathbf{k}_x, \mathbf{k}_e \rangle. \quad (3.5)$$

Once matrix elements have been computed, the line width $\Gamma(n; \mathbf{k}_x)$ is calculated by summing over all final exciton states,

$$\Gamma(n, \mathbf{k}_x) = \frac{\hbar A}{(2\pi)^2} \int d^2q w(\mathbf{q}; n, \mathbf{k}_x). \quad (3.6)$$

Here, $w(\mathbf{q}; n, \mathbf{k}_x)$ is a partial scattering rate computed for fixed momentum transfer using Fermi's Golden Rule,

$$w(\mathbf{q}; n, \mathbf{k}_x) = \frac{2\pi}{\hbar} \sum_{\mathbf{k}_e} |V(\mathbf{q}, \mathbf{k}_e, \mathbf{k}_x)|^2 f(k_e) [1 - f(|\mathbf{k}_e - \mathbf{q}|)] \\ \times \delta \left(\frac{\hbar^2 k_x^2}{2M_x} + \frac{\hbar^2 k_e^2}{2m_e} - \frac{\hbar^2 |\mathbf{k}_x + \mathbf{q}|^2}{2M_x} - \frac{\hbar^2 |\mathbf{k}_e - \mathbf{q}|^2}{2m_e} \right), \quad (3.7)$$

where the Fermi-Dirac distribution

$$f(k) = \left[e^{(\hbar^2 k^2 / 2m_e - \mu) / k_B T} + 1 \right]^{-1} \quad (3.8)$$

contains doping-density (n) dependence through the chemical potential

$$\mu = k_B T \ln [\exp\{\varepsilon_F / k_B T\} - 1], \quad (3.9)$$

and the Fermi energy of a 2D electron gas, $\varepsilon_F = \pi \hbar^2 n / m_e$. In order to simplify the calculations, the parameter $\mathbf{k}_x = 0$ is taken in all computations, effectively choosing a reference frame in which the exciton is at rest. For further details, we refer the reader to Ref. 128, where similar calculations are performed for anisotropic 3D systems.

3.2.2 Electron-trion scattering

Computation of the trion-free electron elastic scattering line width contribution is similar to that of the excitonic case in all ways except for the determination of the scattering matrix elements.

The trion-free electron scattering state is constructed similarly to that of Eq. (3.2), with a few key distinctions to be noted below. Explicitly, we write this scattering state as,

$$|\mathbf{k}_t, \mathbf{k}_e\rangle = \sum_{\substack{\mathbf{k}_1, \mathbf{k}_2 \\ s_1, s_2, s_e}} \xi_S^*(s_1, s_2) \Phi_{\alpha_t \mathbf{k}_t + \mathbf{k}_1, \alpha_t \mathbf{k}_t + \mathbf{k}_2}^* \psi_{\mathbf{k}_e}^* c_{-\mathbf{k}_1}^{s_1^\dagger} c_{-\mathbf{k}_2}^{s_2^\dagger} d_{\mathbf{k}_t + \mathbf{k}_1 + \mathbf{k}_2}^\dagger c_{\mathbf{k}_e}^{s_e^\dagger} |0\rangle. \quad (3.10)$$

Note the introduction of a spin wave function which constrains the trion to the singlet spin configuration, $\xi_S(s_1, s_2) = \langle s_1 s_2 | S \rangle$, via the projection of a two-fermion spin state $\langle s_1 s_2 |$ on the singlet state $|S\rangle$. The projection satisfies the properties, $\sum_{s_1, s_2} \xi_S^*(s_1, s_2) \xi_S(s_1, s_2) = \langle S | S \rangle = 1$, and $\xi_S(s_1, s_2) = -\xi_S(s_2, s_1)$ as per Fermionic anti-commutation rules. Given that the trion triplet state is, at most, weakly bound, we only consider only singlet to singlet scattering.

The trion wave function, Φ , is given by

$$\Phi_{\mathbf{k}_1, \mathbf{k}_2} = \mathcal{N} \frac{8\pi\lambda_1\lambda_2}{A} g(\lambda_1 k_1) g(\lambda_2 k_2). \quad (3.11)$$

Here, λ_1 and λ_2 are variational parameters associated with the Chandrasekhar-type wave function [103, 132], and the constant \mathcal{N} is a normalization factor

$$\mathcal{N} = \frac{1}{\sqrt{1 + \kappa^2}}, \quad \kappa = \frac{4\lambda_1\lambda_2}{(\lambda_1 + \lambda_2)^2} \quad (3.12)$$

which arises during the variational minimization of the trion binding energy [133].

Once the matrix elements

$$\mathcal{V}(\mathbf{q}, \mathbf{k}_e, \mathbf{k}_t) = \langle \mathbf{k}_t + \mathbf{q}, \mathbf{k}_e - \mathbf{q} | V | \mathbf{k}_t, \mathbf{k}_e \rangle \quad (3.13)$$

are computed, the trion line width may be determined using Eqs. (3.6) and (3.7) in the same way as for the exciton case (with the appropriate substitutions, e.g. the initial QP momentum $\mathbf{k}_x \rightarrow \mathbf{k}_t$, the mass ratio $\alpha_x \rightarrow \alpha_t = m_e/M_t$, etc.). Line widths for low doping densities are reported in Section 3.3, computational details of this calculation are given in Appendix 3.D and the physical

parameters used may be found in the caption of Fig 3.1.

3.2.3 Dielectric Function

The dielectric function $\varepsilon(q, \omega)$ takes into account properties of the monolayer TMDC, the surrounding medium, and the excess electron gas [134], respectively, and may be broken down into distinct contributions as [135]

$$\varepsilon(q, \omega) = \varepsilon_I(q) + \varepsilon_{II}(q, \omega). \quad (3.14)$$

We follow previous work [136] and screen the direct and exchange interactions, in contrast to the usual Bethe-Salpeter treatment of bound state formation where the exchange interaction is unscreened [137–140]. The first term consists of a static contribution from the monolayer TMDC and surrounding layers in the absence of doping,

$$\varepsilon_I(q) = \varepsilon_0(1 + 2\pi\chi_{2D}q), \quad (3.15)$$

where $\varepsilon_0 = (\varepsilon_a + \varepsilon_b)/2$ is the dielectric constant of the surrounding medium [130, 131] (the average of the two encapsulating dielectrics) and χ_{2D} is the dielectric polarizability of the 2D material.

The second term is due to the presence of doping electrons and is generally frequency dependent. We follow Stern [134] and treat the excess electrons as a 2-dimensional homogeneous electron gas (HEG). In the static ($\omega = 0$) approximation, this yields

$$\varepsilon_{II}(q, 0) = \frac{2me^2}{\hbar^2 q} \begin{cases} 1 & \text{if } q \leq 2k_F \\ 1 - \sqrt{1 - (2k_F/q)^2} & \text{if } q > 2k_F \end{cases}. \quad (3.16)$$

Note that Eq. (3.16) implicitly carries a doping density (n) dependence through the Fermi momentum $p_F = \hbar k_F = \hbar \sqrt{2\pi n}$.

To motivate this choice, we observe that $\varepsilon(q, 0)$ captures the correct behavior in both the small-wavelength and low-doping limits. In the low-doping limit, the Stern-like term vanishes and the

dielectric function $\varepsilon(q, 0) \rightarrow \varepsilon_I(q)$, which is the dielectric function of the material and its surroundings. The low q -limit suppresses the term containing the polarizability and diverges like $1/q$, correctly screening the 2D Coulomb interaction at small q [141].

If doping levels are large enough, the static approximation presented above will fail [141, 142]. Although this signals one aspect of the high-doping density breakdown of the Golden Rule, one way to potentially extend its domain of validity is to utilize a frequency-dependent scattering matrix element as discussed in Ref. 136. This leads to a dielectric function derived from the 2D Lindhard function, $\varepsilon_{II}(q, E_{\text{eff}}/\hbar)$ [134, 136], evaluated at the effective energy

$$E_{\text{eff}} \equiv E(k_e) - E(|\mathbf{k}_e - \mathbf{q}|) = \frac{\hbar^2(2\mathbf{k}_e \cdot \mathbf{q} - q^2)}{2m_0}, \quad (3.17)$$

which is the energy difference between the initial and final states of the scattering electron. The details of $\varepsilon_{II}(q, \omega)$ are presented in Ref. 134 and in Appendix 3.A.

3.3 Results and Discussion

The Fermi Golden Rule is expected to be valid only in the ultra-low doping regime ($\varepsilon_F \ll \varepsilon_t \sim 10^{12} \text{ cm}^{-2}$), where ε_t denotes the trion binding energy in the limit of zero doping. As the doping level increases, many-body, multi-scattering effects become prominent [143], and a Fermi-polaron-like picture appears to be required [120, 122, 123]. Since the low doping regime is now potentially controllable and accessible in encapsulated samples with graphene gating layers, a Golden Rule approach is useful in enabling a fully microscopic treatment in this restricted regime.

Line widths versus doping level for both the exciton and trion lines are displayed for 5 and 25 K in Fig. 3.1. Results are presented for the experimentally-relevant case of a layer encapsulated by dielectric media with properties mimicking that of boron nitride. We also note that in a hypothetical suspended sample ($\varepsilon_0 = 1$), Γ is enhanced compared to results presented in Fig. 3.1 (e.g. roughly 5 meV at 10^{11} cm^{-2} , compared to only 1 meV in the encapsulated case) and is comparable, or even larger than that associated with phonon-induced broadening, since the scaling of

Γ with respect to the background dielectric function varies roughly as ε_0^{-2} . It should also be noted that in experiments the encapsulating layers are of finite thickness, and while this situation can be handled theoretically [144, 145], we do not do so here as it complicates the treatment of the dielectric function. We thus expect the true magnitude of line width values to be somewhat larger than the values presented in Fig. 3.1. Additionally, while we have also carried out an investigation of inelastic electron-capture scattering, we find that elastic scattering dominates the line widths in the regimes we consider. Thus, we only focus on the elastic scattering contributions.

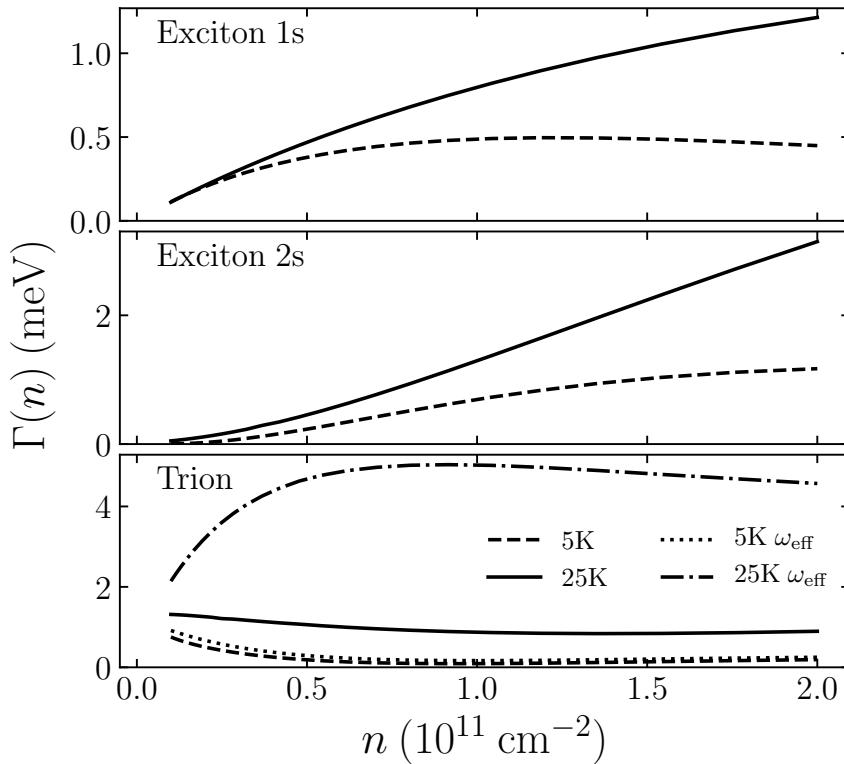


Figure 3.1: Line width broadening of monolayer MoSe_2 as a function of electron doping density for BN-encapsulated ($\varepsilon_0 = 4.5$) [146] monolayers. The following parameters were used: in the exciton calculation, the effective Bohr radii $\lambda_0 = 10.3$ and in the case of the trion, $\lambda_1 = \lambda_0$ and $\lambda_2 = 25.2 \text{ \AA}$ [103]. In the exciton 2s elastic scattering, $a = 7.79 \text{ \AA}$ and $b = 6.20 \text{ \AA}$ (see Appendix 3.B.4). The electron (hole) effective masses employed were 0.49 (0.61) (in units of m_0) [147], and the polarizability $\chi_{2D} = 8.23 \text{ \AA}$ [103]. In the case of exciton elastic scattering, the singlet and triplet contributions are identical as the exchange contribution to the potential dominates; trion triplet states are not considered. Additionally, screening using the effective frequency-dependent dielectric function (see Eq. 3.17) are presented for the trion, as the effective screening does not appear to affect the exciton line width.

We first discuss trion line broadening. For doping levels $n > 0.4 \times 10^{11} \text{ cm}^{-2}$, the trion line width in all cases is largely independent of doping density. The upturn seen in the static screening trion line width as doping density decreases is likely an artifact of behavior embedded in the function $\varepsilon_2(q)$. Indeed, a suppression of the q^{-3} behavior for large q of this function leads to an essentially flat trion line width as a function of doping level, similar to that seen in Fermi-polaron-like theories and in some experiments [148, 149]. It should be noted that in these approaches, however, the trion line broadening is controlled by a phenomenological input parameter. For example, the width of the doping-independent trion line in the work of Efimkin and MacDonald [123] is given by the parameter γ which is input phenomenologically by hand is not derived directly from the microscopic interactions in the system.

Here, our fully microscopic approach allows for the microscopic extraction of the magnitude and temperature dependence of the trion line width. While the static and effective frequency-dependent screening cases are largely in agreement at low T , the same cannot be said for results at 25K. Given the subtle changes in the scattering matrix elements except at small q , this difference likely arises from the larger accessible density of states available at higher densities away from $\omega = 0$ in the screening function.

We now turn to the broadening of the exciton line. Unlike the trion case, the exciton line width monotonically increases as a function of doping density at low values of n in the 25 K case. This is again in agreement with experimental expectations [150, 151].¹ as well as the behavior found in many-body approaches [120, 123, 143]. In particular, in these latter theoretical approaches, an approximately linear dependence of the line width on doping manifests over a much wider doping density range for the exciton line. The very same behavior arises from the Golden Rule at extremely low doping. The decrease of the slope of the line width as n increases, most clearly

¹Recent unpublished experimental results by K. Wagner, E. Wietek, and J. Zipfel in the group of Alexey Chernikov have given access to line widths as a function of doping in the low doping regime in clean, encapsulated WSe₂ at 5 K. While the data fills in the upper end of the doping regime we present in our Fig. 3.1, they are in reasonable agreement with our results, including monotonically increasing (with doping level n) exciton lines of width 5-10 meV, a somewhat larger growth for the 2s line as compared to 1s, and a nearly flat trion line that initially slightly decreases with increasing n . We thank Alexey Chernikov for sharing these results to us after the submission of this work. A comparison of the theoretically calculated vs. experimental line widths is presented in Appendix 3.E.

demonstrated in the near-plateau of the 5 K exciton line widths above $n = 0.8 \times 10^{11} \text{ cm}^{-2}$, is a signature of the breakdown of the Golden Rule. Specifically, due to the $\epsilon_F/k_B T$ term in the Fermi-Dirac distribution function, the crossover from the non-degenerate to the degenerate electron gas limit will induce a change in the doping dependence of the excitonic line width from a linear scaling $\Gamma \sim n$ at low doping to an eventual plateau $\sim k_B T$, and then an unphysical decline with increasing n . This same trend is reported in Ref. 128 for the quantum well case. We systematically examine this behavior in Fig. 3.2, which shows the doping and temperature dependence of this behavior. If one focuses on the more physically-described regime $n < 0.8 \times 10^{11} \text{ cm}^{-2}$, it is observed that, unlike in the trion case, doping-induced exciton line broadening is largely insensitive to temperature variations in the range $T = 5\text{-}25 \text{ K}$. Furthermore, given the fact that phonon-induced line broadening is suppressed at these temperatures, doping induced line broadening effects may be observable at $T = 5 \text{ K}$ in clean, encapsulated samples even for doping densities as low as $n \sim 2 \times 10^{11} \text{ cm}^{-2}$, especially with respect to the 2s line, where the line broadening effects appear to be slightly enhanced compared to the ground state.

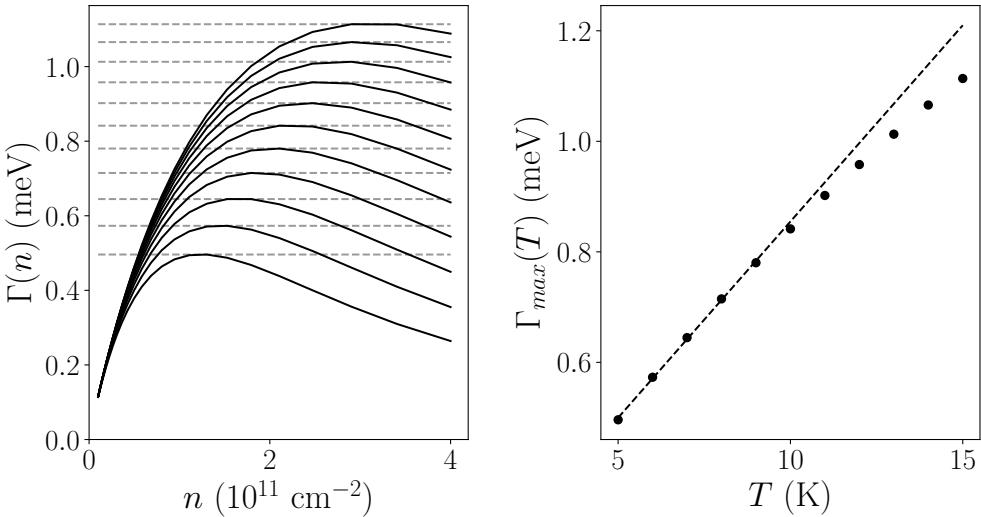


Figure 3.2: Doping dependence of the 1s exciton line width at temperatures $T = 5, 6, \dots, 15 \text{ K}$ (left). Parameters describe monolayer MoSe₂, as seen in Fig 3.1. Lower line widths correspond to lower temperatures. The horizontal dashed lines show the plateau location. The value of the maximum (plateau) is also plotted as a function of temperature, showing clear linear behavior at low T (right).

3.4 Conclusion

In this work we have employed perturbation theory to calculate the rates of electron-exciton and electron-trion scattering in monolayer TMDCs in the low doping density limit. Our approach is fully microscopic with respect to all input parameters and functions, including matrix elements and the dielectric screening model. On the other hand, it is expected that the Fermi Golden Rule should break down at low doping densities close to the degeneracy crossover of the electron gas in the monolayer, and some caution must be exercised with respect to the use the forms of the dielectric screening functions employed here [141]. Avoiding these approximations allows for the description of a much broader range of doping, but requires a full frequency-dependent many-body treatment [120, 122, 123].

Accepting the above limitations, the calculations presented here still allow for some important conclusions to be drawn. First, we find that with a reasonable treatment dielectric environment, exciton line widths arising from exciton-electron scattering on the order of 1 meV or higher are possible at low temperatures in the low doping regime accessible in encapsulated, graphene-gated samples. Thus, even mild doping may provide a line broadening mechanism that can compete with (but not necessarily exceed) lifetime and phonon-related broadening in this regime. As expected from previous many-body calculations in the very low doping regime, the growth of the excitonic line width is monotonic with increasing n , while the trion line width is largely insensitive to doping. However we find that the trion line width is sensitive to temperature variations even over the small range $T = 5\text{--}25$ K, a somewhat unexpected feature from the standpoint of many-body theories such as that of Ref. 123 where the trion line width is partly described by a phenomenological input parameter. Lastly, we find that excited state exction line broadening is somewhat larger and shows more sensitivity to increases in doping levels. Future work should be devoted to testing the veracity of these predictions and to understand how they merge with many-body approaches which have been applied to study the higher doping density regime [152].

In conclusion, we have provided a microscopic model for understanding how the scattering of

excitons and trions surrounded by an electron gas in monolayer TMDCs may induce line broadening in the very low doping density limit at low temperatures. A more detailed effort aimed at placing these contributions in the context of other mechanisms, such as exciton-phonon scattering, is worth of future study. In addition, the approach adopted here may be of use for the calculation of the rates of processes such as Auger recombination [153–155] in dimensionally-confined systems. These and related topics will be the subject of future investigations.

3.A RPA Polarizability

Following the definition in Stern [134], in this appendix we present the frequency-dependent 2D electron gas polarizability, χ , and its $\omega \rightarrow 0$ limit. The general form of χ is $\chi(q, \omega) = \chi_1(q, \omega) + i\chi_2(q, \omega)$, where

$$\chi_1(z, \tilde{u}) = \frac{e^2 m}{\hbar^2 q^2 \pi} \left\{ 1 - \frac{C_-(z, \tilde{u})}{2} \sqrt{(1 - \tilde{u})^2 - z^{-2}} - \frac{C_+(z, \tilde{u})}{2} \sqrt{(1 + \tilde{u})^2 - z^{-2}} \right\} \quad (3.18)$$

and

$$\chi_2(z, \tilde{u}) = \frac{e^2 m}{\hbar^2 q^2 \pi} \left\{ \frac{D_-(z, \tilde{u})}{2} \sqrt{z^{-2} - (1 - \tilde{u})^2} - \frac{D_+(z, \tilde{u})}{2} \sqrt{z^{-2} - (1 + \tilde{u})^2} \right\}, \quad (3.19)$$

where $z \equiv q/2k_F$ and $\tilde{u} \equiv 2\omega m/\hbar q^2$. Note that the quantities in the braces, $\{\cdot\}$, are dimensionless. The functions C and D are defined as follows,

$$C_{\pm}(z, \tilde{u}) \equiv \begin{cases} (z \pm \tilde{u})/|z \pm \tilde{u}| & \text{if } |z \pm \tilde{u}| > 1 \\ 0 & \text{otherwise} \end{cases} \quad (3.20)$$

and

$$D_{\pm}(z, \tilde{u}) \equiv \begin{cases} 0 & \text{if } |z \pm \tilde{u}| > 1 \\ 1 & \text{otherwise} \end{cases}. \quad (3.21)$$

In the static approximation we note that $\chi_2(q, 0) = 0$ and χ_1 reduces to Eq. (3.16), where generally

$$\varepsilon_2(q, \omega) = 2\pi B(q, \omega)\chi(q/2k_F, 2\omega m/\hbar q^2), \quad (3.22)$$

and $B(q, \omega) = \sqrt{q^2 - \varepsilon_0 \omega^2 c^{-2}}$.

3.B Exciton-electron elastic scattering

In this appendix, we outline the details of the $X + e^- \rightarrow X + e^-$ scattering calculation, including accounting for electron spin. Here, and in Appendix 3.C, we follow closely with the approach of Ref. 128, generalizing to the strict 2D limit and filling in necessary details.

In the following, it will be useful to keep in mind the electron and hole anti-commutation relations

$$\{d_{\mathbf{k}}^{s\dagger}, c_{\mathbf{k}'}^{s'\dagger}\} = \{d_{\mathbf{k}}^{s\dagger}, c_{\mathbf{k}'}^{s'}\} = \{d_{\mathbf{k}}^s, c_{\mathbf{k}'}^{s'\dagger}\} = \{d_{\mathbf{k}}^s, c_{\mathbf{k}'}^{s'}\} = 0, \quad (3.23)$$

(electrons and holes always anti-commute) and,

$$\{x_{\mathbf{k}}^s, x_{\mathbf{k}'}^{s'}\} = \{x_{\mathbf{k}}^{s\dagger}, x_{\mathbf{k}'}^{s'\dagger}\} = 0; \quad \{x_{\mathbf{k}}^s, x_{\mathbf{k}'}^{s'\dagger}\} = \delta_{\mathbf{kk}'}\delta_{ss'}, \quad (3.24)$$

where $x = c, d$. Also, recall that $\psi_{\mathbf{k}_e}$ ends up as a global phase factor in the expression for the scattering rate, and will be ignored in the following derivations.

3.B.1 General form of the matrix elements

A prudent first step to computing Eq. (3.5) is to split up V into its constituent parts and evaluate them independently on the initial state $|\mathbf{k}_x^\alpha, \mathbf{k}_e^\beta\rangle$, where spin indexes have been added as superscripts (the exciton spin references the electron; hole spin will not be important). In the case of the

electron-hole component, we have

$$V_{eh} \left| \mathbf{k}_x^\alpha, \mathbf{k}_e^\beta \right\rangle = \sum_{\mathbf{k}'\mathbf{q}'} v_{q'} \phi^* \underbrace{c_{-\mathbf{k}'+\mathbf{q}'}^{\alpha\dagger} d_{\mathbf{k}_x+\mathbf{k}'-\mathbf{q}'}^\dagger}_{\text{self-interaction}} c_{\mathbf{k}_e}^{\beta\dagger} |0\rangle - \sum_{\mathbf{k}'\mathbf{q}'} v_{q'} \phi^* c_{-\mathbf{k}'}^{\alpha\dagger} d_{\mathbf{k}_x+\mathbf{k}'-\mathbf{q}'}^\dagger c_{\mathbf{k}_e+\mathbf{q}'}^{\beta\dagger} |0\rangle, \quad (3.25)$$

where $\phi^* \equiv \phi_{\alpha_x \mathbf{k}_x + \mathbf{k}'}^*$. The first term in the above equation only contains information about the exciton interacting with itself (self-interaction) and is therefore discarded. The electron-electron component is calculated in a similar fashion and does not contain self-interaction terms:

$$V_{ee} \left| \mathbf{k}_x^\alpha, \mathbf{k}_e^\beta \right\rangle = \sum_{\mathbf{k}'\mathbf{q}'} v_{q'} \phi^* c_{-\mathbf{k}'-\mathbf{q}'}^{\alpha\dagger} d_{\mathbf{k}_x+\mathbf{k}'}^\dagger c_{\mathbf{k}_e+\mathbf{q}'}^{\beta\dagger} |0\rangle. \quad (3.26)$$

From here by direct computation we find the general matrix elements of the electron-exciton elastic scattering process to be

$$\left\langle (\mathbf{k}_x + \mathbf{q})^\theta, (\mathbf{k}_e - \mathbf{q})^\omega \middle| V_{eh} \left| \mathbf{k}_x^\alpha, \mathbf{k}_e^\beta \right\rangle \right\rangle = - \sum_{\mathbf{k}''\mathbf{k}'\mathbf{q}'} \phi_1 \phi_2^* v_{q'} (\delta_{\mathbf{k}'', \mathbf{k}'}^{\theta\alpha} \delta_{-\mathbf{q}, \mathbf{q}'}^{\omega\beta} - \delta_{-\mathbf{k}'', \mathbf{k}_e + \mathbf{q}'}^{\theta\beta} \delta_{\mathbf{k}', \mathbf{k}_e - \mathbf{q}}^{\omega\alpha}), \quad (3.27)$$

where $\phi_1 \phi_2^* \equiv \phi_{\alpha_x \mathbf{k}_x + \alpha_x \mathbf{q} + \mathbf{k}''} \phi_{\alpha_x \mathbf{k}_x + \mathbf{k}'}^*$. Explicitly, this is

$$\begin{aligned} \left\langle (\mathbf{k}_x + \mathbf{q})^\theta, (\mathbf{k}_e - \mathbf{q})^\omega \middle| V_{eh} \left| \mathbf{k}_x^\alpha, \mathbf{k}_e^\beta \right\rangle \right\rangle = \\ v_q \delta_{\theta\alpha} \delta_{\omega\beta} \sum_{\mathbf{k}'} \phi_{\alpha_x \mathbf{k}_x + \alpha_x \mathbf{q} + \mathbf{k}'} \phi_{\alpha_x \mathbf{k}_x + \mathbf{k}'} - \phi_{\alpha_x \mathbf{k}_x - \mathbf{k}_e + \mathbf{q}} \delta_{\theta\beta} \delta_{\omega\alpha} \sum_{\mathbf{k}'} \phi_{\alpha_x \mathbf{k}_x + \alpha_x \mathbf{q} - \mathbf{k}_e - \mathbf{k}'} v_{k'}, \end{aligned} \quad (3.28)$$

which can be separated into direct (corresponding to v_q) and exchange ($v_{k'}$) contributions. It is also observed that for practical computation $\phi = \phi^*$ and thus the complex conjugation is dropped. The electron-electron term is computed

$$\begin{aligned} \left\langle (\mathbf{k}_x + \mathbf{q})^\theta, (\mathbf{k}_e - \mathbf{q})^\omega \middle| V_{ee} \left| \mathbf{k}_x^\alpha, \mathbf{k}_e^\beta \right\rangle \right\rangle = \\ \sum_{\mathbf{k}''\mathbf{k}'\mathbf{q}'} \phi_1 \phi_2^* v_{q'} (\delta_{\mathbf{k}'', \mathbf{k}'}^{\theta\alpha} \delta_{\mathbf{q}, -\mathbf{q}'}^{\omega\beta} - \delta_{-\mathbf{k}'', \mathbf{k}_e + \mathbf{q}'}^{\theta\beta} \delta_{-\mathbf{k}' - \mathbf{q}', \mathbf{k}_e - \mathbf{q}}^{\omega\alpha}) \end{aligned} \quad (3.29)$$

and simplified in a similar fashion,

$$\begin{aligned} \left\langle (\mathbf{k}_x + \mathbf{q})^\theta, (\mathbf{k}_e - \mathbf{q})^\omega \middle| V_{ee} \middle| \mathbf{k}_x^\alpha, \mathbf{k}_e^\beta \right\rangle = \\ v_q \delta_{\theta\alpha} \delta_{\omega\beta} \sum_{\mathbf{k}'} \phi_{\alpha_x \mathbf{k}_x - \beta_x \mathbf{q} + \mathbf{k}'} \phi_{\alpha_x \mathbf{k}_x + \mathbf{k}'} - \delta_{\theta\beta} \delta_{\omega\alpha} \sum_{\mathbf{k}'} \phi_{\alpha_x \mathbf{k}_x + \alpha_x \mathbf{q} - \mathbf{k}_e - \mathbf{k}'} \phi_{\alpha_x \mathbf{k}_x + \mathbf{q} - \mathbf{k}_e - \mathbf{k}'} v_{k'}. \end{aligned} \quad (3.30)$$

Combining terms into direct and exchange contributions, we have

$$V^D(\mathbf{q}, \mathbf{k}_e, \mathbf{k}_x) = v_q \delta_{\theta\alpha} \delta_{\omega\beta} \sum_{\mathbf{k}'} \phi_{\alpha_x \mathbf{k}_x + \mathbf{k}'} [\phi_{\alpha_x \mathbf{k}_x - \beta_x \mathbf{q} + \mathbf{k}'} - \phi_{\alpha_x \mathbf{k}_x + \alpha_x \mathbf{q} + \mathbf{k}'}], \quad (3.31)$$

where the Kronecker delta functions ensure the proper spins are paired, and

$$V^{XC}(\mathbf{q}, \mathbf{k}_e, \mathbf{k}_x) = -\delta_{\theta\beta} \delta_{\omega\alpha} \sum_{\mathbf{k}'} v_{k'} \phi_{\alpha_x \mathbf{q} - \Delta \mathbf{k}_x + \mathbf{k}'} [\phi_{\mathbf{q} - \Delta \mathbf{k}_x + \mathbf{k}'} - \phi_{\mathbf{q} - \Delta \mathbf{k}_x}], \quad (3.32)$$

where $\Delta \mathbf{k}_x \equiv \mathbf{k}_e - \alpha_x \mathbf{k}_x$.

3.B.2 Spin contributions

Both the V_{ee} and V_{eh} terms can be split into clear direct and exchange contributions such that in the individual electron spin basis,

$$\langle \theta\omega | V | \alpha\beta \rangle = \delta_{\theta\alpha} \delta_{\omega\beta} V^D + \delta_{\theta\beta} \delta_{\omega\alpha} V^{XC}.$$

If the incident and exciton electrons are in a singlet configuration, we have to consider all contributions from the singlet state $|S\rangle = (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)/\sqrt{2}$,

$$\langle S | V | S \rangle = \frac{1}{2} (\langle \uparrow\downarrow | V | \uparrow\downarrow \rangle + \text{cc.} - \langle \uparrow\downarrow | V | \downarrow\uparrow \rangle - \text{cc.}),$$

which in the specified basis is

$$V_S \equiv \langle S | V | S \rangle = V^D - V^{XC}.$$

By inspection, any of the triplet configurations are

$$V_T \equiv \langle T|V|T \rangle = V^D + V^{XC}.$$

In the case of the exciton case, the singlet and triplet contributions are essentially identical, since the exchange contribution dominates, meaning $|V_S|^2 \approx |V_T|^2$; for the trion, we do not consider triplet states.

3.B.3 1s-1s scattering

With the assumption that the exciton wave function ϕ is in the parameterized ground state (1s) given by Eq. (3.3), the direct interaction has an analytic form. Noting that

$$\sum_{\mathbf{k}} \rightarrow \frac{A}{(2\pi)^2} \int d^2k, \quad \mathbf{k} \in \mathbb{R}^2,$$

and that the convolution

$$\int d^2k' g(\lambda k') g(\lambda' |\mathbf{q} \pm \mathbf{k}'|) = \frac{2\pi}{(\lambda + \lambda')^2} g\left(\frac{\lambda\lambda' q}{\lambda + \lambda'}\right), \quad (3.33)$$

the direct terms simplify to (dropping the spin Kronecker deltas)

$$V_{1s}^D(q) = \frac{2\pi e^2}{A q \varepsilon(q)} [g(\lambda \beta_x q / 2) - g(\lambda \alpha_x q / 2)]. \quad (3.34)$$

The exchange terms do not simplify and must be evaluated numerically,

$$\begin{aligned} V_{1s}^{XC}(\mathbf{q}, \mathbf{k}_e, \mathbf{k}_x) &= -\frac{4e^2 \lambda^2}{A} \int \frac{d^2k'}{k' \varepsilon(k')} g(\lambda |\alpha_x \mathbf{q} - \Delta \mathbf{k}_x + \mathbf{k}'|) \\ &\quad \times [g(\lambda |\mathbf{k}_1 + \mathbf{q} - \Delta \mathbf{k}_x|) - g(\lambda |\mathbf{q} - \Delta \mathbf{k}_x|)]. \end{aligned} \quad (3.35)$$

These results have been previously derived for scattering in finite quantum wells [128] and match the results above in the 2D analytic limit. Here, $\lambda = \lambda_0 = 10.3$ Å is the exciton effective Bohr radius, and $\alpha_x = m_e/M_x$ is the mass ratio of the exciton, $\beta_x = 1 - \alpha_x$, $m_e = 0.49m_0$ and $M_x = m_e + m_h$, where $m_h = 0.61m_0$.

3.B.4 2s-2s scattering

To compute the excited state (2s) exciton elastic scattering line width, we parameterize a radial 2s hydrogen wave function

$$\phi^{2s}(r, a, b) \propto \left(2 - \frac{r}{b}\right) e^{-r/2a} \quad (3.36)$$

in terms of an effective Bohr radius a and a secondary parameter b chosen to ensure orthogonality to the 1s state. An initial fit to the first excited state exact-diagonalization result of the Wannier exciton in a 2D Keldysh potential yielded length scales $a = 7.79$ Å and $b = 5.33$ Å, the latter of which was modified to $b = 6.20$ Å to ensure orthogonality. Fourier transforming to momentum-space yields

$$\phi_k^{2s} = N_{2s} \left[\frac{16\pi a^2}{(1 + 4a^2 k^2)^{3/2}} - \frac{2\pi \left(\frac{1}{2a^2} - k^2\right)}{b \left(\frac{1}{4a^2} + k^2\right)^{5/2}} \right] \quad (3.37)$$

with normalization

$$N_{2s} = \sqrt{\frac{b^2}{4\pi a^2 A(3a^2 - 4ab + 2b^2)}}.$$

Matrix elements are computed by making the substitution $\phi \rightarrow \phi^{2s}$ in Eqs. (3.31) and (3.32) numerically performing the 2D integrals.

3.C Trion-electron elastic scattering

The details of the $T+e^- \rightarrow T+e^-$ scattering process are significantly more involved than the exciton case. The introduction of an extra electron manifests as another pair of creation and annihilation operators in the matrix element evaluation and adds many more terms. While the calculation is

longer, it is no more conceptually difficult. In this appendix, we present the detailed derivation of the matrix elements for a two dimensional system, which coincide with the results for the 3D quantum well in the $L \rightarrow 0$ limit [128].

The total elastic scattering matrix element $\mathcal{V}(\mathbf{q}, \mathbf{k}_e, \mathbf{k}_t)$ is calculated by first computing the action of $V|\mathbf{k}_t, \mathbf{k}_e\rangle$. This not only simplifies the number of operator contractions, it also allows for removal of self-interaction terms (those characterized by internal interactions between electrons and holes within the trion), as they do not contribute to the scattering matrix elements (similar to that of the exciton scattering case). To begin, we first evaluate the general contraction, which is used during the evaluation of Eq. (3.13),

$$c_1 c_2^\dagger c_3^\dagger c_4^\dagger |0\rangle = [\delta_{12} c_3^\dagger c_4^\dagger - \delta_{13} c_2^\dagger c_4^\dagger + \delta_{14} c_2^\dagger c_3^\dagger] |0\rangle, \quad (3.38)$$

where in Eq. (3.39), $1 \equiv (\mathbf{k}'_1, z'_1, s')$, $2 \equiv (-\mathbf{k}_1, z_1, s_1)$, $3 \equiv (-\mathbf{k}_2, z_2, s_2)$, and $4 \equiv (\mathbf{k}_e, z_e, s_e)$, as this will be useful in computing both $V_{eh}|\mathbf{k}_t, \mathbf{k}_e\rangle$ and $V_{ee}|\mathbf{k}_t, \mathbf{k}_e\rangle$. In following calculations, hole operators will be ignored, as they do not contribute additional constraints or prefactors to the line width calculations. Moreover, $V_{eh}|\mathbf{k}_t, \mathbf{k}_e\rangle$ evaluates to

$$\begin{aligned} V_{eh} |\mathbf{k}_t, \mathbf{k}_e\rangle &= - \sum_{\substack{\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}'_1, \mathbf{q}' \\ s_1, s_2, s_e, s'}} v_{q'} \xi_S^*(s_1, s_2) \Phi_{\alpha_t \mathbf{k}_t + \mathbf{k}_1, \alpha_t \mathbf{k}_t + \mathbf{k}_2}^* \psi_{\mathbf{k}_e}^* c_{\mathbf{k}'_1 + \mathbf{q}'}^{s' \dagger} c_{\mathbf{k}'_1}^{s' \dagger} c_{-\mathbf{k}_1}^{s_1 \dagger} c_{-\mathbf{k}_2}^{s_2 \dagger} c_{\mathbf{k}_e}^{s_e \dagger} |0\rangle \\ &= - \sum_{\substack{\mathbf{k}_1, \mathbf{k}_2, \mathbf{q}' \\ s_1, s_2, s_e}} \xi_S^*(s_1, s_2) \Phi_{\alpha_t \mathbf{k}_t + \mathbf{k}_1, \alpha_t \mathbf{k}_t + \mathbf{k}_2}^* \psi_{\mathbf{k}_e}^* v_{q'} \\ &\quad \times \left\{ c_{-\mathbf{k}_1 + \mathbf{q}'}^{s_1 \dagger} c_{-\mathbf{k}_2}^{s_2 \dagger} c_{\mathbf{k}_e}^{s_e \dagger} - c_{-\mathbf{k}_2 + \mathbf{q}'}^{s_2 \dagger} c_{-\mathbf{k}_1}^{s_1 \dagger} c_{\mathbf{k}_e}^{s_e \dagger} + c_{\mathbf{k}_e + \mathbf{q}'}^{s_e \dagger} c_{-\mathbf{k}_1}^{s_1 \dagger} c_{-\mathbf{k}_2}^{s_2 \dagger} \right\} |0\rangle. \end{aligned} \quad (3.39)$$

The first two terms correspond to self-interactions between the internal electrons and holes of the trion. This is most easily seen by observing that after the action of V_{eh} on the trion-free electron state, the initial incident electron momentum \mathbf{k}_e remains unchanged in the final creation operator. In the last term, however, we see that a momentum exchange of \mathbf{q}' has taken place.

The electron-electron terms corresponding to $V_{ee}|\mathbf{k}_t, \mathbf{k}_e\rangle$ are calculated similarly. As in the

electron-hole case, we begin by performing the right-most contraction

$$c_1 c_2 c_3^\dagger c_4^\dagger c_5^\dagger |0\rangle = [\delta_{15}\delta_{24} - \delta_{14}\delta_{25}] c_3^\dagger |0\rangle + [\delta_{13}\delta_{25} - \delta_{15}\delta_{23}] c_4^\dagger |0\rangle + [\delta_{14}\delta_{23} - \delta_{13}\delta_{24}] c_5^\dagger |0\rangle, \quad (3.40)$$

where in Eqs. (3.40) and (3.41) $1 \equiv (\mathbf{k}'_2, s'_2)$, $2 \equiv (\mathbf{k}'_1, s'_1)$, $3 \equiv (-\mathbf{k}_1, s_1)$, $4 \equiv (-\mathbf{k}_2, s_2)$, $5 \equiv (\mathbf{k}_e, s_e)$. In similar fashion, $V_{ee} |\mathbf{k}_t, \mathbf{k}_e\rangle$ is thus found to be

$$\begin{aligned} V_{ee} |\mathbf{k}_t, \mathbf{k}_e\rangle &= \frac{1}{2} \sum_{\substack{\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}'_1, \mathbf{k}'_2, \mathbf{q}' \\ s_1, s_2, s_e, s'_1, s'_2}} \xi_S^*(s_1, s_2) \Phi_{\alpha_t \mathbf{k}_t + \mathbf{k}_1, \alpha_t \mathbf{k}_t + \mathbf{k}_2}^* \psi_{\mathbf{k}_e}^* v_{q'} c_{\mathbf{k}'_1 + \mathbf{q}'}^{s'_1 \dagger} c_{\mathbf{k}'_2 - \mathbf{q}'}^{s'_2 \dagger} c_{\mathbf{k}'_1}^{s'_1} c_{-\mathbf{k}_1}^{s_1 \dagger} c_{-\mathbf{k}_2}^{s_2 \dagger} c_{\mathbf{k}_e}^{s_e \dagger} |0\rangle \\ &= \frac{1}{2} \sum_{\substack{\mathbf{k}_1, \mathbf{k}_2, \mathbf{q}' \\ s_1, s_2, s_e}} \xi_S^*(s_1, s_2) \Phi_{\alpha_t \mathbf{k}_t + \mathbf{k}_1, \alpha_t \mathbf{k}_t + \mathbf{k}_2}^* \psi_{\mathbf{k}_e}^* v_{q'} \left\{ c_{-\mathbf{k}_2 + \mathbf{q}'}^{s_2 \dagger} c_{\mathbf{k}_e - \mathbf{q}'}^{s_e \dagger} c_{-\mathbf{k}_1}^{s_1 \dagger} - c_{\mathbf{k}_e + \mathbf{q}'}^{s_e \dagger} c_{-\mathbf{k}_2 - \mathbf{q}'}^{s_2 \dagger} c_{-\mathbf{k}_1}^{s_1 \dagger} \right. \\ &\quad \left. + c_{\mathbf{k}_e + \mathbf{q}'}^{s_e \dagger} c_{-\mathbf{k}_1 - \mathbf{q}'}^{s_1 \dagger} c_{-\mathbf{k}_2}^{s_2 \dagger} - c_{-\mathbf{k}_1 + \mathbf{q}'}^{s_1 \dagger} c_{\mathbf{k}_e - \mathbf{q}'}^{s_e \dagger} c_{-\mathbf{k}_2}^{s_2 \dagger} + c_{-\mathbf{k}_1 + \mathbf{q}'}^{s_1 \dagger} c_{-\mathbf{k}_2 - \mathbf{q}'}^{s_2 \dagger} c_{\mathbf{k}_e}^{s_e \dagger} - c_{-\mathbf{k}_2 + \mathbf{q}'}^{s_2 \dagger} c_{-\mathbf{k}_1 - \mathbf{q}'}^{s_1 \dagger} c_{\mathbf{k}_e}^{s_e \dagger} \right\} |0\rangle \\ &= \sum_{\substack{\mathbf{k}_1, \mathbf{k}_2, \mathbf{q}' \\ s_1, s_2, s_e}} \xi_S^*(s_1, s_2) \Phi_{\alpha_t \mathbf{k}_t + \mathbf{k}_1, \alpha_t \mathbf{k}_t + \mathbf{k}_2}^* \psi_{\mathbf{k}_e}^* v_{q'} \\ &\quad \times \left\{ c_{-\mathbf{k}_1}^{s_1 \dagger} c_{-\mathbf{k}_2 - \mathbf{q}'}^{s_2 \dagger} c_{\mathbf{k}_e + \mathbf{q}'}^{s_e \dagger} + c_{-\mathbf{k}_1 - \mathbf{q}'}^{s_1 \dagger} c_{-\mathbf{k}_2}^{s_2 \dagger} c_{\mathbf{k}_e + \mathbf{q}'}^{s_e \dagger} + c_{-\mathbf{k}_1 - \mathbf{q}'}^{s_1 \dagger} c_{-\mathbf{k}_2 + \mathbf{q}'}^{s_2 \dagger} c_{\mathbf{k}_e}^{s_e \dagger} \right\} |0\rangle \end{aligned} \quad (3.41)$$

In order to move from the second to the third equality in Eq. (3.41), we have made use of the substitutions $\mathbf{q}' \rightarrow -\mathbf{q}'$ in the first, third and fifth terms. The last term is a self-interaction exchange of momentum \mathbf{q}' between the two electrons on the trion.

With the self-interaction terms removed, and ignoring hole operators, the operation of $V = V_{eh} + V_{ee}$ acting on the trion-electron scattering state is

$$\begin{aligned} V |\mathbf{k}_t, \mathbf{k}_e\rangle &= \sum_{\substack{\mathbf{k}_1, \mathbf{k}_2, \mathbf{q}' \\ s_1, s_2, s_e}} \xi_S^*(s_1, s_2) \Phi_{\alpha_t \mathbf{k}_t + \mathbf{k}_1, \alpha_t \mathbf{k}_t + \mathbf{k}_2}^* \psi_{\mathbf{k}_e}^* v_{q'} \\ &\quad \times \left\{ c_{-\mathbf{k}_1 - \mathbf{q}'}^{s_1 \dagger} c_{-\mathbf{k}_2}^{s_2 \dagger} + c_{-\mathbf{k}_1}^{s_1 \dagger} c_{-\mathbf{k}_2 - \mathbf{q}'}^{s_2 \dagger} - c_{-\mathbf{k}_1}^{s_1 \dagger} (z_1) c_{-\mathbf{k}_2}^{s_2 \dagger} \right\} c_{\mathbf{k}_e + \mathbf{q}'}^{s_e \dagger} |0\rangle. \end{aligned} \quad (3.42)$$

The trion-electron elastic scattering matrix elements are given by the action of $\langle \mathbf{k}_t + \mathbf{q}, \mathbf{k}_e - \mathbf{q} |$ on Eq. (3.42). Executing all possible integrals analytically produces a series of terms which can be

broken into a direct component and two exchange components. We first adopt some notation: $\tilde{\lambda} = \lambda_1\lambda_2/(\lambda_1 + \lambda_2)$, is a harmonic-mean-like term which arises during convolutions e.g. in Eq. (3.33), and $\Delta\mathbf{k}_t = \mathbf{k}_e - \alpha_t\mathbf{k}_t$, where \mathbf{k}_t is the initial trion momentum and $\alpha_t = m_e/M_t$ is the ratio of the effective electron mass to that of the trion's $M_t = 2m_e + m_h$. Finally, the elastic scattering matrix elements, $\mathcal{V}(\mathbf{q}, \mathbf{k}_e, \mathbf{k}_t)$, are given by the sum of Eq. (3.43) and Eq. (3.45). The direct terms are

$$\mathcal{V}^D(q) = \frac{2\pi e^2}{q\varepsilon(q)A(1+\kappa^2)} \sum_{j=1}^5 g_j(q), \quad (3.43)$$

where

$$\begin{aligned} g_1(q) &= g(\lambda_1\alpha_t q/2)g(\lambda_2\beta_t q/2), \\ g_2(q) &= g(\lambda_2\alpha_t q/2)g(\lambda_1\beta_t q/2), \\ g_3(q) &= 2\kappa^2 g(\tilde{\lambda}\alpha_t q)g(\tilde{\lambda}\beta_t q), \\ g_4(q) &= -g(\lambda_1\alpha_t q/2)g(\lambda_2\alpha_t q/2), \\ g_5(q) &= -\kappa^2 g^2(\tilde{\lambda}\alpha_t q), \end{aligned} \quad (3.44)$$

and the exchange terms are

$$\mathcal{V}^{XC}(\mathbf{q}, \mathbf{k}_e, \mathbf{k}_t) = \frac{2e^2}{A(1+\kappa^2)} \int \frac{d^2 k'}{k' \varepsilon(k')} \sum_{j=1}^6 [G_j(\mathbf{q}, \mathbf{k}'; \lambda_1, \lambda_2) + G_j(\mathbf{q}, \mathbf{k}'; \lambda_2, \lambda_1)], \quad (3.45)$$

where

$$\begin{aligned} G_1(\mathbf{q}, \mathbf{k}'; \lambda_1, \lambda_2) &= \lambda_2^2 g(\lambda_1\alpha_t q/2)g(\lambda_2|\Delta\mathbf{k}_t - \mathbf{q}|)g(\lambda_2|\alpha_t\mathbf{q} - \Delta\mathbf{k}_t + \mathbf{k}'|), \\ G_2(\mathbf{q}, \mathbf{k}'; \lambda_1, \lambda_2) &= -\lambda_2^2 g(\lambda_2|\alpha_t\mathbf{q} - \Delta\mathbf{k}_t + \mathbf{k}'|)g(\lambda_1|\alpha_t\mathbf{q} - \mathbf{k}'|/2)g(\lambda_2|\Delta\mathbf{k}_t - \mathbf{q}|), \\ G_3(\mathbf{q}, \mathbf{k}'; \lambda_1, \lambda_2) &= -\lambda_2^2 g(\lambda_1\alpha_t q/2)g(\lambda_2|\mathbf{q} - \Delta\mathbf{k}_t + \mathbf{k}'|)g(\lambda_2|\alpha_t\mathbf{q} - \Delta\mathbf{k}_t + \mathbf{k}'|), \\ G_4(\mathbf{q}, \mathbf{k}'; \lambda_1, \lambda_2) &= \kappa\lambda_1\lambda_2 g(\tilde{\lambda}\alpha_t q)g(\lambda_1|\alpha_t\mathbf{q} - \Delta\mathbf{k}_t + \mathbf{k}'|)g(\lambda_2|\Delta\mathbf{k}_t - \mathbf{q}|), \\ G_5(\mathbf{q}, \mathbf{k}'; \lambda_1, \lambda_2) &= -\kappa\lambda_1\lambda_2 g(\lambda_1|\alpha_t\mathbf{q} - \Delta\mathbf{k}_t + \mathbf{k}'|)g(\tilde{\lambda}|\alpha_t\mathbf{q} - \mathbf{k}'|)g(\lambda_2|\Delta\mathbf{k}_t - \mathbf{q}|), \\ G_6(\mathbf{q}, \mathbf{k}'; \lambda_1, \lambda_2) &= -\kappa\lambda_1\lambda_2 g(\tilde{\lambda}\alpha_t q)g(\lambda_2|\mathbf{q} - \Delta\mathbf{k}_t + \mathbf{k}'|)g(\lambda_1|\alpha_t\mathbf{q} - \Delta\mathbf{k}_t + \mathbf{k}'|). \end{aligned} \quad (3.46)$$

Applying $\langle \mathbf{k}_t + \mathbf{q}, \mathbf{k}_e - \mathbf{q} |$ on Eq. (3.42) produces a series of integrals, many of which may be evaluated analytically. In addition, the signs, and in some cases the prefactor, of the various terms are determined by summing over the spin degrees of freedom.

As in previous calculations, it is helpful to evaluate the contraction of Fermionic operators

$$\langle 0 | c_1 c_2 c_3 c_4^\dagger c_5^\dagger c_6^\dagger | 0 \rangle = \delta_{16}(\delta_{34}\delta_{25} - \delta_{24}\delta_{35}) + \delta_{15}(\delta_{24}\delta_{36} - \delta_{34}\delta_{26}) + \delta_{14}(\delta_{35}\delta_{26} - \delta_{25}\delta_{36}), \quad (3.47)$$

where for the electron-hole interaction,

$$\begin{aligned} 1 &\equiv (\mathbf{k}_e - \mathbf{q}, s'_e), \\ 2 &\equiv (-\mathbf{k}'_2, s'_2), \\ 3 &\equiv (-\mathbf{k}'_1, s'_1), \\ 4 &\equiv (-\mathbf{k}_1, s_1), \\ 5 &\equiv (-\mathbf{k}_2, s_2), \\ 6 &\equiv (\mathbf{k}_e + \mathbf{q}', s_e). \end{aligned} \quad (3.48)$$

As an example, the terms including δ_{16} produce $g_4(q)$ and $g_5(q)$ in Eq. (3.44), and the remainder of the exchange terms correspond to G_1 and G_4 in Eq. (3.46).

To evaluate the first (second) electron-electron interactions, we replace $\mathbf{k}_1 \rightarrow \mathbf{k}_1 + \mathbf{q}'$ and $\mathbf{k}_2 \rightarrow \mathbf{k}_2 + \mathbf{q}'$ in Eq. (3.48). Each of the six terms generated by the contraction in Eq. (3.47) is evaluated individually for the hole and two electrons, generating 18 total terms and producing Eqs. (3.44) and (3.46). Note that $g_3(q)$ in Eq. (3.44) accounts for two identical electron-electron interaction terms.

Instead of presenting a derivation of all 18 terms, we present a detailed derivation of one of

them. The others follow similarly. Consider the electron-hole term corresponding to $\delta_{24}\delta_{15}\delta_{36}$,

$$-\frac{1}{2} \sum_{\substack{\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}'_1, \mathbf{k}'_2, \mathbf{q}' \\ s_1, s_2, s_e, s'_1, s'_2, s'_e}} v_{q'} \xi_S^*(s_1, s_2) \xi_S(s'_1, s'_2) \psi_{\mathbf{k}_e}^* \psi_{\mathbf{k}_e - \mathbf{q}} \\ \times \Phi_{\alpha_t \mathbf{k}_t + \mathbf{k}_1, \alpha_t \mathbf{k}_t + \mathbf{k}_2}^* \Phi_{\alpha_t(\mathbf{k}_t + \mathbf{q}) + \mathbf{k}'_1, \alpha_t(\mathbf{k}_t + \mathbf{q}) + \mathbf{k}'_2} \delta_{\mathbf{k}'_2, \mathbf{k}_1} \delta_{s'_2, s_1} \delta_{\mathbf{k}_e - \mathbf{q}, -\mathbf{k}_2} \delta_{s'_e, s_2} \delta_{-\mathbf{k}'_1, \mathbf{k}_e + \mathbf{q}'} \delta_{s'_1, s_e}. \quad (3.49)$$

Note that the factor of $1/2$ is due to an average over the initial free-electron spin states. We may sum over the dummy variables $s'_1, s'_2, s'_e, \mathbf{k}'_2, \mathbf{k}_2$ and \mathbf{q}' . This yields

$$-\frac{1}{2} \sum_{\substack{\mathbf{k}_1, \mathbf{k}'_1 \\ s_1, s_2, s_e}} v_{\mathbf{k}_e + \mathbf{k}'_1} \xi_S^*(s_1, s_2) \xi_S(s_e, s_1) \psi_{\mathbf{k}_e}^* \psi_{\mathbf{k}_e - \mathbf{q}} \Phi_{\alpha_t \mathbf{k}_t + \mathbf{k}_1, \alpha_t \mathbf{k}_t + \mathbf{q} - \mathbf{k}_e}^* \Phi_{\alpha_t(\mathbf{k}_t + \mathbf{q}) + \mathbf{k}'_1, \alpha_t(\mathbf{k}_t + \mathbf{q}) + \mathbf{k}_1}. \quad (3.50)$$

The spin factors are evaluated first:

$$\sum_{s_1, s_2, s_e} \xi_S^*(s_1, s_2) \xi_S(s_e, s_1) = \begin{cases} -1 & \text{if } s_2 = s_e \\ 0 & \text{if } s_2 \neq s_e. \end{cases} \quad (3.51)$$

After sorting each term in Eq. (3.50) by integration variable and making the substitutions $\mathbf{k}'_1 \rightarrow \mathbf{k}'_1 - \mathbf{k}_e$ and $\mathbf{k}_1 \rightarrow \mathbf{k}_1 - \alpha_t \mathbf{k}_t$, one integral may be evaluated analytically using the convolution in Eq. (3.33), yielding $G_4(\mathbf{q}, \mathbf{k}'; \lambda_1, \lambda_2)$ after the substitution $\Delta \mathbf{k}_t = \mathbf{k}_e - \alpha_t \mathbf{k}_t$ is made.

3.D Computational Details

All integrations were performed using the Cubature adaptive integration package [156]. Integrals over $(0, \infty)$ were mapped to the finite range $(0, 1)$ and performed using adaptive integration. Additionally, in order to avoid exhausting available memory, integrals were nested in the following way: First, scattering matrix elements were computed on the fly and converged to some relative error tolerance ϵ . This results in a computation of a two-dimensional integral for the exchange terms. Once the matrix element V is computed, the Golden Rule integration which contains $|V|^2$,

along with the integration over all final states, is performed (this is a three-dimensional integral), and converged to some error tolerance $c\epsilon$, where c is typically on the order of 100 – 1000. Finally, to ensure convergence, the entire computation is converged with respect to the decreasing of ϵ .

3.E Comparison to experiment

As mentioned in a footnote in Section 3.3, recent now published work from the Chernikov Group has appeared to have successfully accessed the homogeneous line of the scattering processes studied in this chapter [157]. In Fig. 3.3, we compare the results as presented in Fig. 3.1 to that of the results in Ref. 157.²

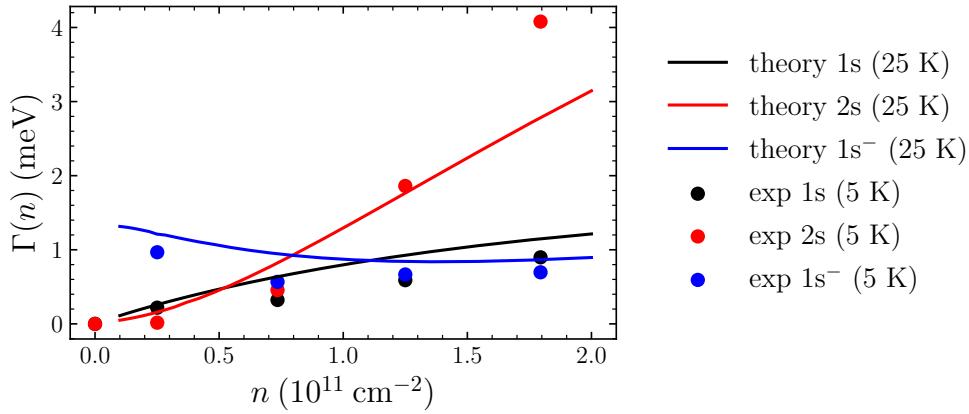


Figure 3.3: Comparison of our theoretical results with the experimental results [157]. To compare qualitatively, we performed the following transformations to the experimental data: First, the intrinsic line widths are subtracted off of the experimental results. Second, while the experiments were conducted at 5 K, we compare the experimental results to our theoretical calculations at 25 K so as to better visualize the qualitative dependence on n .

Possible thermal broadening mechanisms notwithstanding (hence the difference in temperature between the compared theoretical and experimental results so as to allow for a clearer visual), there are two qualitative similarities to note. First, the relative magnitudes of the computed broadening are extremely accurate, especially in this low-doping regime. For example, at ultra-low doping, the exciton 1s and 2s broadenings are similar, with the trion's significantly larger. At slightly larger

²We thank Alexey Chernikov for allowing us access to the raw data for presentation here.

dopings, the 2s exciton broadening is much larger than both the trion and 1s exciton. Second, the doping-dependent behavior is correctly captured. Specifically, both the 1s and 2s excitons are effectively monotonic functions of n , whereas the trion experiences a dip at low doping, and then begins to increase.

Part II

Machine learning for chem-informatics application

Chapter 4

Classification of local chemical environments from x-ray absorption spectra using supervised machine mearning

The content presented in this chapter is based closely on the following work (with primarily aesthetic changes made for additional clarity):

- Classification of local chemical environments from x-ray absorption spectra using supervised machine mearning [[arXiv](#)]

M. R. Carbone, S. Yoo, M. Topsakal* & D. Lu,* Phys. Rev. Mater **3**, 033604 (2019).

Editors' suggestion

Copyright 2019, American Physical Society

* Indicates corresponding authors.

4.1 Introduction

Knowledge of material structures at the atomic scale is essential to understanding physical phenomena and material properties that can lead to practical applications. Specifically, key information about the local chemical environment (LCE) surrounding an atom, including symmetry, coordination number, bond length and bond angle, forms the fundamental basis that determines the electronic properties of materials. In order to resolve the structure-property relationship, the characterization of atomic structures and their dynamic changes under different thermodynamic conditions has become a primary target of experimental studies. Such efforts have made tremendous impact on many research fields, including superconductivity [158], ultrafast dynamics [159], energy storage [160], and photocatalysis [161]. Recent progress in materials discovery using smart automation [162, 163] and *in situ* and *operando* experiments [164] further highlights emerging challenges and opportunities of materials characterization in *real time*.

Amongst many experimental techniques (e.g. imaging, diffraction, and spectroscopy), the x-ray absorption near edge structure (XANES) is a premier tool for probing LCEs, because it is element specific, sensitive to local structural and electronic properties, and applicable under harsh experimental conditions [165–167], making it a robust structure refinement method [165, 168–172]. Given the atomic arrangement of a sample (\mathbf{x}), its XANES spectra (\mathbf{y}) can be determined through quantum mechanical laws (f) via the mapping $\mathbf{y} = f(\mathbf{x})$. Extracting the information of the LCE ($\tilde{\mathbf{x}}$) as a subset of \mathbf{x} from spectral data can be formulated as an inverse problem: $\tilde{\mathbf{x}} = f^{-1}(\mathbf{y})$. The solution of this inverse problem is highly nontrivial, because the spectral information in experimental XANES is not only abstract, but also averaged over the whole sample. Consequently, much of the success in the past has been achieved by using fingerprints established from empirical observations.

In this study, we focus on $3d$ transition metal K-edge XANES, which carries rich information about the electronic transitions from the $1s$ core level of the absorbing atom to unoccupied states. Since the 1940’s, extensive research has been carried out to correlate spectral features of K-edge

XANES spectra, especially in the pre-edge region, to LCEs [173, 174]. For example, Hanson *et al.* [175] observed distinct chemical shifts in the absorption edge of Mn K-edge XANES in Mn, MnS, MnO₂ and KMnO₄. Wong *et al.* [176] showed linear relationships between the oxidation state of V and both pre-edge and absorption edge positions in the K edge. Farges *et al.* [177–179] and Jackson *et al.* [180] conducted comprehensive studies of the correlation between pre-edge features and the coordination number in Ti, Fe and Ni compounds; they found that the pre-edge peak intensity decreases with increasing coordination number. For fixed coordination number, early 3d transition metal elements (Ti, V, Cr and Mn) have stronger pre-edge peaks than late transition metal elements (Fe, Co, Ni and Cu) overall [174]. Furthermore, while both pre-edge peak locations and intensities in Ti [177] and Ni species [179] exhibit a significant dependence on the coordination number, the pre-edge peak positions in Fe compounds are independent of coordination number [180].

From a theoretical standpoint, the pre-edge peak intensity can be understood qualitatively from quantum mechanical selection rules. The dominant contribution in K-edge XANES comes from $s \rightarrow p$ dipole transitions, as the $s \rightarrow d$ quadrupole terms are generally orders of magnitude smaller. The density of states corresponding to the pre-edge regions of 3d transition metals are derived primarily from their empty 3d bands, and direct $s \rightarrow d$ transitions are dipole-forbidden, which implies a vanishing peak intensity. However, pre-edge peak intensity is enhanced when atomic, unoccupied p and d states hybridize. According to group theory, atomic $p - d$ mixing is allowed under T_d symmetry, but is forbidden under O_h symmetry [181, 182]. As a result, 3d transition metals with tetrahedral geometries tend to exhibit stronger pre-edge peak intensities than those with octahedral geometries. To this end, empirical diagrams have been compiled to classify four-, five- and six-coordinated Ti, Ni and Fe based on pre-edge peak positions and intensities [177–180]; we will refer to this method as the empirical fingerprint approach.

Despite the wide range of applications of the empirical fingerprint approach, including classifying LCEs in crystals, amorphous systems [178, 183] and catalysts [184], it has several limitations that may hinder its practical applications in the broader materials domain. First, coordination num-

ber is not the only factor that affects pre-edge peak features. Quantitative pre-edge features are determined by multiple factors, including coordination number, local distortion, oxidation state, and the nature of the ligands [185]. For example, local distortions, e.g. displacements from the inversion center in octahedral geometries, under the crystal field can lower the local symmetry and enable atomic $p - d$ mixing, resulting in dramatic enhancement of pre-edge peak intensity [185]. Such local distortion-induced pre-edge peak intensity enhancement has been reported in the V K-edges of six-coordinated MgV_2O_6 [186] and $\text{NaV}_{10}\text{O}_{28}$ [187], and in the Ti K-edge of six-coordinated $\text{Li}_4\text{Ti}_5\text{O}_{12}$ [188]. Therefore, isolating pure LCE effects and extracting robust correlations between the LCE and simple spectral descriptors, although valid for exemplary systems, may not be feasible for more structurally complex ones.

Secondly, the empirical fingerprint approach relies on human knowledge to engineer spectral descriptors, which may introduce bias. For example, existing spectral descriptors are primarily derived from the pre-edge region (e.g. peak positions and intensities). However, it is known that pre-edge features are much less visible in late transition metals than early transition metals [174]. Therefore, the existing empirical fingerprint approach may not work effectively for late transition metals due to poor spectral contrast in the pre-edge region. One may need to systematically explore main- and post-edge spectral features in order to engineer and optimize new descriptors, which may not necessarily be simple ones, to tackle this problem.

Machine learning (ML) methods are a promising candidate to solve this inverse materials characterization problem. Instead of relying on empirical features derived from a small number of human observations, ML methods are data-driven approaches that make predictions based on large training sets, eliminating human bias from the feature selection process. There are myriad successful examples of the utilization of ML methods in condensed matter physics, materials science and chemistry, including methods to solve many-body problems [189], predict quantum phase transitions [190], generate force field potentials [8], design new catalysts [191], and perform structure refinement [192, 193]. In the context of XANES, one expects ML algorithms to *learn* spectral descriptors in the full energy range of the spectrum and weight them appropriately for robust LCE

predictions.

In this study, we tackled the LCE classification problem using supervised ML applied to a wide energy range of the XANES spectra (~ 50 eV above the onset) as input. In this way, the spectral feature space was systematically explored in order to establish the relationships between XANES spectra and LCE classes, specifically the local atomic geometries. As proof-of-principle, we applied ML algorithms to synthetic K-edge XANES spectra obtained from high throughput *ab initio* calculations. This study serves as a precursor to a potentially very powerful tool for real time structure refinement using experimental XANES, which will require in-depth understanding of the accuracy of the theory and further improvement of the ML algorithms.

4.2 Methods

The workflow of the element-specific, spectrum-based LCE classification framework is summarized in Figure 4.1, which contains three core modules: *data acquisition*, *LCE class labeling* and *training of machine learning models*. We stress that the workflow we developed can be adapted to a wide range of elements characterized by different spectroscopic techniques, as long as the spectral information is element specific and sensitive to the LCE such that there exists distinguishable spectral contrast associated with different LCEs. Below, we describe each module in detail.

4.2.1 Data acquisition

For any given element, the first step is to extract atomic structures representing different LCEs from existing materials structure databases. The structural database must be large enough to build a reasonably-sized training set for machine learning models. To demonstrate the applicability of the LCE classification framework, we have considered eight 3d transition metal elements (Ti, V, Cr, Mn, Fe, Co, Ni, and Cu) and extracted all available oxide structures that have been structurally optimized using density functional theory (DFT) from the Materials Project Database [194–196].

Once the structural database is established, the next step is to generate the corresponding spec-

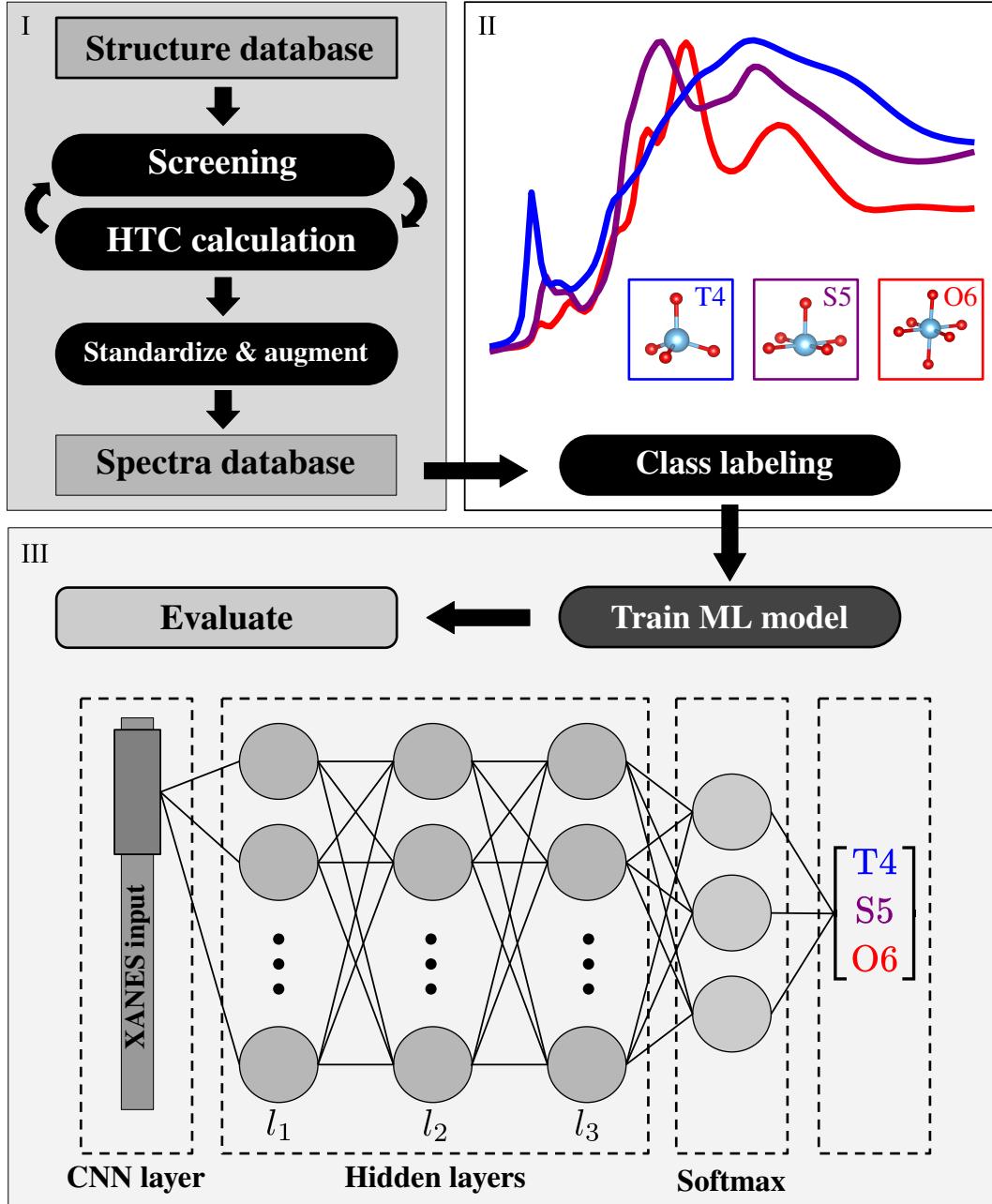


Figure 4.1: Workflow of the spectrum-based local chemical environment classification framework using supervised machine learning, which contains three modules: (I) data acquisition supplemented by high throughput computing (HTC) calculations, (II) labeling and (III) training of machine learning models. The machine learning architecture (set of hyperparameters) used in this work is shown in Module III. Notably, the model consists of an optional convolutional layer (shown in orange) followed by three hidden layers l_1 , l_2 and l_3 consisting of 90, 60 and 20 neurons, respectively, ending with a softmax output. Further details of the network are described in Subsection 4.2.3.

tral database. We focused on the K-edge XANES, as it is element specific and sensitive to the LCE (e.g., symmetry, charge state, and coordination number). In principle, one may populate the spectral database entirely with experimental spectra, but this strategy suffers from several drawbacks. First, experimental XANES spectra represent an average of signals (site-averaged signals) from each absorbing site (site-specific signals). In order to identify the correlations between spectra and local structures using ML, it is necessary to use site-specific spectra for training, as they possess much stronger spectral contrast than the site-averaged spectra. Second, when developing LCE classifiers using supervised ML methods, one needs to label XANES spectra with LCE descriptors, which means that only experimental spectra of known structures can be selected for the database. This requirement severely limits the pool of candidates for the database to mostly well-characterized crystal structures. As a result, qualified experimental spectra represent only a small fraction of the targeting LCEs in the local configuration space, which are heavily weighted in known crystals and under-represent the materials space of amorphous systems, surfaces, interfaces and nanoparticles. Consequently, a pure experimental spectral database suffers from data availability and data heterogeneity issues.

On the other hand, combined with available structural databases and well-established structure sampling methods, computational XANES have a clear advantage in exploring the LCE space and producing site-specific spectra. Furthermore, recent development in computational XANES modeling [197–207] has made it feasible to contrast experimental spectra quantitatively, enabling accurate local structure refinement of nanoparticles [192, 208], interfaces [209], dopant sites [210, 211] and structural phase transformation [188, 193] using computational techniques. An accurate and computationally efficient first-principles XANES method would be an ideal choice for sampling the vast LCE parameter space and mitigating data availability and heterogeneity issues. Indeed, computational XANES databases have recently emerged as a new tool for fast structure screening through data mining [212–214].

In this study, we generated the computational XANES database with the FEFF9 code [215], which is a popular and computationally efficient method based on multiple scattering theory.

We utilized the existing site-specific x-ray absorption spectroscopy FEFF library in the Materials Project [212, 213] and only calculated spectra not contained in this library. In the first data standardization step, site-specific spectra that failed sanity checks were automatically discarded, such as when the FEFF calculations did not converge with the default input or when the output FEFF spectra are not physical (e.g., with negative absorption coefficients). In the second step, we removed “duplicates”, which otherwise would have introduced bias if nearly identical structures were selected for both training and testing. We used a site-symmetry finder from the PYMATGEN library [195] to determine which sites in a crystal structure are symmetrically equivalent. For every pair of spectra, one was removed from the dataset if the average mean absolute difference between them was less than 0.015, a number chosen based on visual inspection of a large number of similar spectra. The process of removing duplicates also has the benefit of reducing the total number of necessary FEFF calculations to populate the XANES database. Finally, calculated XANES spectra were spline-interpolated onto an absorbing site-specific energy grid, so that the input feature vector was standardized for each spectrum. For each type of absorbing site, the energy grid was chosen such that it contains the maximum amount of available information with an energy resolution of approximately 0.5 eV.

All standardized data before augmentation are henceforth referred to as the base dataset. Training data were augmented by shifting the spectra by ± 1 and ± 2 eV. The size of the augmented training set thus becomes five times the size of the base dataset. We found that data augmentation improves the accuracy and robustness of the machine learning model.

4.2.2 Local chemical environment class labeling

Feature engineering of LCEs is an active research topic with increasing applications in a variety of areas including, for example, neural network potential development [8, 216–218]. Among many possible choices, labels based on the coordination environment (e.g. tetrahedral, square pyramidal, and octahedral geometries), although simple, provide key information on chemical bonding and have been widely used in the x-ray spectroscopy community. In this study, we utilized the

continuous symmetry measure (CSM), developed by Avnir and Pinsky [219] and hosted in the ChemEnv package [220], to measure the similarity between an input local geometry and a particular polyhedron. The smaller the CSM for a polyhedron, the more the input geometry resembles it. We applied a cutoff such that atoms further away than 1.2 times the nearest neighbor distance from the absorbing site were not considered. This cutoff was chosen as a balance between prediction accuracy and computational cost. The CSM was applied to each absorbing site, and the polyhedron with the lowest CSM value was chosen as the site LCE label.

We restricted the LCE labels to only tetrahedral (T4), square pyramidal (S5) and octahedral (O6) geometries, because across the eight transition metal families these are the most abundant LCEs, often by an order of magnitude more than the rest. The class breakdown of the dataset is presented in Table 4.1. The total number of site-specific spectra is on average a few thousand per atom type, with V (3366) and Mn (3493) the most abundant and Cu (839) the least abundant. It should be noted that one can expand the dataset by adding new structures from other material databases, generating artificial structures or introducing additional class labels. Furthermore, amongst all three chosen classes, O6 dominates, making up about 64% of the entire structure database. The impact of the inhomogeneity of the data distribution on the predictive power of the ML model is discussed in Section 4.3.

Table 4.1: The distribution of classes in the structure database used for training the machine learning models.

Absorber	T4	S5	O6	Total
Ti	271	359	1562	2192
V	948	412	2006	3366
Cr	396	121	902	1419
Mn	502	657	2334	3493
Fe	797	319	1874	2990
Co	583	227	1428	2238
Ni	246	163	1238	1647
Cu	290	183	366	839
Total	4033	2441	11710	18184

4.2.3 Training machine learning models

The core machine learning algorithm (see Figure 4.1) consists of an optional 1-D convolutional layer followed by three fully connected, feed-forward hidden layers with 90, 60 and 20 neurons, ending with a softmax output layer of 3 neurons. The input layer of the neural network is the XANES spectrum scaled to zero mean and unit variance on a standardized grid of 100 entries, and the output determines the target vector, which contains the probabilities of the three LCE classes (T4, S5 and O6) computed from the softmax function. All neurons use the rectified linear unit (ReLU) activation function and a 30% dropout to guard against over-fitting. ML models with and without the 1-D convolutional layer are referred to as the convolutional neural network (CNN) and multi-layer perceptron (MLP), respectively. The optional convolutional layer contains 8 filters and a kernel (sliding window) size of 10, stride of 1 and max-pooling size of 2, and takes as input spectral data processed in an identical manner to that of the MLP. CNNs inherently assume correlations between nearby data points, and being a down-sampling and pooling technique, sacrifice resolution in favor of invariance to the precise location of input data. The algorithm determines trained parameters by minimizing a categorical cross-entropy loss function using the Adam optimizer [221]. Mini-batch sizes of 32 were used during 50 full passes (epochs) of the training data. All training and evaluations were performed using Keras [222] with a TensorFlow [223] backend.

For each absorbing site, we used statistical bootstrapping: 90% of the database was used for training ML models and the remainder for testing. These subsets were selected randomly in a stratified manner, meaning that the proportion of each class in both the training and testing sets was always the same. In order to generate a statistical estimate on the accuracy of the classifier, we sampled the testing data with replacement over 10 folds and report the averaged results in Figure 4.3. To make full use of all available information, data included once in a testing set, were not used in any future testing sets. We found that testing results are mostly invariant to the chosen neural network architecture assuming enough training parameters were included. Therefore, the fixed 3-layer MLP with an optional convolutional layer (and associated hyperparameters) was used

throughout all experiments.

4.3 Results and discussion

In the following, we present our LCE classification study through visual inspection, principal component analysis (PCA) [224] and analysis of the machine learning classifiers (MLCs). We demonstrate that MLCs can accurately predict LCE classes from synthetic XANES data generated by the FEFF9 code. We further discuss the relevance of this study based on synthetic data to the real challenge of the LCE classification of experimentally measured XANES spectra.

4.3.1 Visual inspection of the spectral database

The FEFF K-edge XANES database of eight 3d transition metal elements (from Ti to Cu) is shown in Figure 4.2, color coded by LCE class (T4: blue; S5: green; O6: red). There are noticeable trends in the raw spectra that can be detected by visual inspection, prior to a more in-depth analysis. Overall, early 3d transition metals (e.g. Ti, V, Cr, and Mn) show more intense pre-edge peaks than late 3d transition metals (e.g. Ni and Cu), consistent with the trend from experiment [174]. Notably, T4 in Ti, V and Cr oxides exhibit sharp pre-edge peaks at about 4970, 5470, and 5995 eV, respectively. The pre-edge peak intensity decreases as the coordination number increases, consistent with the observations of Farges *et al.* [177–179] and Jackson *et al.* [180]. Such qualitative agreement between theory and experiment suggests that spectral analysis of the FEFF database is physically insightful, especially for the LCE classification problem.

In addition to pre-edge features, across the eight elements T4 exhibits the highest *post-edge* intensity, followed by S5 and finally O6. However, the role of post-edge features in LCE classification has not yet been explored in the literature, which could be an important supplement to existing pre-edge based methods. We expect that algorithms including a wide energy range in the XANES spectra can in principle improve the spectral sensitivity to the LCE as compared to those relying solely on pre-edge features.

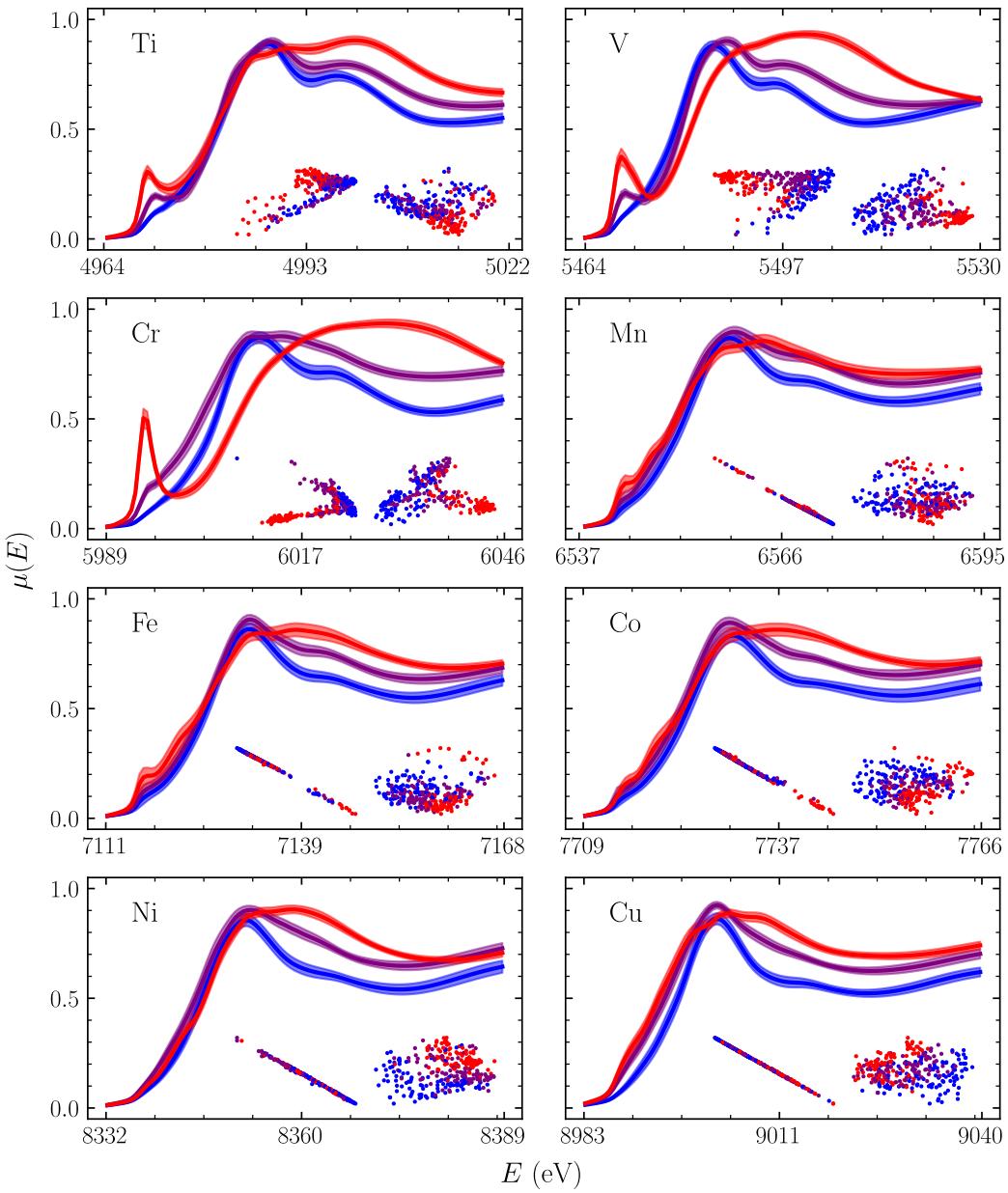


Figure 4.2: FEFF K-edge XANES database of eight 3d transition metal families. Spectra were spline-interpolated onto discretized grids of 100 points, and scaled on the vertical axis such that the maximum value is 1 (prior to shifting the mean to 0 and scaling to unit variance). The average spectrum for each class, with its scaled spread $\sigma/5$ is shown. Principal component analysis of two regions is shown in the two insets: the full feature space (lower right), and only the pre-edge region (lower center, discussed in Section 4.3), with an equal sampling of the three classes displayed to aid visual clarity. The x and y -axes correspond to the first and second principal axes. Classes are color coded as follows: blue for tetrahedral (T4), purple for square pyramidal (S5) and red for octahedral (O6).

4.3.2 Principal component analysis of the spectral database

We further analyzed the spectral database with PCA. Following the standard notation, X^k is defined as the full set of spectral data for the k th absorbing species with \mathbf{x}^{jk} being the j th spectrum in the dataset (a single feature vector input) after taking zero sample mean and unit variance. Denote \mathbf{w}^{1k} and \mathbf{w}^{2k} as the first two principal axes in the feature space. We computed coordinates of spectrum j in the PCA plot, $\mathbf{z}^{jk} = z_1^{jk} \hat{x} + z_2^{jk} \hat{y}$, as

$$z_\alpha^{jk} = \frac{\mathbf{x}^{jk} \cdot \mathbf{w}^{\alpha k}}{\max_l |z_\alpha^{lk}|}, \quad \alpha = 1, 2, \quad (4.1)$$

where for clarity the denominator scales z_α^{jk} within $[-1, 1]$.

To evaluate the significance of the pre-edge features, we truncated the principal axes by applying a cutoff n_c to the spectra, such that $x_{n_c}^{jk}$ correspond to the vertical dashed lines in Figure 4.2. Then PCA was performed for only the pre-edge region along the truncated principal axes,

$$\tilde{z}_\alpha^{jk} = \frac{\sum_{n=1}^{n_c} x_n^{jk} w_n^{\alpha k}}{\max_l |\tilde{z}_\alpha^{lk}|}, \quad \alpha = 1, 2, \quad (4.2)$$

by excluding features beyond $x_{n_c}^{jk}$. The axes in the plots generated by Eqs. 4.1 and 4.2 are scaled in the same way, so that their clustering patterns can be compared directly. Similar patterns are expected from the full and pre-edge PCA plots, if the pre-edge features dominate the spectral contrast. On the other hand, if the pre-edge features are less significant, there will be weak correlations between two sets of PCA patterns.

Full PCA plots are shown in the lower right insets of Figure 4.2. Overall, a large degree of clustering is realized, consistent with the observation of distinguishable spectral features from visual inspection. In Ti, V, Cr, Mn, Fe, and Co, most of the T4 points are located in the lower right corner and O6 points in the upper left corner. The T4 and O6 points of the Ti, V, and Cr can be easily separated in PCA plots due to their sharp (T4) and negligible (O6) pre-edge features. In the PCA plots of Mn, Fe and Co, there is also a secondary cluster of T4 points located in the upper

right corner. The data distributions in Ni and Cu are similar to the others, but with a slightly smaller packing density. In most cases, S5 clusters are intertwined with the other two classes, flanked by T4 on one side and O6 on the other.

The lower center insets in Figure 4.2 show the PCA of the pre-edge region. All classes appear less clustered than the PCA patterns of the full feature space. While in this case information contained in the feature space has clearly been reduced, this effect is less prominent in the early transition metal elements, which still exhibit a large degree of clustering due to the significant spectral contrast in the pre-edge region. On the contrary, elements such as Co, Ni and Cu exhibit such severe information loss that PCA data points collapse into a linear pattern, which is detrimental to the MLC performance, especially in systems that already exhibit weak spectral contrast.

In summary, visual inspection and PCA suggest that a MLC is likely able to accurately learn the trends in XANES spectra and correlate them to their respective classes. However, it is unclear whether MLCs can perform equally well for every class in all of the transition metal species we studied.

4.3.3 Machine learning classifier performance

The accuracy of the MLCs for each LCE class is reported using the F_1 score, which is the harmonic mean of the precision P and recall R ,

$$F_1 = \frac{2PR}{P+R}, \quad P = \frac{t_+}{t_+ + f_+}, \quad R = \frac{t_+}{t_+ + f_-}, \quad (4.3)$$

where t_+ , f_+ and f_- represent the true positives, false positives and false negatives in a two-class (2 by 2) confusion matrix.

In order to make a fair assessment of the MLC performance, we need to address the data imbalance issue in our training set. As seen in Table 4.1, the number of LCEs that conform to the O6 geometry vastly outnumbers the others, indicating that the accuracy of each class alone might not be the most reliable metric, as it may be biased due to class imbalances in the training data. We

address this problem in two ways: by using the F_1 score instead of the accuracy on a class-by-class basis, and reporting the macro F_1 score as a representative metric.

In general, the F_1 score is a much stricter metric than the accuracy and is a better indicator of performance. It accounts for both the precision P (of all predicted positives, cases that are actually positive) and recall R (out of all the actual positives, cases that are correctly identified) and dramatically penalizes poor scores in either category (contrary to the mean of the precision and recall). To demonstrate why this is important, consider the F_1 score of the relatively underrepresented S5 class. Suppose that there are 10 S5 and 100 T4 and O6 in the data set, and that the classifier has a 10% false negative and false positive rate. Accuracy would naturally be 90%, but this is a poor representation of the classifier, since 10 non-S5 data points were predicted as S5, unnaturally inflating the number of predicted positives. On the contrary, the F_1 score of 62% accounts for this by incorporating the low precision (47%) into the metric. In addition to a breakdown by class, the *macro* F_1 score (\bar{F}_1) is reported, which is the average of the class-wise F_1 scores computed using a one-versus-all approach. The \bar{F}_1 score treats each class on equal footing and further penalizes classifying data in an underrepresented class incorrectly relative to a class with many data points.

Table 4.2: \bar{F}_1 scores for different absorbing species, as the averages of the class-wise F_1 scores (red, green and blue bars) presented in Figure 4.3, both with (CNN) and without (MLP) the convolutional layer. Comparisons are made between models trained from the full feature space and reduced feature space corresponding to only the pre-edge region of the spectra.

Element	MLP	CNN	pre-MLP	pre-CNN
Ti	0.83(2)	0.84(2)	0.73(4)	0.78(3)
V	0.86(1)	0.86(2)	0.77(2)	0.79(3)
Cr	0.87(2)	0.87(3)	0.72(4)	0.75(4)
Mn	0.83(2)	0.85(2)	0.62(4)	0.68(3)
Fe	0.85(2)	0.86(3)	0.57(2)	0.63(3)
Co	0.85(3)	0.87(2)	0.59(3)	0.64(3)
Ni	0.87(1)	0.88(3)	0.61(5)	0.66(4)
Cu	0.86(2)	0.86(2)	0.50(4)	0.64(7)
Average	0.85(1)	0.86(1)	0.64(1)	0.70(1)

As clearly shown from the \bar{F}_1 scores in Table 4.2, MLCs can classify the LCEs of all eight

3d transition metal families very accurately. Uncertainties reported in the last digit of Table 4.2 and error bars in Figure 4.3 correspond to the standard deviation calculated from ten different trained models. CNNs and MLPs perform equally well, with very close \bar{F}_1 scores of 0.86 and 0.85, respectively. The class-wise F_1 scores are plotted in Figure 4.3. Notably, MLCs can reach over 90% accuracy on the T4 (CNN: 0.92; MLP: 0.92) and O6 classes (CNN: 0.96; MLP: 0.95), which can be understood from the observation of raw spectra. The strong pre-edge peak intensity is a signature of the T4 configuration in, e.g., Ti, V and Cr. Conversely, the lack of a significant pre-edge peak is a clear indicator of an O6 configuration. In these cases, it is likely that the pre-edge features are sufficient to distinguish O6 from T4. On the contrary, the spectral contrast between T4 and O6 is very low in the pre-edge region in late transition metal elements, especially in Ni and Cu. It is remarkable that MLCs can achieve the same accuracy for T4 and O6 in late transition metal elements. Such a universally good performance underscores the ability of the MLCs to extract spectral descriptors without human bias in the full energy range, including the pre-, main- and post-edge regions. Moreover, the relatively small overall error margin is a testament to the reliability and robustness of the classifier across many trained models.

Relative to T4 and O6, the S5 classification is less successful, with an overall accuracy of ~ 0.70 (CNN: 0.71; MLP: 0.68), as shown in Figure 4.3. The weaker performance of MLCs on the S5 class can be explained by Figure 4.2, where data associated with the S5 class lay between those in T4 and O6 in both the spectral and principal component space, making them more difficult to identify.

4.3.4 Importance of features beyond the pre-edge

The ability of MLCs to accurately classify late transition metal oxides that lack prominent pre-edge features suggests that features beyond the pre-edge region play an important role in the neural network model. This hypothesis is supported by the PCA results shown in the insets of Figure 4.2. In the late transition metals, while the full spectra can be effectively clustered in two-component PCA, the same analysis of pre-edge spectra displays a completely different linear pattern resulting

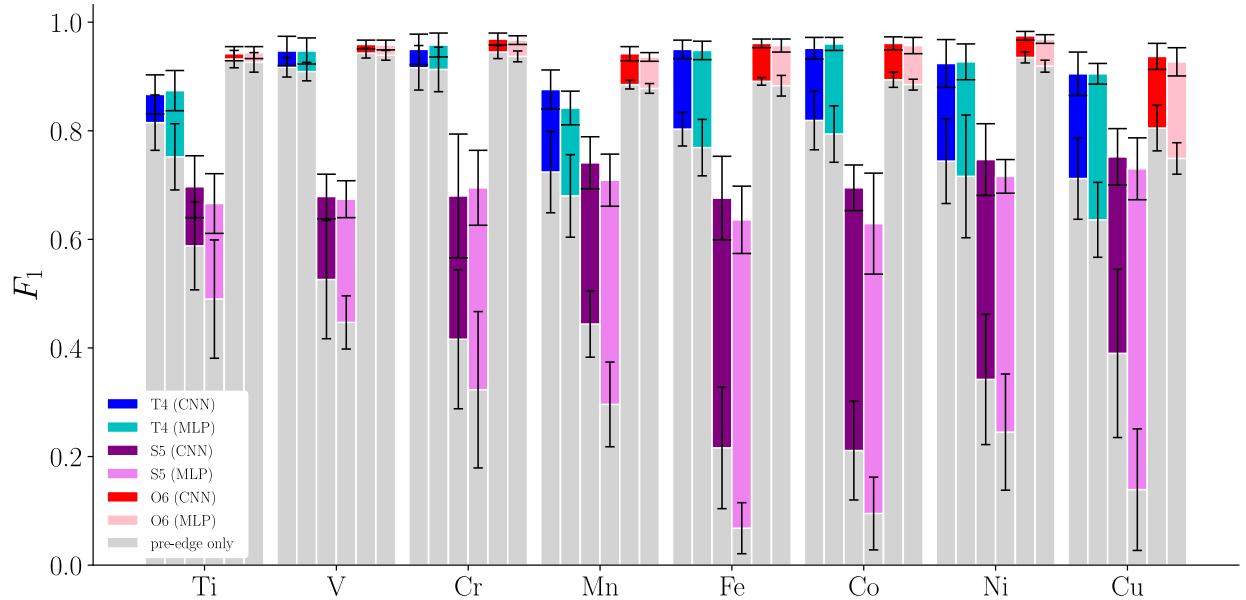


Figure 4.3: Class-wise F_1 scores calculated using different machine learning models (CNN/MLP) for T4 (blue/cyan), S5 (purple/light purple) and O6 (red/pink) local coordination environments in eight 3d transition metal elements. While the full height of each bar represents the results trained on the full feature space, gray bars overlaid on top with lower F_1 values represent the results for the models trained only on the pre-edge region. For example, for the Co S5 CNN, the F_1 score reported for training on the full spectral space is about 0.7, but decreases sharply to about 0.2 when trained on the pre-edge region only. Error bars correspond to standard deviation; the ones with wider caps correspond to the full feature space.

from substantial information loss.

To gain further insight, we train MLCs with identical architectures using *only* the pre-edge region defined by energies below the dashed lines in Figure 4.2, which we refer to as the pre-MLCs (pre-CNNs and pre-MLPs). In principle, if \bar{F}_1 scores of the pre-MLCs are close to those trained on the full spectra, then the pre-edge features are sufficient to classify the LCE for that absorbing element. Conversely, a significant drop in the \bar{F}_1 scores of the pre-MLCs would be a clear indication that features beyond the pre-edge region play a significant role in the MLCs.

As shown in Table 4.2, the average \bar{F}_1 score in pre-MLCs drops significantly by about 20% (from 0.86 to 0.70 in CNN and from 0.85 to 0.64 in MLP), as compared to MLCs trained on the full spectral space. The class-wise F_1 scores of pre-MLCs which are consistently lower than that of regular MLCs are shown in Figure 4.3 as the gray bars. We quantify this accuracy degradation

by

$$\Delta = F_1(\text{MLC}) - F_1(\text{pre-MLC}) \quad (4.4)$$

and summarize the results averaged over early (Ti, V, Cr, and Mn) and late transition metal elements (Fe, Co, Ni and Cu) in Table 4.3. First, Δ in late transition metals is more than doubled compared to early transition metals. Second, among the three classes, Δ of O6 is the smallest (< 0.10). It increases significantly for T4 in late transition metals to 0.16 in the pre-CNN (0.21 in the pre-MLP) and finally reaches the largest values for S5, at 0.22 (0.30) in early transition metals and 0.44 (0.57) in late transition metals. The results in Figure 4.3 and Tables 4.2-4.3 highlight the critical importance of features beyond the pre-edge region in accurately classifying LCEs, especially for late transition metals. The effects are the largest in the S5 class, which is rather characterless in the pre-edge region, showing neither very strong (like T4) nor very weak (like O6) pre-edge intensities.

Table 4.3: The class-wise difference between the F_1 score evaluated over the entire feature space and over only the pre-edge region (Δ). Results are averaged over the early (Ti, V, Cr and Mn) and late (Fe, Co, Ni and Cu) transition metal elements.

		T4	S5	O6
pre-MLP	Early	0.09(3)	0.30(5)	0.03(1)
	Late	0.21(4)	0.57(5)	0.10(1)
pre-CNN	Early	0.07(3)	0.22(5)	0.03(1)
	Late	0.16(3)	0.44(7)	0.08(1)

We note that unlike the case of regular MLCs, the pre-edge CNN averaged over all absorbing species ($\bar{F}_1 = 0.70$) outperforms the corresponding pre-edge MLP ($\bar{F}_1 = 0.64$) substantially. In the most extreme situation of S5, the pre-CNN ($F_1 = 0.48$) is 23% more accurate than the pre-MLP ($F_1 = 0.39$) in early transition metals, and it is more than doubled in late transition metals with $F_1 = 0.28$ (0.12) for the pre-CNN (pre-MLP). The substantially better performance of the pre-CNN is likely caused by the use of the convolutional filter, which makes the CNN able to learn subtle pre-edge features from augmented data more effectively than the MLP.

4.3.5 Discussion

The MLCs described so far were trained on computational FEFF XANES spectra. Developing MLCs that can classify the LCE of a broad range of material families using experimental XANES spectra is a more challenging task that is beyond the scope of the current work. Nonetheless, in this section we discuss several key issues that need to be addressed in order to achieve this goal, including validation of the theory, edge alignment of the simulated spectra, and the variations in the spectral intensity.

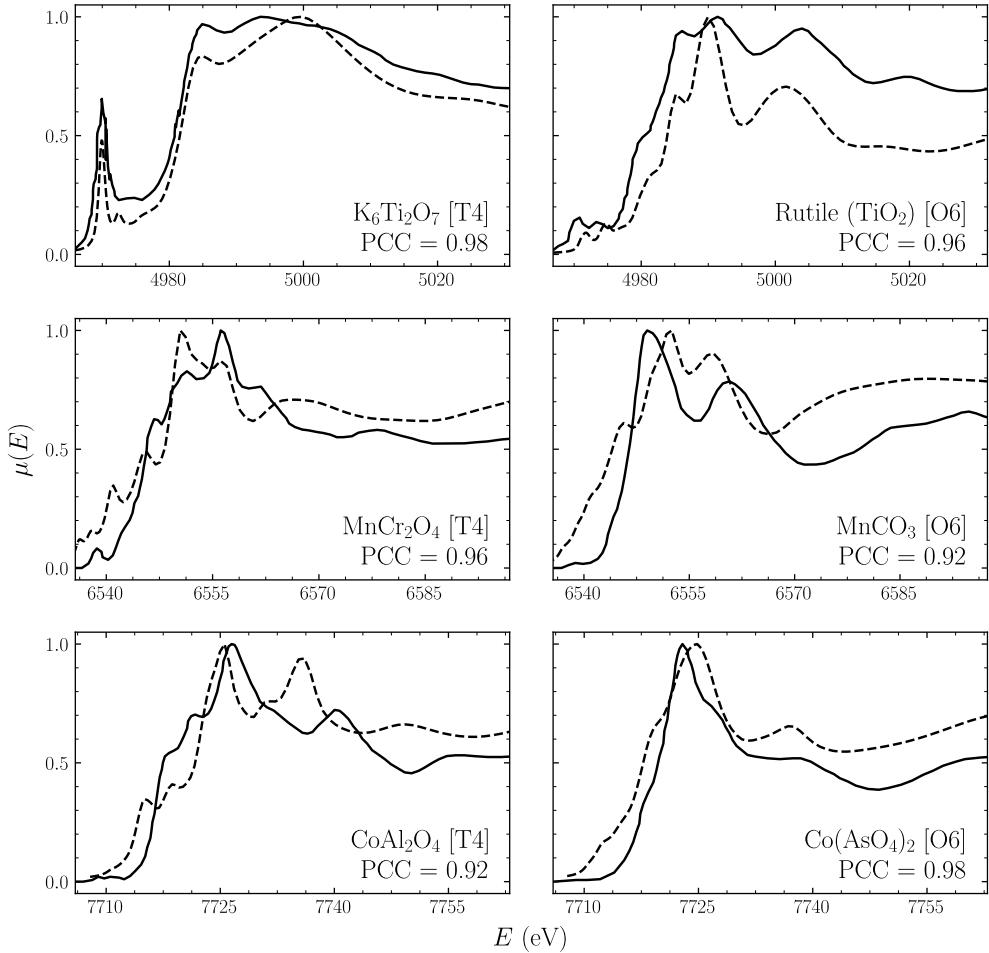


Figure 4.4: Comparison between representative experimental (solid) and FEFF (dashed) XANES spectra. FEFF calculations are shifted on the energy axis to maximize such that the Pearson correlation coefficient (PCC) in order to find the best match between experimental and theoretical XANES. The experimental spectra of $\text{K}_6\text{Ti}_2\text{O}_7$ and rutile were extracted from Farges *et al.* [178], and the experimental spectra of both pairs of Mn and Co oxides from Manceau *et al.* [225].

In order to apply MLCs trained on synthetic data to experimental spectra, it is very important to validate the theory such that the computational spectra can faithfully reproduce experimental spectral features. To this end, we compare FEFF spectra with experimental spectra on a small number of oxides: $K_6Ti_2O_7$, rutile (TiO_2), $MnCr_2O_4$, $MnCO_3$, $CoAlO_4$ and $Co(AsO_4)_2$ in Figure 4.4. This list is not meant to be exhaustive. Within this small sample, while the overall shape and major peaks are well reproduced by FEFF, there are noticeable differences in the spectral details, including the peak positions and relative intensities of different peaks. Furthermore, the degree of agreement is system-dependent. As shown in Figure 4.4, the optimal Pearson correlation coefficients (PCCs) between FEFF and experiment range from 0.92 to 0.98.

Despite the relatively high PCC scores, MLCs trained on spectra at the FEFF level of theory as such cannot reliably classify experimental spectra. It is necessary to generate the computational XANES database with more accurate methods and conduct a systematic benchmark of theory against experiment. However, the computational expense of these calculations grows quickly with the complexity of the methods, which could in practice limit the level of theory used to generate the training set. A good compromise would involve developing robust ML algorithms for a physically sound but numerically imperfect training set. It may also be possible to augment the computational spectral database with a subset of experimental data and apply ML techniques that can handle a hybrid database, such as transfer learning.

Another open question in computational XANES is the edge alignment of computational spectra with experimental spectra, because current first-principles electronic structure methods have difficulty predicting accurate absolute onset energies. This issue stems from the use of pseudopotentials and/or approximations to the electron self-energy and core-hole final state effects. Therefore, XANES calculations are often analyzed with the relative energy scale or after they are manually aligned with reference experimental spectra. However, the energy shift in the spectral alignment with respect to reference experimental spectra could be system-dependent, which warrants further study.

In order to investigate the impact of the edge alignment on the performance of MLCs, we shift

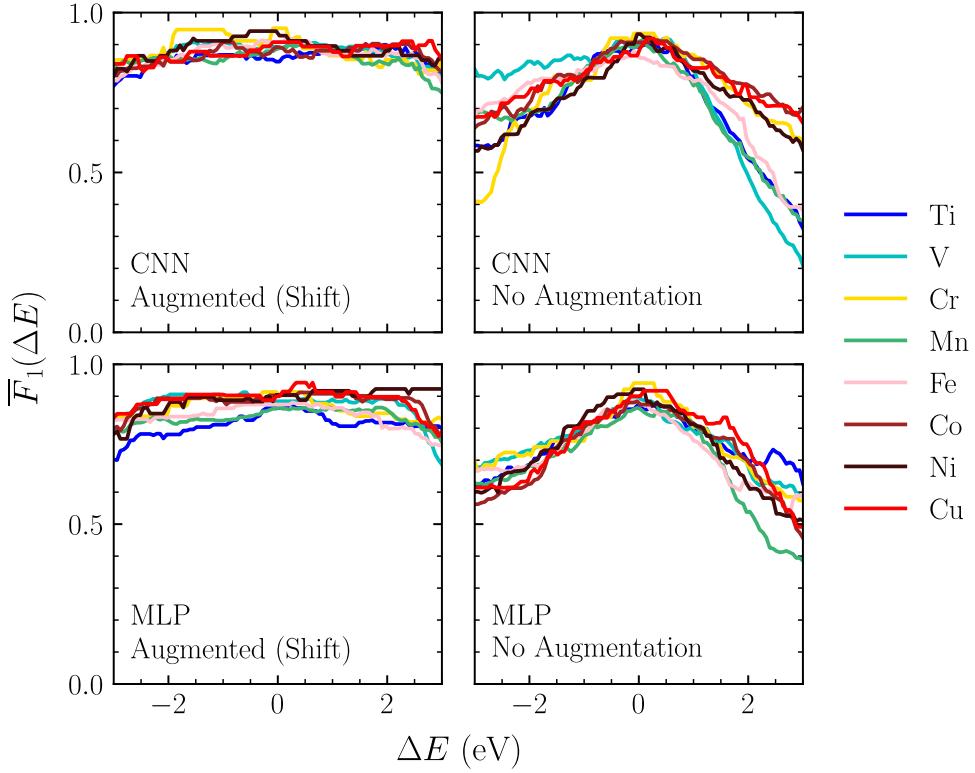


Figure 4.5: \bar{F}_1 score as a function of the amount of energy shift (ΔE) applied to the test set. The two graphs on the left illustrate the results of MLCs trained on augmented data (as presented thus far: ± 1 and 2 eV, generating 5 times the base amount of training data), while those on the right illustrate MLCs trained without augmenting the training set.

the test set by up to ± 3 eV to study the transferability of MLCs against the shifted data. We note that a shift of 3 eV in energy is quite significant, as the energy range of the pre-edge region is about 10 to 15 eV. As shown in Figure 4.5, the MLCs are very robust against the energy shift, as the $\bar{F}_1(\Delta E)$ curves are almost flat. The CNN slightly outperforms MLP with $\bar{F}_1(\text{CNN}) > 0.8$ in most of the range of ΔE for all eight elements. The robustness of the MLCs results from the data augmentation we applied in the training set with energy shifts of ± 1 and ± 2 eV as described in the Section 4.2. If we apply the same test on MLCs developed from the base training set without data augmentation, the accuracy deteriorates quickly after $|\Delta E| > 1$ eV, as shown in Figure 4.5.

Additionally, we have seen that the spectral intensity of theoretical XANES spectra may not match perfectly with experiments. On top that, the intensity of experimental XANES spectra is subject to several uncertainties from sample preparation and various instrumental factors, such

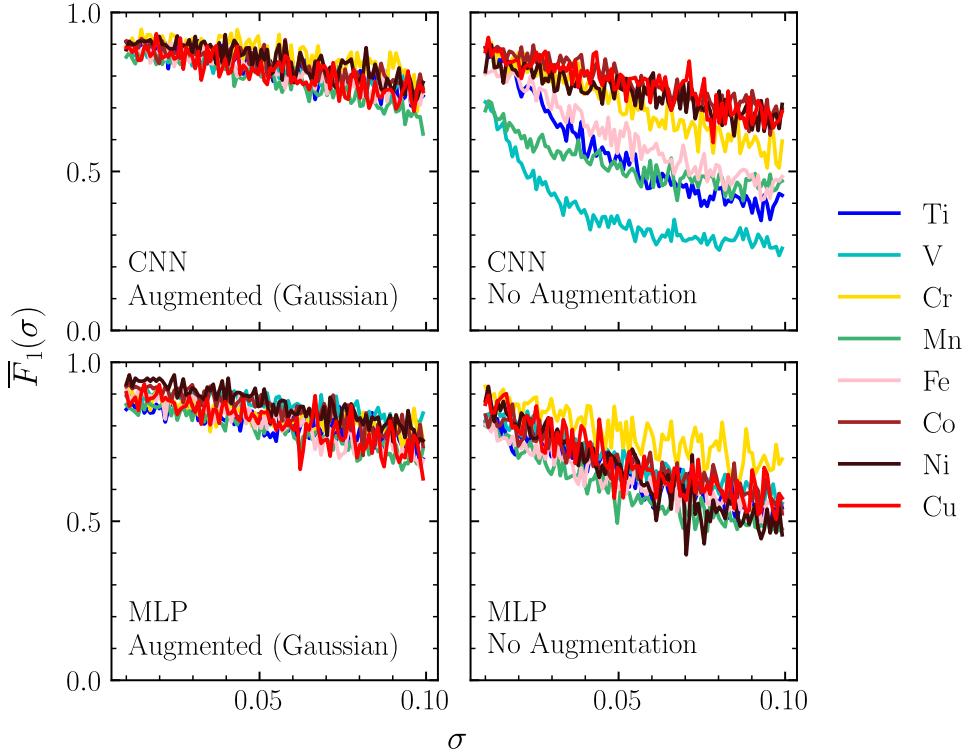


Figure 4.6: \bar{F}_1 score as a function of the standard deviation (σ) used to generate Gaussian random noise, introduced into every point E in the spectra $\mu(E)$. The two graphs on the left illustrate the results of MLCs trained on a training set augmented with Gaussian random noise of $\sigma = 0.03$ (also generating 5 times the base amount of training data), and those on the right illustrate MLCs trained without data augmentation.

as type and mosaic spread of the monochromator crystals, source sizes, slit heights and beam instabilities [178] and the resolution of the apparatus [226]. Therefore, experimental spectra of the same materials that are measured using different samples or collected at different beamline settings may have slightly different intensity profiles. To investigate the impact of the uncertainties in the spectral intensity on the MLCs, we introduced Gaussian random noise with standard deviation σ to the spectral intensity centered around $\mu(E)$ for every E on the energy grid. To isolate the effects of the Gaussian random noise, we test the MLCs trained without augmentation accounting for the energy shift. As shown in Figure 4.6, the overall \bar{F}_1 score decays quickly with the increasing σ in both the CNN and MLP, with V trained using the CNN suffering the most. After we augmented the training set with spectra containing the Gaussian random noise with $\sigma = 0.03$, the \bar{F}_1 score

decays much more slowly with increasing σ .

From the analysis above, one can clearly see that MLCs perform dramatically better when the training set is augmented, which is a sensible result. It is interesting to note that the CNN underperforms relative to the MLP without data augmentation, specifically for early transition metals that exhibit strong pre-edge peaks. For example, the \bar{F}_1 score of the unaugmented CNN for Ti and V drops to ~ 0.45 and 0.25, respectively, at $\sigma = 0.1$. This trend is not entirely counterintuitive for two reasons. First, the CNN as described in Subsection 4.2.3 is not completely shift-invariant, and without proper data augmentation it may not learn how to account for small perturbations. Second, as shown in Table 4.3, the CNN relies more on the pre-edge region than the MLP. Since shifting the location of the pre-edge peak strongly affects the pre-edge spectral features, a sizable drop in performance is to be expected. A similar argument may be made for the effects of Gaussian random noise, which can artificially distort both the shape and location of peaks.

4.4 Conclusion

We propose a new computational framework to perform element-specific classification of local chemical environments from XANES spectra. In addition to the construction of structure and spectral databases and structural labels, a central element of this framework is unraveling the correlation between spectral features and local chemical environments systematically using machine learning classifiers. As proof-of-principle, we applied our method to the computational XANES database of eight 3d transition metal elements generated by the FEFF code and achieved a high average macro F_1 score of 0.86. Our method can reliably capture not only the prominent pre-edge features, but also the less characteristic spectral features beyond the pre-edge region. We showed that features beyond the pre-edge region turn out to be very important to the accuracy of the classification, especially for late transition metal elements. The ability to extract key structural information in the full spectral range makes our machine learning-based method more robust and transferable than empirical fingerprint methods based solely on the pre-edge region. As an impor-

tant starting point, our work will motivate future research on the problem of classification of local chemical environments on experimental measured spectra.

Chapter 5

Machine-learning x-ray absorption spectra to quantitative accuracy

The content presented in this chapter is based closely on the following work (with primarily aesthetic changes made for additional clarity) and its supplemental material:

- Machine-learning x-ray absorption spectra to quantitative accuracy [arXiv]
M. R. Carbone, M. Topsakal, * D. Lu* & S. Yoo,* Phys. Rev. Lett. **124**, 156401 (2020).
Copyright 2020, American Physical Society

* Indicates corresponding authors.

5.1 Introduction

The last decade has witnessed exploding developments in artificial intelligence, specifically deep learning applications, in many areas of our society [227], including image and speech recognition, language translation and drug discovery, just to name a few. In scientific research, deep learning methods allow researchers to establish rigorous, highly non-linear relations in high-dimensional data. This enormous potential has been demonstrated in, e.g., solid state physic and materials science [228, 229], including the prediction of molecular [216, 230] and crystal [231] properties, infrared [232] and optical excitations [233], phase transitions [234] and topological ordering [235] in model systems, *in silico* materials design [236] and force field development [8, 237].

One high-impact area of machine learning (ML) applications is predicting material properties. By leveraging large amounts of labeled data consisting of feature-target pairs, ML models, such as deep neural networks, are trained to map features to targets. The ML parameters are optimized by minimizing an objective loss criterion, and yields a locally optimal interpolating function [238]. Trained ML models can make accurate predictions on unknown materials almost instantaneously, giving this approach a huge advantage in terms of fidelity and efficiency in sampling the vast materials space as compared to experiment and conventional simulation methods. So far, existing ML predictions mostly focus on simple quantities, such as the total energy, fundamental band gap and forces; it remains unclear whether ML models can predict complex quantities, such as spectral functions of real materials, with high accuracy. Establishing such capability is in fact essential to both the physical understanding of fundamental processes and design of new materials. In this study, we demonstrate that ML models can predict x-ray absorption spectra of molecules with quantitative accuracy, capturing key spectral features, such as locations and intensities of prominent peaks.

X-ray absorption spectroscopy (XAS) is a robust, element-specific characterization technique widely used to probe the structural and electronic properties of materials [239]. It measures the intensity loss of incident light through the sample caused by core electron excitations to unoccupied

states [240]. In particular, the x-ray absorption near edge structure (XANES) encodes key information about the local chemical environment (LCE), e.g. the charge state, coordination number and local symmetry, of the absorbing sites [166, 240, 241]. Consequently, XANES is a premier method for studying structural changes, charge transfer, and charge and magnetic ordering in condensed matter physics, chemistry and materials science.

To interpret XANES spectra, two classes of problems need to be addressed. In a *forward* problem, one simulates XANES spectra from given atomic arrangements using electronic structure theory [200, 201, 240, 242–245]. In an *inverse* problem, one infers key LCE characteristics from XANES spectra [192, 193, 246]. While the solution of the forward problem is limited by the accuracy of the theory and computational expense, it is generally more complicated to solve the inverse problem, which often suffers from a lack of information and can be ill-posed [247]. Standard approaches typically rely on either empirical fingerprints from experimental references of known crystal structures or verifying hypothetical models using forward simulation [179, 248].

When using these standard approaches, major challenges arise from material complexity associated with chemical composition (e.g., alloys and doped materials) and structure (e.g., surfaces, interfaces and defects), which makes it impractical to find corresponding reference systems from experiment and incurs a high computational cost of simulating a large number of possible configurations, with hundreds or even thousands of atoms in a single unit cell. Furthermore, emerging high-throughput XANES capabilities [164] poses new challenges for fast, even on-the-fly, solutions of the inverse problem to provide time-resolved materials characteristics for *in situ* and *operando* studies. As a result, a highly accurate, high-throughput XANES simulation method could play a crucial role in tackling both forward and inverse problems, as it provides a practical means to navigate the material space in order to unravel the structure-spectrum relationship. When combined with high-throughput structure sampling methods, ML-based XANES models can be used for the fast screening of relevant structures.

Recently, multiple efforts have been made to incorporate data science tools in x-ray spectroscopy. Exemplary studies include database infrastructure development (e.g. the computational

XANES database in the Materials Project [194–196, 212]), building computational spectral fingerprints [249], screening local structural motifs [250], predicting LCE attributes in nano clusters [192] and crystals [193, 246] from XANES spectra using ML models. However, predicting XANES spectra directly from molecular structures using ML models has, to the best of our knowledge, not yet been attempted.

As a proof-of-concept, we show that a graph-based deep learning architecture, a message passing neural network (MPNN) [251], can predict XANES spectra of molecules from their molecular structures to quantitative accuracy. Our training sets consist of O and N K-edge XANES spectra (simulated using the FEFF9 code [215]) of molecules in the QM9 molecular database [252], which contains $\sim 134k$ small molecules with up to nine heavy atoms (C, N, O and F) each. The structures were optimized using density functional theory with the same functional and numerical convergence criteria. This procedure, together with the atom-restriction of the QM9 database, ensures a consistent level of complexity from which a ML database can be constructed and tested. Although our model is trained on computationally inexpensive FEFF data, it is straightforward to generalize this method to XANES spectra simulated at different levels of theory.

5.2 Dataset construction

The MPNN inputs (feature space) are derived from a subset of molecular structures in the QM9 database, henceforth referred to as the *molecular structure space*, \mathcal{M} . Two separate databases are constructed by choosing molecules containing at least one O (\mathcal{M}_O , $n_O \approx 113k$) or at least one N atom (\mathcal{M}_N , $n_N \approx 81k$) each; note that $\mathcal{M}_O \cap \mathcal{M}_N \neq \emptyset$, as many molecules contain both O and N atoms. The molecular geometry and chemical properties of each molecule are mapped to a graph ($\mathcal{M}_A \rightarrow \mathcal{G}_A$, $A \in \{O, N\}$) by associating atoms with graph nodes and bonds with graph edges. Following Ref. 251, each $g_i \in \mathcal{G}_A$ (i the index of the molecule) consists of an adjacency matrix that completely characterizes the graph connectivity, a list of atom features (absorber, atom type, donor/acceptor status, and hybridization), and a list of bond features (bond type and length).

A new feature, “absorber”, is introduced to distinguish the absorbing sites from the rest of the nodes. Each graph-embedded molecule in \mathcal{G}_A corresponds to a K-edge XANES spectrum in the *spectrum or target space*, $S_A \in \mathbb{R}^{n_A \times 80}$, which is the average of the site-specific spectra of all absorbing atoms, A , in that molecule, spline interpolated onto a grid of 80 discretized points and scaled to a maximum intensity of 1. For each database $\mathcal{D}_A = (\mathcal{G}_A, S_A)$, the data is partitioned into training, validation and testing splits. The latter two contain 500 data points each, with the remainder used for training. The MPNN model is optimized using the mean absolute error (MAE) loss function between the prediction $\hat{\mathbf{y}}_i = \text{MPNN}(g_i)$ and ground truth $\mathbf{y}_i \in S_A$ spectra. During training, the MPNN learns effective atomic properties, encoded in hidden state vectors at every atom, and passes information through bonds via learned messages. The output computed from the hidden state vectors is the XANES spectrum discretized on the energy grid as a length-80 vector. Additional details regarding the graph embedding procedure, general implementation [253–255] and MPNN operation can be found in Ref. 251 and in Appendix 5.A.

5.2.1 Preliminary analysis

Prior to the training, we systematically examine the distribution of the data. Following common chemical intuition, the data are labeled according to the functional group that the absorbing atom belongs to. In order to efficiently deconvolute contributions from different functional groups, we only present results on molecules with a *single* absorbing atom each; this subset is denoted as $\mathcal{D}'_A = (\mathcal{G}'_A, S'_A) \subset \mathcal{D}_A$, and the distribution of common functional groups in \mathcal{D}'_A are elaborated upon in the captions of Figs. 5.1 and 5.2, where the most abundant compounds are ethers and alcohols in \mathcal{D}'_O , and tertiary (III°) and secondary (II°) amines in \mathcal{D}'_N . From the presented averaged spectra, distinct spectral contrast (e.g., number of prominent peaks, peak locations and heights) can be identified between different functional groups. In fact, several trends in the FEFF spectra qualitatively agree with experiment, such as the sharp pre-edge present in ketones (black) but absent in alcohols (red) [256], and the general two-peak feature of primary (I°) amines (blue) [257].

Although XANES is known as a local probe that is sensitive to the LCE of absorbing atoms,

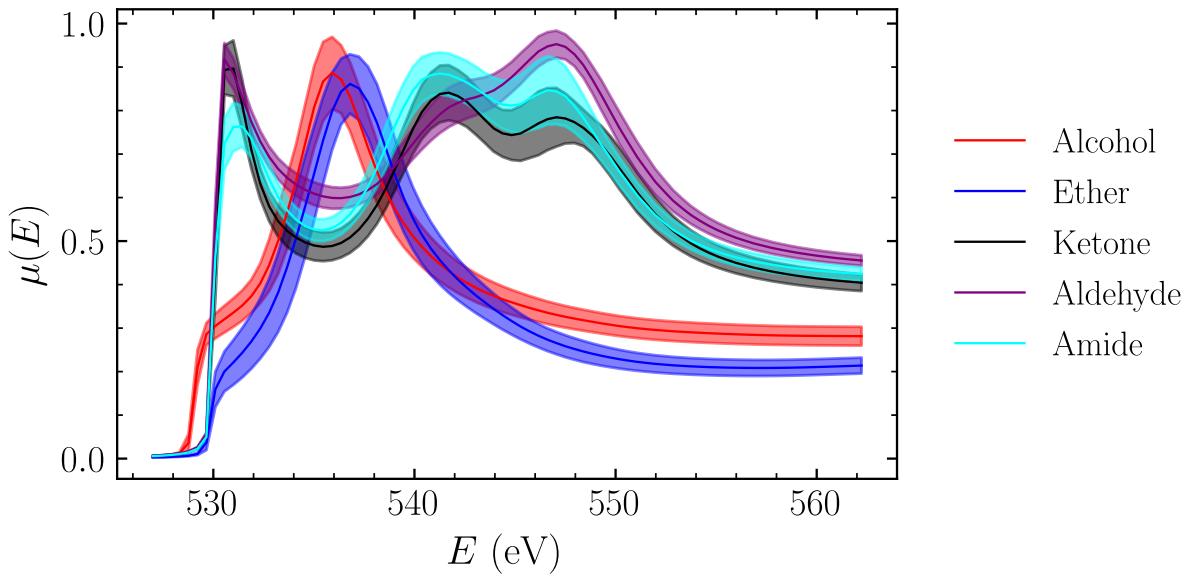


Figure 5.1: The average oxygen spectrum for each considered functional group in S'_O , with the scaled standard deviation $\sigma/2$ presented to highlight regions of high variance. The approximate percent distribution of functional groups throughout the presented dataset is 30.0/42.9/10.5/8.1/8.6 (using the order from the legend) with a total of 47k in S'_O .

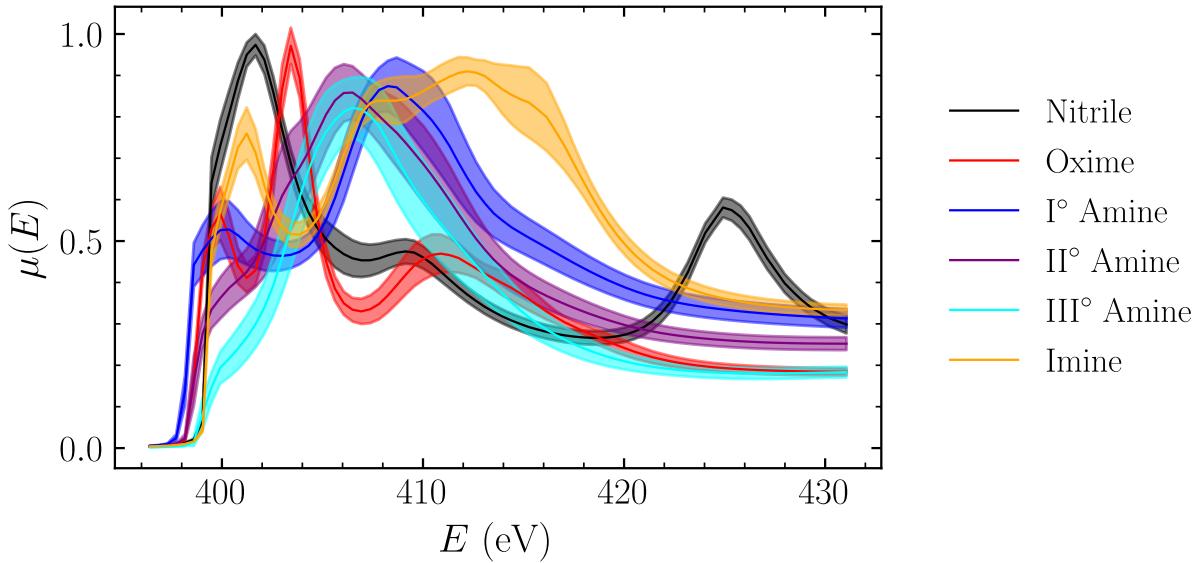


Figure 5.2: The average nitrogen spectrum for each considered functional group in S'_N , with the scaled standard deviation $\sigma/2$ presented to highlight regions of high variance. The approximate percent distribution of functional groups throughout the presented dataset is 19.3/3.3/6.0/32.4/37.3/1.7 (using the order from the legend) with a total of 35k in S'_N .

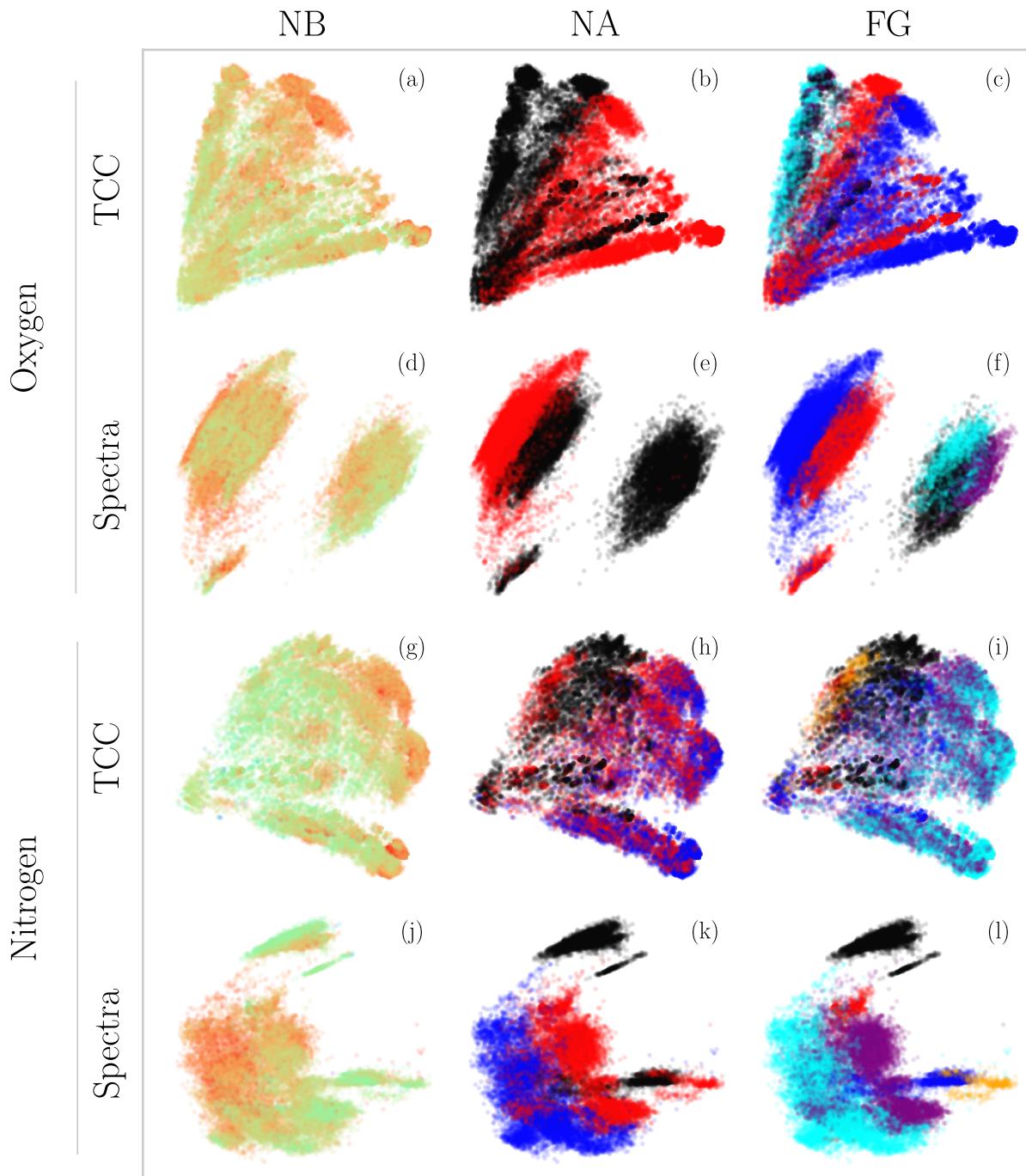


Figure 5.3: PCA plots for both the TCC and spectra proxies for the molecules in \mathcal{D}'_A labeled by NB, NA and FG. The total number of non-hydrogenic bonds (NB, top) range from 1 (violet) to 13 (red). The total number of atoms bonded to the absorbing atom (NA, center) takes on one of three values: 1, 2 or 3 (black, red and blue, respectively). The color legends for the functional group of the absorbing atom (FG, bottom) are the same as in Figs. 5.1 and 5.2.

a systematic study of the degree of such correlation on a large database has not yet been performed. To investigate this structure-spectrum correlation, we perform principal component analysis (PCA) [258] on both the features and targets in \mathcal{D}_A , and visually examine the clustering patterns after the data in \mathcal{D}'_A is labeled by different chemical descriptors. To provide a baseline, we consider the total number of non-hydrogenic bonds in the molecule (NB), which is a generic, global property, supposedly having little relevance to the XANES spectra. Next we consider two LCE attributes: the total number of atoms bonded to the absorbing atom (NA) and the functional group of the absorbing atom (FG). While spectra on a discrete grid can be processed directly, molecular structures, with different number of atoms and connectivity, need to be pre-processed into a common numerical representation before clustering. Thus, the molecular fingerprint of each molecule in \mathcal{M}_A is calculated from its SMILES code using the RDKit library [259]. Then an arbitrarily large subset of 10^4 molecules, $\widetilde{\mathcal{M}}_A \subset \mathcal{M}_A$, is randomly selected to construct a molecular similarity matrix of Tanimoto correlation coefficients (TCCs) [260], $T_A \in [0, 1]^{N_A \times 10^4}$, from the molecular fingerprints such that $T_{A,ij} = \text{TCC}(m_i, m_j)$, where $m_i \in \mathcal{M}_A$ and $m_j \in \widetilde{\mathcal{M}}_A$. $\text{TCC}(m_i, m_i) = 1$ defines perfect similarity. The T_A matrix therefore provides a uniform measure of structural similarity of every molecule in \mathcal{M}_A to each one of the 10^4 references, serving as a memory-efficient proxy to \mathcal{M}_A .

Results of the PCA dimensionality reduction are presented for both data sets and all three descriptor labels (NB, NA and FG) in Fig. 5.3. Specifically, after PCA is performed on unlabeled data, the data are colored in by their respective labels. While some degree of structure is manifest in NB, it is clear that the overall clustering is much inferior to both NA and FG, confirming that NB is largely irrelevant to XANES. On the other hand, both NA and FG exhibit significant clustering, with the latter, as expected, slightly more resolved; while NA can only distinguish up to 2 (3) bonds in the O (N) data sets, FGs reveal more structural details of the LCE, and encode more precise information, such as atom and bond types. For NA and FG, clustering in the TCC-space is more difficult to resolve, as it is only a course-grained description of the molecule, missing detailed information about, e.g., molecular geometry, which will be captured by the MPNN. Despite this,

visual inspection reveals significant structure, such as in Fig. 5.3(c), where alcohols (red), ethers (blue) and amides (cyan) appear well-separated.

Spectra PCA of FG in Figs. 5.3(f) and 5.3(l) can also be directly correlated with the sample spectra in Figs. 5.1 and 5.2. For instance, the shift in the main peak position between ketones/aldehydes/amides (black/purple/cyan) and alcohols/ethers (red/blue) in S'_O reflects the impact of a double versus a single bond on the XANES spectra. As a result, groups of these structurally different compounds are well-separated in the spectra PCA as shown in Fig. 5.3(f); even compounds with moderate spectral contrast, e.g., between alcohols (red) and ethers (blue), are well-separated. Similar trends are observed in S'_N , where, e.g., nitrile groups (black) show a distinct feature around 425 eV, which clearly distinguishes itself from the other FGs, and, likely because of that, one observes a distinct black cluster in Fig. 5.3(l).

5.3 Results and discussion

The PCA suggests that the FG is a key descriptor of XANES. As the MPNN can fully capture the distinction of FGs through node features, edge features and the connectivity matrix, we expect that an MPNN can learn XANES spectra of molecules effectively. Randomly selected testing set results from the trained MPNN for both \mathcal{D}_O and \mathcal{D}_N are presented in Fig. 5.4 and ordered according to MAE, with the best decile at the top and worst decile at the bottom. It is worth noting that MPNN predictions not only reproduce the overall shape of the spectra, but, more importantly, predict peak locations and heights accurately. In the best decile, the MPNN predictions and ground truth spectra are nearly indistinguishable. Even in the worst decile, the main spectral features (e.g. three peaks between 530 and 550 eV in the oxygen K-edge and two peaks between 400 and 410 eV in nitrogen K-edge) are correctly reproduced with satisfactory relative peak heights.

As shown in Table I, the MAE of the prediction is 0.023 (0.024) for the oxygen (nitrogen) test set, which is an order of magnitude smaller than the spectral variation defined by the mean absolute deviation of the oxygen (0.131) and nitrogen (0.123) test sets. To provide an additional

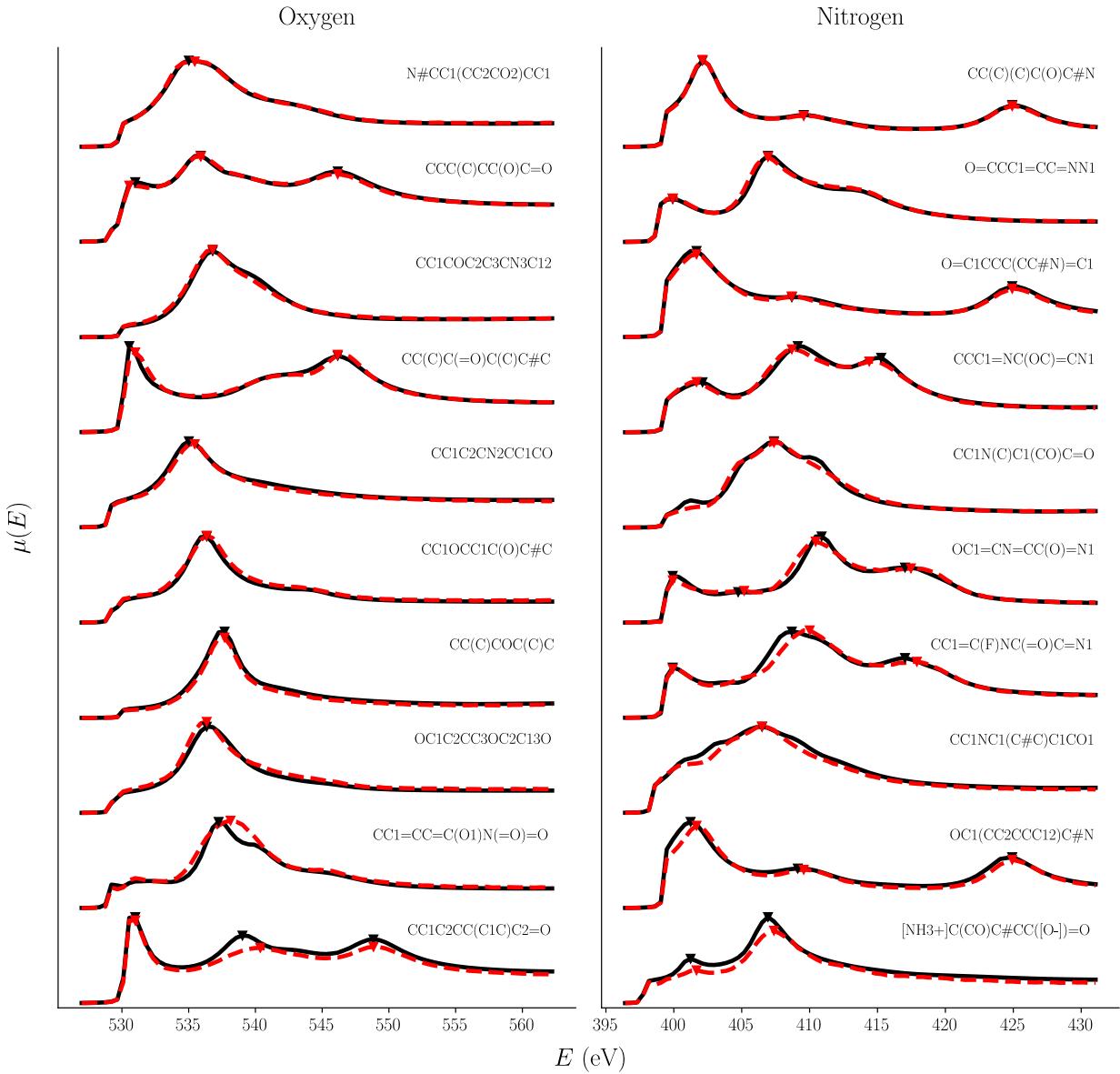


Figure 5.4: Performance metrics for the MPNN evaluated on the \mathcal{D}_A testing sets. Top: waterfall plots of sample spectra (labeled by their SMILES codes) of ground truth (black) and predictions (dashed red), where prominent peaks (see text) are indicated by triangles. One randomly selected sample from every decile is sorted by MAE (first: best; last: worst).

quantification of the model’s accuracy, we select prominent peaks, defined by those with height above half the maximum height of the spectrum and separated by a minimum 12 grid points (≈ 6 eV) in energy. We find that the number of prominent peaks in 95% (90%) of predicted spectra correspond with that of the ground truth for the oxygen (nitrogen) testing set. Peak locations and

heights are predicted with average absolute difference of $\overline{\Delta E} = 0.49$ (0.48) eV and $\overline{\Delta \mu} = 0.045$ (0.041), respectively (see Fig. 5.5, Fig. 5.6 and Table 5.1). The predicted peak heights display a very narrow distribution around $\Delta\mu = 0$, as the total population in the tail region with $\Delta\mu > 0.1$ is only 7%. As shown at the insets, the vast majority ($\sim 90\%$) of the predicted peak locations fall within ± 1 eV of the ground truth, with the coefficient of determination, $R^2 \geq 0.96$. The exceptional accuracy of the MPNN model results on predicting both peak location and intensity underscores its predictive power and its ability to capture essential spectral features.

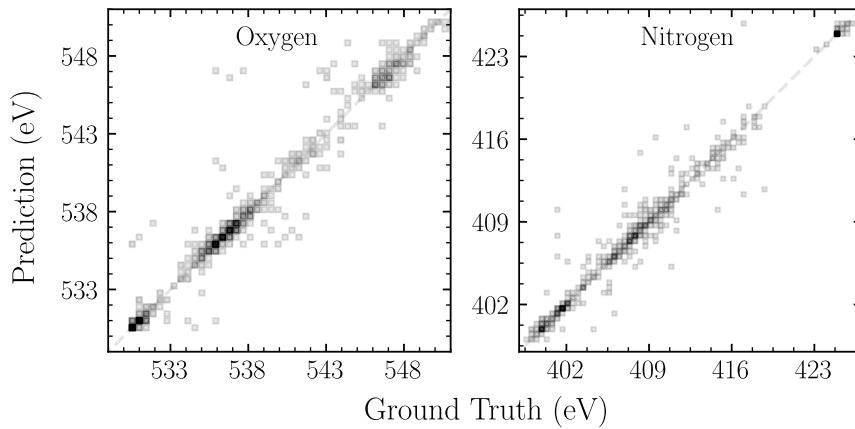


Figure 5.5: Comparison between the prediction and ground truth in peak locations. Contains a total of 824 and 868 peak predictions and ground truths for the oxygen and nitrogen datasets, respectively. The lightest squares represent one data point, with the darkest representing roughly 40. The R^2 values for the presented data are 0.96 and 0.98, respectively.

It is also important to understand the robustness of the network for practical applications; specifically, we examine how distorting or removing certain features impacts the model performance. To do so, we train separate MPNN models using “contaminated” features, where either (1) the bond length is randomized (RBL), or (2) the atom type is randomly chosen, and all other atomic features are removed (RAF). In addition, we investigate the impact of the locality in the MPNN prediction of XANES spectra of molecular systems. By default, the MPNN operates on the graph-embedding of the whole molecule, referred to as the core results. However, the significance of the FG as a sound proxy for the XANES spectra (see Fig. 5.3) suggests that local properties, such as the LCE, play a dominant role. Therefore, spatially truncated graphs are likely to be sufficient to

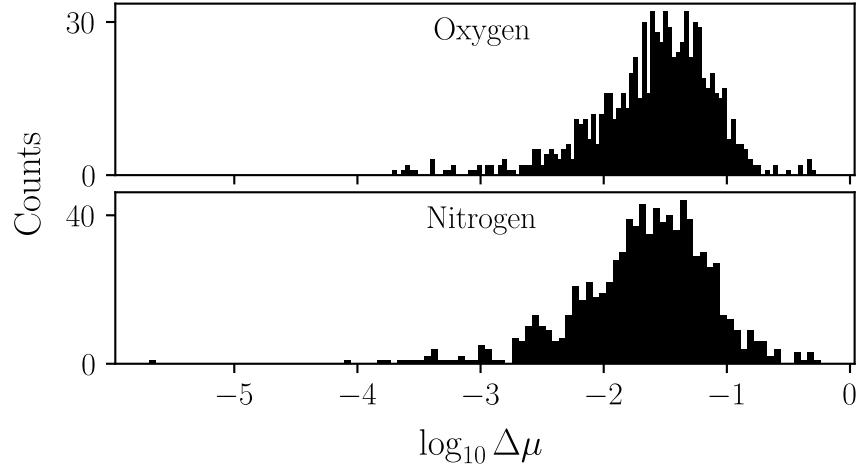


Figure 5.6: Distribution of the absolute error of predicted peak heights, $\Delta\mu$, for both the oxygen and nitrogen datasets.

predict the XANES spectra of molecules accurately. To quantify this effect, we impose different distance cutoffs (d_c) from 2 to 6 Å around the absorbing atoms, and train separate ML models using spatially truncated graphs.

Table 5.1: Performance metrics based on the MAE of the spectra, $\overline{\Delta E}$ and $\overline{\Delta\mu}$.

A	Data	MAE	$\overline{\Delta E}$ (eV)	$\overline{\Delta\mu}$
O	Core	0.023(1)	0.52(4)	0.044(2)
	RBL	0.031(1)	0.55(3)	0.051(2)
	RAF	0.041(2)	0.63(3)	0.068(3)
	$d_c = 4 \text{ \AA}$	0.023(1)	0.45(3)	0.040(2)
	$d_c = 3 \text{ \AA}$	0.025(1)	0.48(3)	0.040(2)
	$d_c = 2 \text{ \AA}$	0.095(4)	0.80(4)	0.179(6)
N	Core	0.024(1)	0.47(3)	0.042(2)
	RBL	0.029(1)	0.57(3)	0.049(2)
	RAF	0.045(2)	0.70(4)	0.084(3)
	$d_c = 4 \text{ \AA}$	0.023(1)	0.43(3)	0.039(2)
	$d_c = 3 \text{ \AA}$	0.027(2)	0.47(3)	0.046(3)
	$d_c = 2 \text{ \AA}$	0.056(4)	0.66(4)	0.099(5)

Independent MPNN models were trained and tested on each database corresponding to either RBL, RAF and different d_c values. As shown in Table 5.1, randomizing the bond length feature does not affect the performance of MPNN, as $\overline{\Delta E}$ and $\overline{\Delta\mu}$ in RBL only worsen slightly. Atomic

features have a larger impact than the bond length, as $\overline{\Delta E}$ and $\overline{\Delta\mu}$ in RAF have a sizable increase from 0.52 (0.47) to 0.63 (0.70) eV and from 0.044 (0.042) to 0.068 (0.084) in \mathcal{D}_O (\mathcal{D}_N). In fact, despite the seemingly large increase, $\overline{\Delta E}$ is still well below 1 eV, i.e., falling within 1-2 grid points, resulting in only a marginal impact on its practical utility. Percentage-wise, the change in $\overline{\Delta\mu}$ is comparable to $\overline{\Delta E}$ for RAF. If we consider relative peak intensity instead of absolute peak intensity as measured by $\Delta\mu$, this difference becomes less significant.

The analysis above leads to a seemingly counter-intuitive conclusion that key XANES features can be obtained with little knowledge about the atomic features and bond length, especially if one considers the importance to know which atoms are the absorption sites. It turns out that this is not entirely surprising, since it has been shown that the distinct chemical information of atoms can be extracted by ML techniques from merely the chemical formula of the compound [261], i.e., specific atomic information can be learned through its environment. In this case, the connectivity matrix likely compensates for a lack of atom-specific information, and supplies enough knowledge about the LCE to make accurate predictions. As for the effect of the locality, we found that the results are statistically indistinguishable from the core results when $d_c \geq 4 \text{ \AA}$, and breaks down at $d_c \approx 2 \text{ \AA}$, indicating that the MPNN architecture requires at least the first two coordination shells to make accurate predictions.

5.4 Concluding remarks

In summary, we show that the functional group carries statistically significant information about the XANES spectra of molecules, and that by using a graph-based deep learning architecture, molecular XANES spectra can be effectively learned and predicted to quantitative accuracy. With proper generalization, this method can be used to provide a general purpose, high-throughput capability for predicting spectral information, which may not be limited to XANES, of a broad range of materials including molecules, crystals and interfaces.

5.A Machine learning details

5.A.1 Graph feature generation

Starting with an .xyz geometry file, we performed the following steps to extract the features described in Table 5.2. To aid the reader in understanding how the feature extraction is performed, we also present an example in Fig. 5.7. Note this can be thought of as performing the mapping from the molecule-space \mathcal{M} to the graph-space \mathcal{G} .

1. Using the geometry SMILES code, an RDKit [259] molecule class was used to determine atom donor and acceptor status, aromaticity, atomic number, and hybridization. An extra feature ("Absorber") was also included, and indexes whether or not the atom is of type A and contributes to computing the target spectra.
2. A NetworkX [254] graph is constructed from the RDKit molecule class. Edges, along with their edge-types (single, double or triple bonded, or part of an aromatic ring) were only initialized between bonded atoms. Bond distances were computed directly from the QM9 geometry files.
3. Adjacency matrices Q_i were initialized for every molecule m_i . Matrix elements $Q_{i,jk} = 1$ if atom j is bonded to atom k , and = 0 otherwise.

5.A.2 Technical details of the Message Passing Neural Network

Here, we summarize the operation of the MPNN used in this work. To aid in visualization, we present a flowchart of the MPNN operation in Fig. 5.8. For further details as well as more examples of different MPNN architectures, we refer the reader to Ref. 251. The general architecture of the MPNN is defined by two phases and three core functions. In the first of the two phases, called the message passing phase (MPP), n_i hidden states, h_v^0 , are initialized for $v = 1, \dots, n_i$, where n_i is the number of atoms in molecule m_i , and v represents an atom in that molecule, for every graph in

Table 5.2: Summary of the input features. The first six denote node (atom) features \mathbf{V}_i , the next two are edge (bond) features \mathbf{E}_i and last is the adjacency matrix Q_i , which encodes all atomic connectivity. *Hydrogen atoms are initialized with a one-hot vector of all zeros. This serves two purposes: both indicating a lack of hybridization and forcing the MPNN to weight hydrogen atoms slightly less compared to the heavy atoms in the molecule. **Each bond type can be a single, double, triple or aromatic. ***For a molecule with z atoms, the adjacency matrix will be $z \times z$.

Feature	Description	Type	Length
Atom Type	H, C, N, O or F	One-hot	5
Absorber	Is the atom type = A	1 or 0	1
Donor	Atom is an e^- donor	1 or 0	1
Acceptor	Atom is an e^- acceptor	1 or 0	1
Aromatic	Atom in aromatic ring	1 or 0	1
Hybridization	sp , sp^2 or sp^3	One-hot*	3
Bond Length	Distance in Å	Float	1
Bond Type	**	One-hot	4
Adjacency	Atomic connectivity	Matrix	***

\mathcal{G}_A . For the remainder of this section, we drop the molecular index and assume features and targets refer to a single molecule. The initial hidden state vector is given by

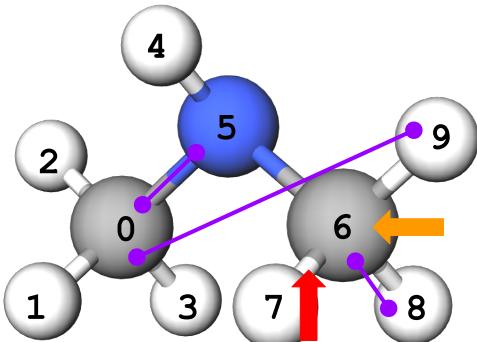
$$h_v^0 = (\mathbf{V}, 0, \dots, 0), \quad \text{LEN}(h_v^0) = d. \quad (5.1)$$

It means that the initial hidden state is simply the list of features for that atom, zero-padded up to some length d , which is a hyperparameter for the MPNN.

The MPP runs for T iterations, and at each iteration, the hidden state vectors h_v^t are reassigned via the following procedure. First, a message

$$m_v^{t+1} = \sum_{w \in N(v)} M^t(h_w^t, e_{vw}) \quad (5.2)$$

is computed by a sum of message functions M^t . The message function incorporates information about all nearest neighbor atoms $N(v)$, and all connecting bonds. Here, we follow the edge network representation for M^t as defined in [251], in which M^t has the following properties: $M^t = M$,



Connectivity Matrix

1	1	1	1	0	1	0	0	0	0
1	1	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0
0	0	0	0	1	1	0	0	0	0
1	0	0	0	1	1	1	0	0	0
0	0	0	0	0	1	1	1	1	1
0	0	0	0	0	0	1	1	0	0
0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	1	0

Atom 0 bonded to 5

Atom 0 not bonded to 9

Atom 6 bonded to 8

Atom Features

0	Not Hydrogen
1	Is Carbon
0	Not Nitrogen
0	Not Oxygen
0	Not Fluorine
0	Not an absorbing atom
0	Not a donor
0	Not an acceptor
0	Not in aromatic ring
0	Not sp hybridized
0	Not sp^2 hybridized
1	Is sp^3 hybridized

Bond Features

1.09	Bond length (pre-scaled)
1	Is single bond
0	Not double bond
0	Not triple bond
0	Not aromatic bond

Figure 5.7: Molecular features of an exemplary molecule as processed by the MPNN. The connectivity matrix (zero-indexed) of the molecule is shown in addition to the features of atom 6 and bond 6-7. Specific connectivity is highlighted with purple lines on the molecule, and the corresponding entries circled in the connectivity matrix.

meaning that the *same* message function is trained regardless of the time step t . M is expressed as

$$M(h_w^t, e_{vw}) = \text{MLP1}(e_{vw})h_w^t, \quad (5.3)$$

where e_{vw} represents the edge (bond) between nodes (atoms) v and w . The message function consists of a neural network trained on bond information only, i.e., for any single atom, it is trained only on the bonds connecting it to its nearest neighbors. The output of MLP1 is a $d \times d$ matrix. In

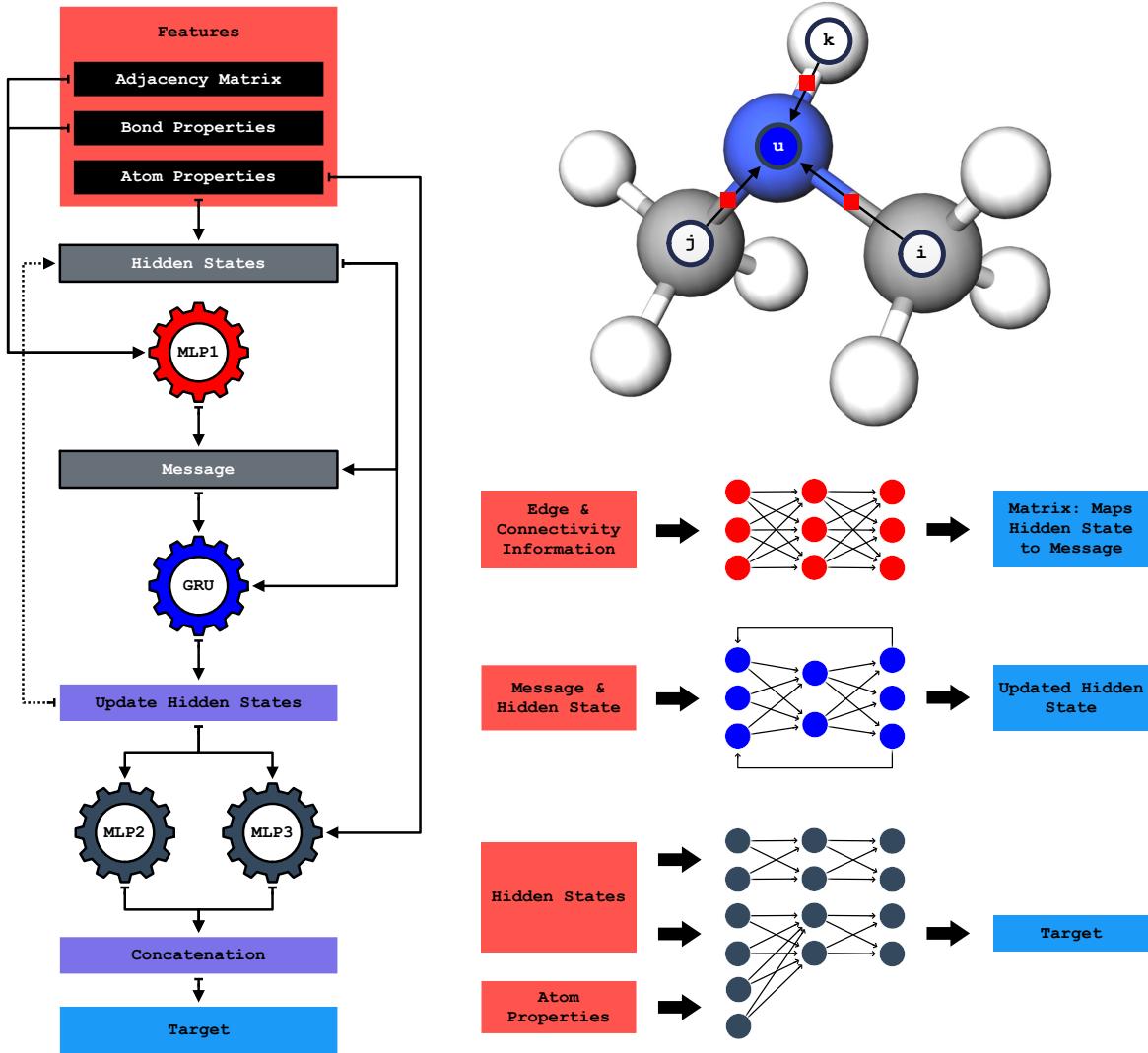


Figure 5.8: Flowchart of the operation of the MPNN. Multi-layer perceptrons (MLPs) are denoted by gears, operations by purple blocks, and internal hidden variables in grey. Initial features are contained in the larger red square, and the target in light blue. Corresponding color-coded cartoons of the networks are shown to the right of the flowchart (note these are not representative of the actual architectures). Finally, above the cartoons is an example graph/molecule during the message passing (black arrows) and update (blue absorbing nitrogen atom) phases. The red squares indicate that edge and nearest neighbor information are used in MLP1 to construct the message.

this way, MLP1 learns a mapping between adjacent atoms h_w and a component of the message that ultimately ends up mutating the hidden state h_v .

For each v , once the message has been computed at time step t , the hidden states are updated

via the update function,

$$h_v^{t+1} = U^t(h_v^t, m_v^{t+1}). \quad (5.4)$$

The update function used in this work is a gated recurrent unit (GRU) introduced by Cho *et al* [255]. It takes as input the current hidden state h_v^t , as well as the constructed message m_v^{t+1} from the message function, and outputs the updated hidden state h_v^{t+1} . Importantly, like the message function, we choose $U^t = U$ such that the GRU is independent of the time step. The independence of both the message and update functions on the time step forces the respective networks to generalize to *any* time step, and can be thought of as a mechanism to guard against over fitting to the training data.

Finally, once the message passing phase has completed, the MPNN enters the readout phase, which takes as input information about the hidden states (generally $h_v^0, h_v^1, \dots, h_v^T$) and predicts the target. The readout function

$$\hat{y} = R(h_v^0, h_v^T) = \sum_v \sigma [\text{MLP3}(h_v^0, h_v^T)] \circ \text{MLP2}(h_v^T) \quad (5.5)$$

inputs the first and last hidden states (where the first hidden states h_v^0 are just the zero-padded atom features) and outputs the length 80 target XAS spectrum. Note that \circ denotes the element-wise product, and σ is the Sigmoid activation function, $\sigma(x) = [1 + e^{-x}]^{-1}$. While MLP2 is trained to make predictions based only on the last output of the MPP, MLP3 effectively learns how to weight those predictions. The sigmoid activation function $\sigma(x)$ has range $(0, 1)$, and thus $\sigma [\text{MLP3}(h_v^0, h_v^T)]$ learns independently from MLP2 how to weight the importance of the outputs, taking as input not only the last output of the MPP, but also the initial atom features.

Training was conducted over 150 epochs using the Adam optimizer, and the Rectified Linear Unit was used for all activation functions. All MLPs contained three hidden layers, each consisting of 92 neurons, and the un-regularized L_1 loss was used as the cost function. No explicit regularization (or dropout) was used during training, but effective regularization was manifest through the use of identical message and update functions regardless of the time step. The specific MPNN

parameters, $T = 3$ and $h = 64$, the maximum time step and length of the hidden state vectors, respectively, were used throughout all calculations. All code was written in PyTorch [253], and training was conducted using a batch size of 512, distributed evenly across parallel GPU architecture.

5.A.3 Learning curves

A hallmark of proper machine learning training is a systematic decrease in the loss metric with respect to training set size until perceptive ability of the model is reached. Here, we show the loss functions for both the O and N data sets. We observe that all losses systematically decrease as the training size is increased.

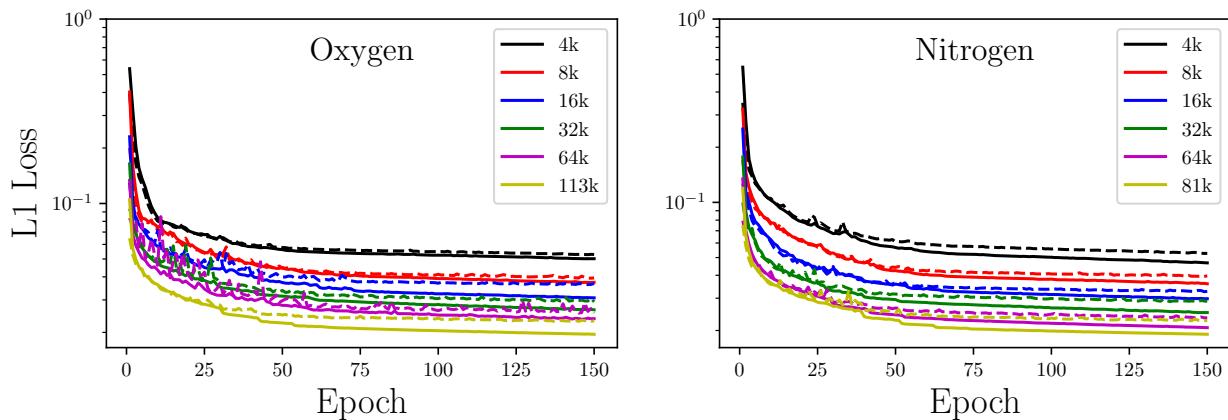


Figure 5.9: L1 loss of the oxygen and nitrogen data sets as a function of training set size (see legend) and epoch. The largest training set size corresponds to the full training set available for that database. Solid lines: training losses; dashed lines: validation losses.

Part III

Monte Carlo dynamics for high-dimensional, continuous phase space models

Chapter 6

Effective trap-like activated dynamics in a continuous landscape

The content presented in this chapter is based on the following published work:

- [Effective trap-like activated dynamics in a continuous landscape. \[arXiv\]](#)
M. R. Carbone,* V. Astuti & M. Baity-Jesi, Phys. Rev. E **101**, 052304 (2020).
Copyright 2020, American Physical Society

* Indicates corresponding authors.

6.1 Introduction

Glasses display a extraordinarily slow dynamics as temperature is decreased [262]. The mean-field picture of this slowing down is elegantly explained as a topological transition in the energy landscape: while at high temperature T the typical configurations are close to the saddle points of the energy landscape, under the *dynamical temperature* T_d the system is confined near the minima of the landscape. Since in mean-field models the energy barriers ΔE diverge in the thermodynamic limit, for $T < T_d$ the system remains confined near the local energy minima and ergodicity is broken [263].

In non-mean-field systems, the barrier heights remain finite, and can be overcome in time scales τ that follow the Arrhenius law, $\tau \propto \exp\left(\frac{\Delta E}{k_B T}\right)$ [264].¹ This barrier-hopping dynamics, which correspond to collective rearrangements of particles, are also called *thermally activated*. The ergodicity breaking induced by the topological transition in the mean field model is hence avoided in many systems of interest, such as 3D glass formers. Activated dynamics must thus be understood in order to characterize the slowing down of glasses.

Given the overwhelming difficulties in the theoretical description of low-dimensional glass formers, a first step towards the understanding of activated dynamics should be done in the mean field approximation. Barrier crossing is in fact also possible in the mean field approximation, provided that the system size N is kept finite [265, 266]. Keeping N finite makes calculations especially hard, so save for some exceptions [267–269], most work on activated dynamics consists of numerical simulations [270–272].

The most popular theoretical framework for the interpretation of activated dynamics is the Trap Model (TM) [273, 274], which consists of a simplified, solvable, version of the energy landscape of glasses, in which the only way to explore the phase space is a purely activated motion between minima in the landscape (called traps), with no notion of distance and with a fixed *threshold* energy that needs to be reached in order to escape a trap.

¹From here on we set the Boltzmann constant $k_B = 1$ and note that unless written with a specified base, all logarithms are the natural logarithm.

The TM yields a wide set non-trivial quantitative and qualitative predictions that have been used to rationalize numerical simulations of low-dimensional glass formers [271, 275, 276], and has been recently shown to serve as an accurate representation of the dynamics of some simple models of glasses in which a threshold energy can be easily identified [268, 269, 277].² A TM description of the dynamics was also shown to be accurate in the Step Model (SM), a model with a single energy minimum, provided that one identifies traps in a dynamical way [279].

Despite these successes, the TM suffers from limitations. On one side, it pictures a phase space motion that is completely unrelated to real-space degrees of freedom, and it is defined on a discrete space of configurations. The first of these two issues was successfully addressed by showing that some problems from number theory can be reformulated as physics problems on a lattice, which behave like the TM [280, 281]. On the other side, the TM paradigm of activated dynamics is probably not suitable for the description of most systems with strong enough correlations [272, 282, 283]. One must therefore try to understand the limits of the applicability of the TM paradigm, and whether it is possible to create a connection between the TM and other systems such as sphere packings.

Much progress along these lines was made in a series of works culminating with the proof that the Random Energy Model (REM), a simple model with glassy behavior, exhibits trap-like dynamics [267–269, 284, 285]. Another consists of studying the influence of phase space connectivity on the dynamics [286, 287]. In our approach, we show that the TM paradigm also applies to a very simple model of a *continuous* N -dimensional landscape, where each dimension represents an independent coordinate in a fictitious space with well-defined metrics. This is done by noting that a TM-like activated behavior can arise due to entropic effects also in the absence of multiple local minima, as was shown for the SM [279].

Our work is thus organized as follows: In Section 6.2 we make simple preliminary observations on how dimensionality induces entropic effects, and in Section 6.3 we introduce the physical model. Furthermore, we study its out-of-equilibrium behavior in Section 6.4 and in Section 6.5

²This connection between trap models had been anticipated decades earlier (see e.g. Refs. 267 and 278).

we show that TM-like dynamics arise. Finally, in Section 6.6 we summarize and discuss our results. We also provide a discussion of the details of our numerical simulations, and mathematical derivations, in the appendices.

6.2 Dimensionality and entropic effects

We study the dynamics of a tracer in an N -dimensional space. This is meant to represent the phase space dynamics of a many-body system in a central potential.³ The potential energy depends only on the distance r of the tracer from the origin,

$$\tilde{E}(r) = \log r, \quad (6.1)$$

and we restrict the phase space to $r \in (0, 1]$. Even though it seems clear that a steepest descent minimization would lead to the origin, any source of noise (temperature, step size, numerical, etc.) would make it almost surely unreachable.

When evolving through a generic equilibrium dynamics algorithm, if N is large, the system would barely feel the presence of the energy funnel, despite its negative divergence. As an example, let us take a Monte Carlo direct sampling dynamics on the N -dimensional unit hypersphere. At every time step a new configuration is proposed with a uniform probability, and a transition towards it is accepted with probability

$$p_{\text{MC}} = \min(1, e^{-\beta \Delta E}), \quad (6.2)$$

where $\beta = T^{-1}$ is the inverse temperature and the energy difference ΔE is negative if the transition decreases r . From any position $\mathbf{x}_0 = (x_{0,1}, \dots, x_{0,N})$ in the landscape, the probability P_\downarrow of moving towards lower energy is set by the relative volume of the sphere of radius $r_0 = |\mathbf{x}_0|$,

$$P_\downarrow(r_0) = \frac{V_N(r_0)}{V_N(1)} = r_0^N, \quad (6.3)$$

³Note that these are N one-dimensional interacting particles, or a single particle in an N -dimensional space.

where $V_N(r)$ is the volume of the N -dimensional sphere of radius r . As also depicted in Fig. 6.1, the probability of proposing a move that decreases the energy goes down exponentially with the dimension of the system N , and this decrease is more severe the closer \mathbf{x}_0 is to the origin. In Fig. 6.1 we can remark that, already for dimensions as small as $N = 10$, P_\downarrow is smaller than single floating point accuracy.

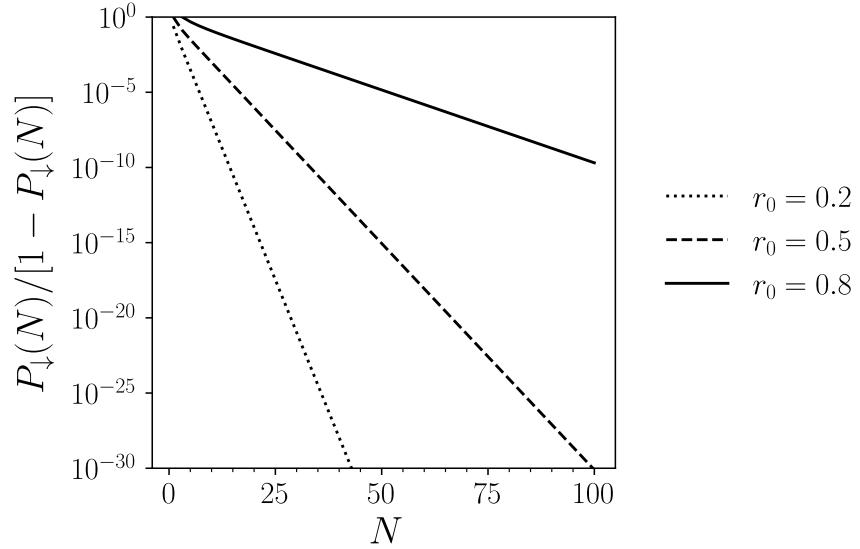


Figure 6.1: Ratio of the probability of the tracer moving towards the well singularity, P_\downarrow , and the tracer moving either away from it or staying in the same location, $1 - P_\downarrow$. Each curve represents a different starting radius, r_0 .

From Eq. (6.3) we can extract a characteristic time scale for decreasing the energy

$$\tau_0 \sim 1/P_\downarrow = e^{-N \log r_0}. \quad (6.4)$$

This kind of slowing down generated by the rarefaction of directions that decrease the energy is called *entropic aging* [288, 289].

At non-zero temperature, for large-enough N the dynamics will always be pushed outwards, because the probability of accepting moves will stay finite (i.e. independent from N), while the probability of proposing moves that decrease the energy is dramatically suppressed. In other words, unless the potential energy also scales with N , at any fixed r_0 there will always exist an N over

which the attraction that $\tilde{E}(r)$ exerts on the tracer becomes irrelevant. One therefore needs to counterbalance by either taking temperatures of order $1/N$, or giving the potential the right scaling N , in order to account for the energetic push towards the center of the N -dimensional sphere with the entropic effects induced by large dimensionality that push the system towards the boundary. This is standard practice in statistical mechanics problems.

6.3 Physical model

If we recall that our model describes the phase space of an N -body system, we should not be surprised that plausible dynamics require the potential to be rescaled with N , as the energy should be an extensive quantity. Thus, in the rest of this work we will use the properly re-scaled potential

$$E(r) = \frac{N}{\beta_c} \log r, \quad (6.5)$$

where $\beta_c > 0$ is a parameter that sets the right physical dimensions, and $r \in (0, 1]$. We normalize the volume so that the portions of phase space map exactly onto their probability, in agreement with a microcanonical interpretation of the space of configurations. The normalized radial volume element is then $d\tilde{V} = \frac{dV}{V_N(1)} = \frac{\Omega^N N r^{N-1} dr}{V_N(1)} = N r^{N-1} dr = \beta_c e^{\beta_c E} dE$, where Ω^N is the N -dimensional solid angle. As a consequence, the density of states $g(E) \equiv |\frac{d\tilde{V}}{dE}|$ is equal to

$$g(E) = \beta_c e^{\beta_c E} \Theta(-E), \quad (6.6)$$

where $\Theta(x)$ is the Heaviside step function.

From $g(E)$ we can calculate the partition function of the Canonical Ensemble,

$$Z(\beta; \beta_c) = \int_{-\infty}^0 dE \beta_c e^{(\beta_c - \beta)E} = \frac{\beta_c}{\beta - \beta_c}, \quad (6.7)$$

which is well-defined only for $\beta < \beta_c$. Thus, the equilibrium phase, with average energy $\langle E(\beta) \rangle =$

$\frac{1}{\beta_c - \beta}$ and radius $\langle r(\beta) \rangle = \exp\left(\frac{\beta_c}{N(\beta - \beta_c)}\right)$, only exists for $\beta \leq \beta_c$. For $\beta > \beta_c$, the system is out of equilibrium and the energy will eventually diverge to $-\infty$ as time goes to infinity.

6.4 Off-equilibrium dynamics

Although the equilibrium phase is trivial, the out-of-equilibrium phase of this model displays rich behavior. Since we are out of equilibrium, we need to define the kind of dynamics used: we analyze both non-local (global) and local dynamics, which are generally equivalent for equilibrium simulations. For global dynamics we use Direct Sampling Monte Carlo (DSMC), and for local dynamics, Markov Chain Monte Carlo (MCMC). Details on simulations and measurements are given in Appendix 6.A.

6.4.1 Direct Sampling Monte Carlo

With DSMC, at every time step a point in the hypersphere is proposed as a move for the algorithm. All points of the phase space are proposed with equal probability, and the moves are accepted with probability p_{MC} [Eq. (6.2)].

In the following, we show that according to the value of β there are several regimes in the dynamics, which were already found in the SM. For $\beta > 2\beta_c$ the energy decreases slowly and steadily, at a rate that follows Eq. (6.4); we call this the entropic aging (EA) regime [288, 289]. For intermediate β , even though the energy as a function of time is decreasing on average,⁴ the trajectory intermittently returns to high energies. Following previous literature on the SM, we call this regime *thermally activated* [279, 290, 291]. In this regime, one can identify a finite *threshold* energy (or, equivalently, radius) towards which the dynamics is spontaneously driven.⁵ Even though the dynamics intermittently returns to the threshold, $\overline{E}(t)$ is decreasing because the

⁴We mean an average over the trajectories, not an ensemble average, which is not well defined for $\beta > \beta_c$. We use an overbar, $\overline{(\dots)}$, to denote average over trajectories.

⁵This concept of threshold energy is related to its definition in the TM, as the energy to reach in order to jump barriers and have renewal dynamics [274]. Operative methods for the detection of the threshold energy based on the behavior of p -spin-like models (e.g. those used in Refs. 272, 292) are not good methods in this context.

system spends short times at high energy, and increasingly longer times at lower energy.

Following Refs. [279, 290], we define the threshold radius r_{th} as the radius from which the probability P_{\uparrow} of increasing the energy equals the probability of decreasing it,

$$P_{\uparrow}(r_{\text{th}}; \beta) \equiv P_{\downarrow}(r_{\text{th}}). \quad (6.8)$$

With Monte Carlo dynamics, in general $P_{\uparrow} + P_{\downarrow} < 1$, since there is also a non-zero probability P_0 that the tracer does not move due to the rejection of movement proposals. However, our calculations of r_{th} are static, so P_0 does not influence them. Neglecting P_0 in a dynamic calculation is equivalent to saying that time does not advance when a move is rejected: this does not change the probability of increasing or decreasing the energy once the move gets accepted.

The probability of increasing the energy from a radius r_0 with DSMC is

$$P_{\uparrow}(r_0; \beta) = \frac{\Omega^N}{V_N(1)} \int_{r_0}^1 r^{N-1} dr e^{-N\frac{\beta}{\beta_c}(\log r - \log r_0)} = \frac{r_0^N - r_0^{N\beta/\beta_c}}{\frac{\beta}{\beta_c} - 1}, \quad (6.9)$$

while $P_{\downarrow}(r_0)$ is given in Eq. (6.3). Equating the two, one obtains no real solution for $\beta > 2\beta_c$. For $\beta_c \leq \beta \leq 2\beta_c$ there is a solution growing continuously from 0 at $\beta = 2\beta_c$ to 1 at $\beta = \beta_c$,

$$r_{\text{th}}(\beta) = \left(\frac{2\beta_c - \beta}{\beta_c} \right)^{\frac{\beta_c}{N(\beta - \beta_c)}}. \quad (6.10)$$

For $\beta < \beta_c$ we are in the equilibrium phase: r_{th} is at distance $\sim 1/N$ from the system boundary and from $\langle r \rangle$. Summarizing, in our simple funnel model we have three regimes (Fig. 6.2):

- $\beta < \beta_c$: Equilibrium Phase (EP)
- $\beta_c \leq \beta < 2\beta_c$: Thermal Activation (TA)
- $\beta \geq 2\beta_c$: Entropic Aging (EA)

Note that the threshold is an attractor of the dynamics, in the sense that the tracer's distance from the center tends to shrink when $r > r_{\text{th}}$, and to expand when $r < r_{\text{th}}$ (Fig. 6.3). This can be

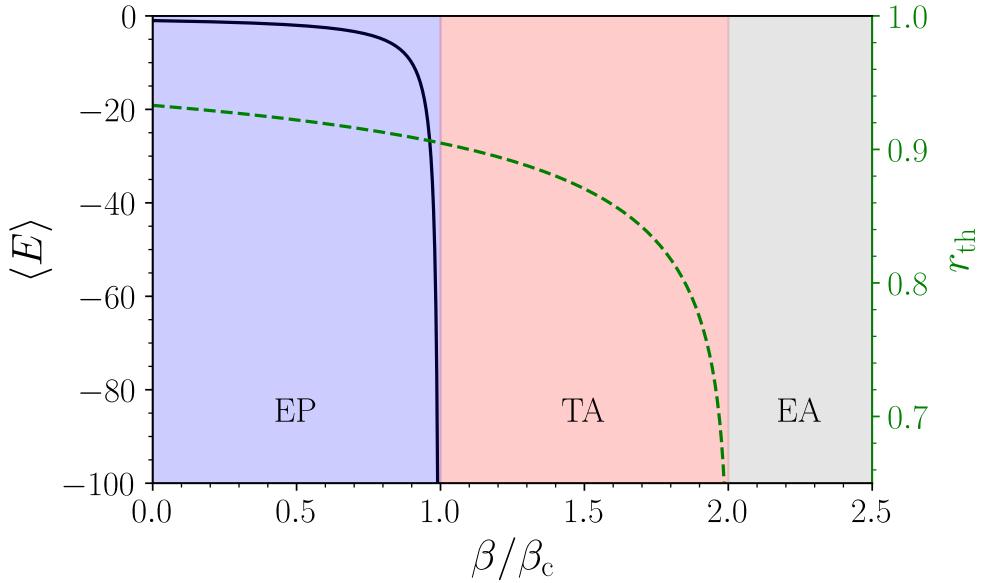


Figure 6.2: Threshold radius (green) and equilibrium value of the energy (black) plotted as a function of β/β_c for $N = 10$. The equilibrium phase (EP) regime, where the system enjoys both a well-defined threshold radius and equilibrium energy is shown in blue ($0 < \beta < \beta_c$). The thermal activation (TA) regime is shown in red. Finally, the entropic aging (EA) regime is shown in grey, where the tracer is relentlessly attracted towards the center of the well ($\beta > 2\beta_c$).

seen clearly from Fig. 6.3, where we show that $P_\uparrow > P_\downarrow \forall r < r_{\text{th}}$, and $P_\uparrow < P_\downarrow \forall r > r_{\text{th}}$. We also show $P_0(r; \beta)$, that goes to 1 as r decreases and can be used as an indicator of the slowness of the dynamics.

In the entropic aging regime the tracer is attracted to the center of the sphere, which it approaches over an infinitely long amount of time. In the thermally activated regime the system fluctuates around $0 < r_{\text{th}} < 1$, but as time passes it becomes increasingly probable that low-energy configurations are reached, where the system will spend very long times before rising to the threshold again. Finally, in the equilibrium phase the system is squeezed on the surface of the hypersphere [$r = 1 - O(1/N)$], and when low-energy configurations are reached the thermal agitation is strong enough to allow for the system to quickly go back to the surface of the hypersphere.

The described dynamical scenario is also encountered in the SM [288, 290, 291]. This is due to the combination of two ingredients: on one side the density of states $g(E)$ in Eq. (6.6) is the same of the SM, and on the other the DSMC algorithm samples directly from $g(E)$. As a consequence,

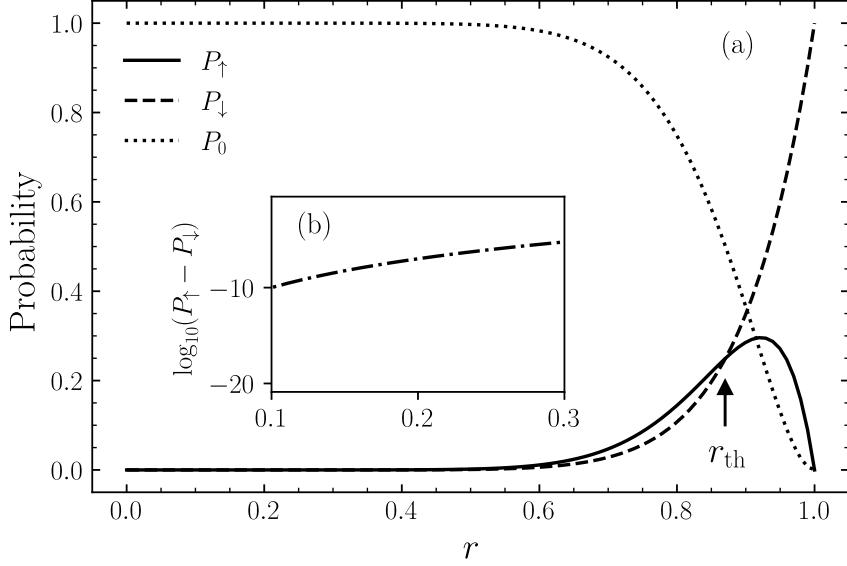


Figure 6.3: Main set: (a) Probability of the tracer moving up ($P_{\uparrow}(r; \beta)$), down ($P_{\downarrow}(r)$) or not moving ($P_0(r; \beta)$) for $N = 10$ and $\beta = 1.5\beta_c$ (TA regime). The threshold radius is identified by $P_{\uparrow} = P_{\downarrow}$. For $r > r_{\text{th}}$, we have $P_{\downarrow} > P_{\uparrow}$. For $r < r_{\text{th}}$, we have $P_{\downarrow} < P_{\uparrow}$ (see the inset (b) for a closeup of the difference between the two). The slowdown of the dynamics with small r is encoded in P_0 going to 1.

the dynamical succession of energies in our funnel is statistically the same as that of the SM, so we note that several results from the SM are realized in DSMC dynamics. For example, in the SM, the distribution of persistence times in a configuration, $\psi_C(\tau_C)$ (distribution of times spent in a configuration), and in a basin, $\psi_B(\tau_B)$ (distribution of times spent under the threshold) both decay as [279, 291]

$$\psi(\tau) \sim 1/\tau^{1+\mu}, \quad (6.11)$$

with $\mu = 2 - \beta/\beta_c$, which is what we find in our high-dimensional funnel (Fig. 6.4, left bottom), and the energy decays logarithmically (Fig. 6.4, left top).

Note, however, that finite-size effects are now different, because in network models such as TM, SM and REM, the lowest available energy depends on N . Therefore, for any finite N , $\bar{E}(t)$ eventually saturates, whereas in our model the energy decreases to $-\infty$ at any system size, so the dynamical phase diagram [285] can differ.

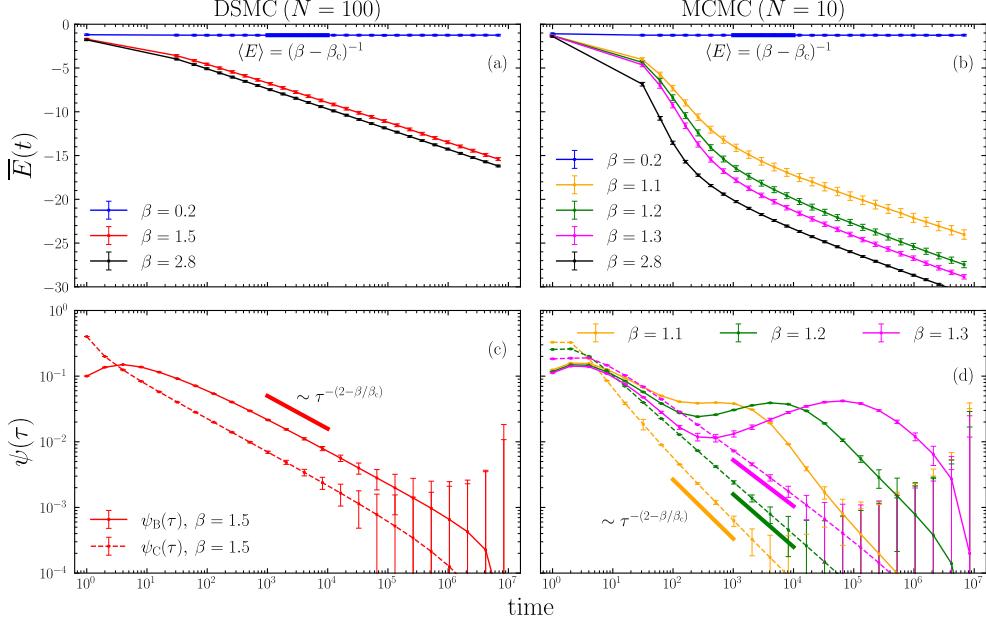


Figure 6.4: Top: We show the energy of three trajectories as a function of the Monte Carlo timestep t , for $N = 100$ and $\delta = 0.1$, each in a different dynamical regime. On the left we show DSMC dynamics (a), and on the right we show MCMC dynamics (b). The EP curve ($\beta = 0.2$) converges quickly to its equilibrium value $\langle E(\beta) \rangle = -1.25$ (horizontal blue dashed line), while TA and EA curves ($\beta = 1.5$ and 2.8 , respectively) diverge logarithmically to $-\infty$. Bottom: Configuration and basin trapping time distributions, $\psi_C(\tau_C)$ (dashed) and $\psi_B(\tau_B)$ (solid), for DSMC (c) and MCMC (d) dynamics. Results are averaged over 200 identical simulations, each with 500 tracers.

6.4.2 Markov Chain Monte Carlo

Especially given that we are introducing spatial effects in the dynamics of network models, it is perhaps more interesting to study also local dynamics. We analyze a Markov Chain sampling of our N -dimensional funnel. From every point \mathbf{x}_t in the hypersphere, a new point \mathbf{x}_{t+1} is proposed by making a Gaussian shift

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \Delta, \quad (6.12)$$

where $\Delta \sim \mathcal{N}_N(0, \delta^2)$ is an N -dimensional Gaussian random variable with variance δ^2 (meaning a diagonal covariance matrix with δ^2 in each entry) centered at the origin and randomly sampled at every time step. The move is accepted with the Monte Carlo rate in Eq. (6.2). The initial configuration is uniformly drawn from the radius 1 hypersphere.

Also in this case, we find the same three dynamical regimes that we found for the DSMC. Since the equilibrium properties are independent of the type of dynamics (provided that it obeys detailed balance), we still have an EP for $\beta < \beta_c$. For higher β , there is a TA regime defined by the presence of a positive threshold radius. In Appendix 6.B we show that the TA regime terminates at $\beta = 2$, where the dynamics is not intermittent anymore, and one reaches an EA regime, where the energy decreases steadily.

We can derive the threshold radius that is valid for large N by imposing that the probability of increasing and decreasing the energy is equal,

$$P_{\downarrow}(r_{\text{th}}; \delta) \equiv P_{\uparrow}(r_{\text{th}}; \beta, \delta), \quad (6.13)$$

where now Eq. (6.13) also accounts for the size of the MCMC step, δ , in the upwards and downwards probabilities, which can be formally written as

$$P_{\downarrow}(x_0; \delta) \propto \int_{0 < |\mathbf{x}| < x_0} d\mathbf{x} e^{-\frac{(\mathbf{x}-\mathbf{x}_0)^2}{2\delta^2}}, \quad (6.14)$$

and

$$P_{\uparrow}(x_0; \beta, \delta) \propto \int_{x_0 < |\mathbf{x}| < 1} d\mathbf{x} e^{-\frac{(\mathbf{x}-\mathbf{x}_0)^2}{2\delta^2}} e^{-\beta \Delta E(\mathbf{x}; \mathbf{x}_0)}. \quad (6.15)$$

In Appendix 6.B we solve Eq. (6.13), see that r_{th} is independent from δ up to $o(1/N)$ terms, and derive an explicit form approximately identical to the DSMC one:

$$r_{\text{th}}(\beta) = \left(\frac{2\beta_c - \beta}{\beta_c} \right)^{\frac{\beta_c}{N(\beta - \beta_c)}} + o\left(\frac{1}{N}\right), \quad (6.16)$$

In the large- N limit the threshold radius for MCMC and DSMC coincides (see also Appendix 6.B and Fig. 6.5). Further, both $\bar{E}(t)$ and $\psi(\tau)$ behave the same at long times (see Fig. 6.4, right).

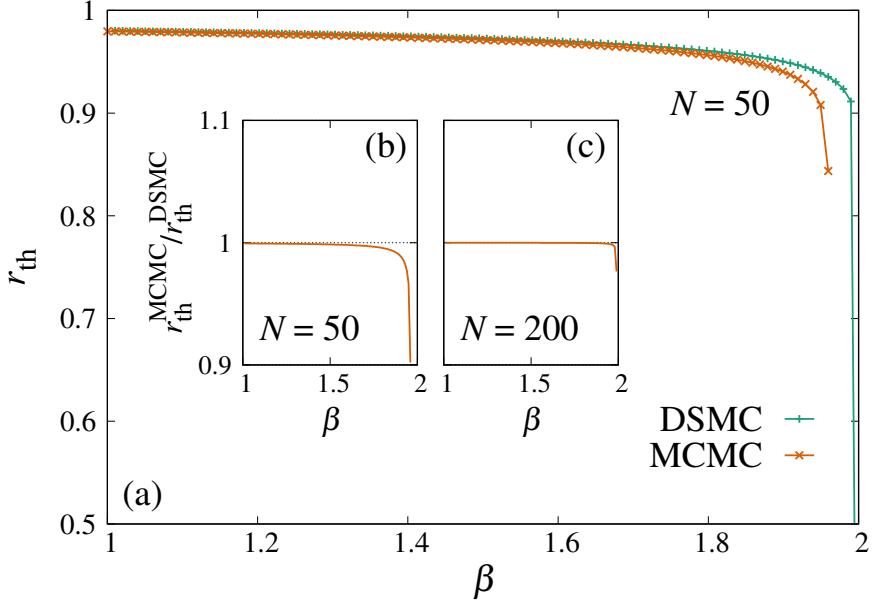


Figure 6.5: Main set: (a) Threshold radius as a function of β in a system of size $N = 50$, for DSMC [Eq. (6.10)] and MCMC [Eq. (6.16)] dynamics. The y axis is truncated at 0.5 to improve the figure's clarity. Points are shown in the curves in order to make overlapping curves visible. Insets: Ratio between the MCMC and the DSMC threshold radii for (b) $N = 50$ and (c) $N = 200$.

6.5 Trap-like behavior

In Ref. 279, it was shown that the SM displays an activated aging dynamics that is effectively like that of the TM. In order to do so, we studied the time evolution of the energy, and defined energy basins dynamically, as the periods of time that the system remains at $E < E_{\text{th}}$. The distributions of trapping times are shown in Fig. 6.4–bottom.

We can use the exponent μ [Eq. (6.11)] to show that the funnel model has TM dynamics, as was done in Ref. 279 for the SM, by studying the aging function $\Pi_B(t_w, t_w + t)$, defined as the probability of not changing basin between the times t_w and $t_w + t$.⁶ To define the basins' threshold we used Eq. (6.10) for both DSMC and MCMC.

In the TM, the aging function has a well-defined limiting value which depends only on the

⁶Details on definition and computation of $\psi(\tau)$ and $\Pi_B(t_w, t_w + t)$ are given in Appendix 6.A.

exponent μ [Eq. (6.11)] and on the ratio $w = t/t_w$ [274, 293],

$$H_\mu(w) = \frac{\sin(\pi\mu)}{\pi} \int_w^\infty \frac{du}{(1+u)u^\mu}. \quad (6.17)$$

In Fig. 6.6-top we show that the aging function in the TA regime converges clearly to the TM prediction in DSMC dynamics. The same is valid for MCMC dynamics which, being local, is much slower than DSMC, so our simulations are restricted to lower N and β .

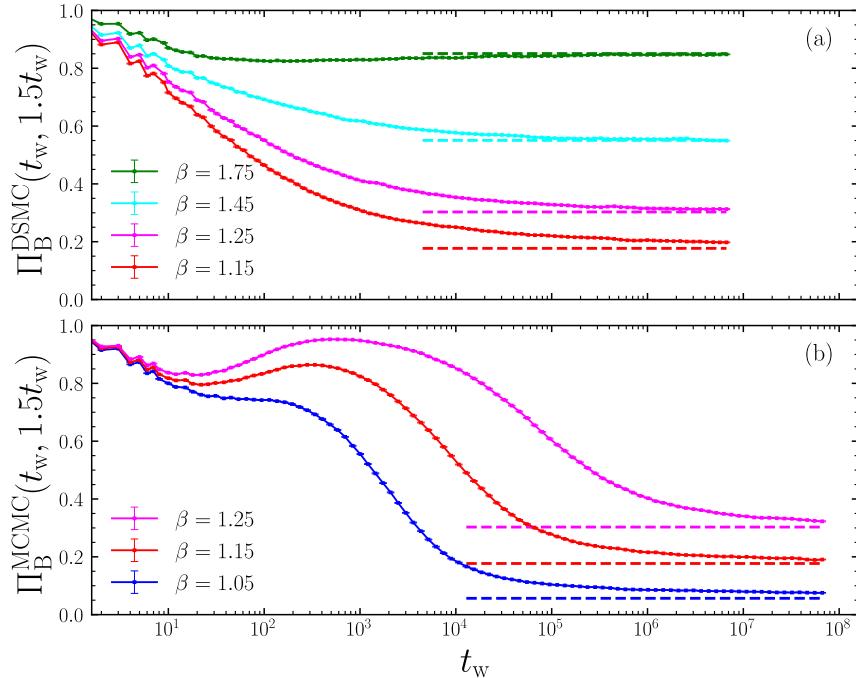


Figure 6.6: Top: (a) Aging functions $\Pi_B(t_w, 1.5t_w)$ for DSMC dynamics, with $N = 100$, $\beta = 1.15, 1.25, 1.45, 1.75$. Bottom: (b) Aging functions $\Pi_B(t_w, 1.5t_w)$ for MCMC dynamics for $N = 10$ and $\beta = 1.05, 1.15, 1.25$. The dashed horizontal lines correspond to the trap value $H_{2-\beta}(0.5)$. Results are averaged over 200 identical simulations, each with 500 tracers.

The strong slowing down in the MCMC dynamics can be appreciated from Fig. 6.7. As one can expect, the global update dynamics has no finite-size effects, whereas the local dynamics is increasingly slower as the system size increases.⁷ This slow down is exponentially large with the

⁷This is true if δ is independent from the system size. If δ increases with the system size (which is not the case in typical local algorithms of many-body systems), the dependence on N can be suppressed. To obtain that the ratio of the volume of configurations accessible in one step, divided by the total volume, should stay constant. We can use Eq. (6.3) to obtain $\delta \sim e^{-\frac{\text{const.}}{N}}$.

system size (Fig. 6.7, inset), which suggests that our funnel model is correctly capturing the nature of activated processes.

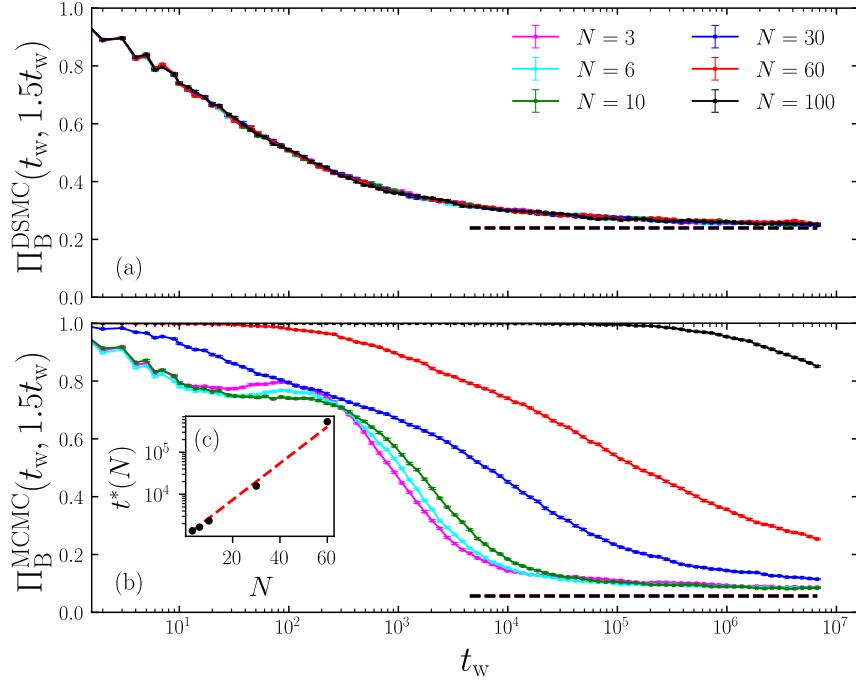


Figure 6.7: Top: (a) Aging functions $\Pi_B(t_w, 1.5t_w)$ for DSMC dynamics with $\beta = 1.2$ and $N = 3, 6, 10, 30, 60, 100$. Bottom: (b) Aging functions $\Pi_B(t_w, 1.5t_w)$ for MCMC dynamics with $\beta = 1.05$ and the same values for N . The dashed horizontal lines correspond to the trap values $H_{2-\beta}(0.5)$. Results are averaged over 100 identical simulations, each with 200 tracers. Inset: (c) The time at which the average $\Pi_B^{\text{MCMC}}(t_w, 1.5t_w)$ function crosses 0.4 as a function of N (obtained through linear interpolation). Note the semi-log scale on the y-axis implies exponential scaling in N .

6.6 Discussion & conclusions

We investigated the out-of-equilibrium dynamics of a tracer in an N -dimensional funnel landscape. The dynamics is dominated by the competition between an energetic pull towards the center, and an entropic push outwards due to the dimension of the space, which turns out to be equivalent to increasing the thermal noise by a factor N . As a consequence, the energetic contribution needs to scale with N (or the temperature needs to be rescaled by $1/N$) for it to be relevant.

The properly rescaled model has a high-temperature equilibrium phase, and two low-temperature out-of-equilibrium regimes: a thermally activated regime in which the tracer intermittently comes up to the surface even though, on average, its energy decreases indefinitely, and an entropic aging regime in which this intermittency disappears. This same phenomenology is found in the Step Model (SM), a network model with random energies and no notion of distance [288, 289, 291], which in our model is recovered in the limit of maximally delocalized and uncorrelated updates.

We find that, besides the system size dependence, this non-equilibrium behavior is independent of the chosen dynamics. We examined a global update method, Direct Sampling Monte Carlo (DSMC), and a local update, Markov Chain Monte Carlo (MCMC), which turned out to be equivalent, providing an example of the equivalence of equilibrium algorithms in out-of-equilibrium contexts. Extensions to other kinds of physical dynamics would arguably give the same results.

Unlike the SM, quenched disorder is not needed in order to have glassy TM-like activated dynamics. This is understood by comparing the SM with our funnel model with DSMC dynamics: the randomness due to disorder in the SM can be incorporated into that due to thermal fluctuations, giving the same kind of long-time activated dynamics.

The funnel model can be seen as an extension of the SM to a continuous landscape, where a notion of space, distance and dimension are now well-defined. This makes it viable to extend the TM paradigm (or the suitable modifications of it) to more realistic models, such as structural glasses.⁸

A critique suffered by models such as TM and SM is that the excessive simplicity of their phase space makes it impossible to use them to describe any Hamiltonian system in realistic terms, which is solved by our funnel model. An alternative approach to associate the TM to models with microscopic degrees of freedom was proposed in Ref. 280, by reformulating the number

⁸One may argue that most models of structural glasses enjoy translation invariance, which the funnel model lacks. However, even though we do not claim that this funnel model can faithfully represent a glass, translation invariance, or lack thereof, does not present with cause for concern. Indeed, translation invariance is not necessarily a feature of glasses, and breaking it often leads to an increased glassiness. For example, randomly pinning particles in a supercooled liquid breaks translation invariance, but exacerbates the glassy behavior [294, 295]. Both in structural and in spin glasses, symmetries such as translation and rotation give rise to Debye modes in the density of states, which are not a signature of glassiness. To the contrary, the glassy low-frequency modes emerge when those symmetries are explicitly broken [296, 297].

partitioning problem (NPP) as a Mattis spin glass, and focusing on single spin flip dynamics. This TM-like behavior is due to the presence of few low-energy configurations that are one spin flip away from typical energies. By changing the dynamics to multiple spin flips, the TM behavior disappears, and the dynamics mimics the SM when all the spins are updated simultaneously [281]. This is conceptually different from what we find for two main reasons:

- In Ref. 280, the TM dynamics is due to an equivalence at a level of the *landscape*, which is composed by rare, point-like, regions with a very low energy in an environment where the system can otherwise move freely. Instead, in the funnel model (and in the SM), the origin of the TM-like behavior is purely due to entropy, and it is effective in the sense that the trapping time distributions are different from those of the TM, but their relationship with the aging functions is the same as that of the TM. We thus have two very different kinds of TM-like dynamics, one which is energy-driven, and another which is entropy-driven, and they should be treated differently. Defining basin hopping through the dynamics has also been done in experiments of glass-formers by looking at particle movement [298]. However, the possible entropic origin of the observed activated dynamics is either dismissed, in favor of energy-based arguments, or it is incorporated into kinetic constraint arguments [299]. Both kinds of barriers (energetic and entropic) induce logarithmically slow dynamics [271], and in realistic systems there is likely competition (or synergy) between the two kinds of effects, due to the presence of a collection of deep wide minima [282].
- We find a TM-like behavior for both local and global dynamics. The type of activation analyzed in the NPP is energy-driven [280], and in the limit of global updates the model resembles the SM [281], which exhibits entropic TM-like activation [279]. Our analysis suggests, therefore, that entropy-driven activation is more robust to changes in the dynamics, and that the NPP is likely to exhibit both energetic and entropic trap-like behaviors at the same time. We highlight these simple models and their activated behaviors in relation to their dynamics in Table 6.1.

Table 6.1: For each model and dynamics, we specify whether we have energy- or entropy-driven activation. Trap model (TM) and Step model (SM) live in a fully connected phase space, so they cannot have local dynamics. The funnel model studied in this paper consists of a single well, so it cannot have energy-driven trap-like behavior. It does have entropy-driven activation for both local and global dynamics. The number partitioning problem (NPP) and the random energy model (REM) have energy-driven trap-like activation when using single spin flip dynamics. With global dynamics, the NPP has (as well as the exponential version of the REM) entropy-driven activation. This makes the NPP and REM models good candidates for having both kinds of trap-like behaviors simultaneously (i.e. with the same dynamics), since we can expect the same phenomenology of the funnel model. ^aThe trap and step models cannot have local dynamics. ^bThe subject of this work. ^cModels like NPP and REM could have entropically activated trap-like behavior also with local dynamics. ^dAn exponential REM with global dynamics is a SM. The usual, Gaussian, REM has not been examined.

Model	Global Energy	Global Entropy	Local Energy	Local Entropy
TM	Yes	No	No ^a	No ^a
SM	No	Yes	No ^a	No ^a
Funnel	No	Yes^b	No	Yes^b
NPP	No	Yes	Yes	? ^c
REM	No	Yes ^d	Yes	? ^c

Our funnel model introduces Euclidean space in the SM, and shows an alternative way of introducing locality, providing the possibility of local moves but without the multiplicity of local minima that characterize energy-driven trap landscapes. As it also happens for the NPP [281], the SM is only the limit for maximally global dynamics of our model, but now locality is different than in the NPP, since it involves a Euclidean metric, and therefore, finite-size effects are also different. In fact, in the SM and the other aforementioned lattice models, N determines the lowest reachable energy, whereas here N is related to the amplitude of the noise. This implies that SM and funnel model are the same model only in the $N \rightarrow \infty$ limit using DSMC dynamics. Another way to see this is that, even for finite N , in the funnel model, the ground state is at $-\infty$, and it is almost impossible for any algorithm with any amount of noise to reach the center of the hypersphere. A direct consequence of this is that, unlike TM, SM, REM and NPP, a finite-size funnel model at $\beta > \beta_c$ will never reach equilibrium. Furthermore, the introduction of local dynamics and the connection to particle systems uncover that entropic TM-like aging also displays an exponential

slowing down with the system size, which is a fundamental trait of activation which needed to be observed.

The interaction potential to which these particles are subject, although exotic, can be found in several situations, such as in bosonic systems [300], when converting into extensive problems with an exponential scaling in the system size [280], or by reformulating number theory problems in terms of a cost function [280]. An interesting development of our work would be the exploration of entropy-driven activation in r^α potentials (for example the case $\alpha = -1$ would represent N particles in a Coulomb potential). Such further directions will be the subject of future work.

6.A Simulation details

In this Appendix we explain our simulation procedures, and explain our parallel code, that we provide for open access at <https://github.com/x94carbone/hdwell>.

For each choice of the parameters and dynamics we simulate M batches (usually 100 to 200) of trajectories of $\tau_{\max} = 10^7$ to 10^8 time steps. Each of the m tracers per batch (usually 200 to 500) is computed in parallel, and are used to compute distribution averages effectively. From each batch we obtain the whole curves $\psi(\tau)$ and $\Pi(t_w, t_w + t)$ which can then be averaged among batches. The specific details of how these values are calculated are summarized in this Appendix.

6.A.1 Monte Carlo procedure

The details of the simulation algorithms are henceforth summarized:

1. Initialize on the N -dimensional sphere with a uniform distribution (so for large N the tracer is initially at $r \simeq 1$). To sample uniformly the hypersphere we use the algorithm in Ref. 301.
2. For each tracer at time t , make a proposal move towards \mathbf{x}_{t+1}^* as follows:
 - (a) If DSMC: \mathbf{x}_{t+1}^* uniform in the hypersphere (using Ref. 301).
 - (b) If MCMC: \mathbf{x}_{t+1}^* according to Eq. (6.12).

3. Accept or reject the move with the Metropolis rule Eq. (6.2).
4. If the timestep is designated for recording quantities of interest, save the value of the energy, ψ and Π -values for each tracer.
5. Update the timestep: $t \leftarrow t + 1$.
6. Repeat 2-5 until $t = t_{\max}$.

6.A.2 Calculation of the trapping time distributions

To compute ψ_C , the following procedure is used: A counter is initialized for each tracer in a simulation which keeps track of the number of time steps that tracer remains in a single configuration. Since this is a continuous landscape, the distance from the center of the well is a sufficient proxy for the exact configuration, since we can neglect the probability of changing configuration maintaining exactly the same radius. Therefore, the trapping time is measured as the number of steps during which the system is at the same r . At every time step, the configuration of each tracer is queried. If $r_t = r_{t+1}$, that tracer's counter increases by 1. If $r_t \neq r_{t+1}$, we immediately update a histogram (with \log_2 -spaced bins) with the the value of the trapping time. This procedure allows for a sizable reduction of the memory devoted to the measurements [277].

A similar procedure is used to calculate the values for ψ_B . A separate counter keeps track of the number of time steps that a tracer is below E_{th} in a basin. As soon as the tracer rises above E_{th} this counter is logged and reset in the same way we described for ψ_C .

6.A.3 Calculation of the aging functions

Finally, we make note of how we calculate values for Π_B during the simulation. Note that the calculation of Π_C is analogous, where instead of the basin index described further on, the configuration proxy r_t is used, in the same way that we described for ψ_C . The quantity Π_B is the probability of

not changing basin between two times. Stated another way, $\Pi_B(t_w, t_w + t)$ is the probability, being the tracer in some basin at t_w , that it is in the *same* basin at $(t + t_w) = t_w(1 + w)$ that the tracer is in that same basin (having not left). In this work we take $w = 0.5$.

To keep track of the particular basin a tracer is in, a basin index \mathcal{B}_j is kept for every tracer j and has the following properties.

1. If at time step t_w , tracer j is in its n^{th} basin (meaning it has entered and left $n - 1$ basins before t_w), then $\mathcal{B}_j = n$.
2. If the tracer has just left its n^{th} basin at t_w , then $\mathcal{B}_j = n + i$, where i is the imaginary number. The choice of using complex numbers to index whether a tracer is in or out of a basin is arbitrary, but allows for simpler notation in the code.
3. When the tracer reenters a basin, the imaginary component of \mathcal{B}_j is set back to 0, and the real part increments: $n \leftarrow n + 1$.

Thus in summary, the real part of \mathcal{B}_j references the index of the last basin that tracer was in, and the presence of an imaginary component is used to index whether or not that tracer is currently in or out of a basin. If $\text{Im}\{\mathcal{B}(t_w)\} \neq 0$, the measurement is discarded in computing the normalization of Π_B , since the tracer is initially not in a basin. If instead $\text{Im}\{\mathcal{B}_j(t_w)\} = 0$, then for a particular tracer,

- if $\mathcal{B}_j(t_w) = \mathcal{B}_j(t_w + t) \Rightarrow \Pi_B(t, t') = 1$
- if $\mathcal{B}_j(t_w) \neq \mathcal{B}_j(t_w + t) \Rightarrow \Pi_B(t, t') = 0$

As we described for the trapping time distributions, we extract a curve $\Pi(t_w, t_w(1 + w))$ from each batch of runs, and compute statistical error bars by comparing batches.

6.B MCMC calculation of the threshold

To evaluate the threshold radius in the MCMC approach we need to solve the equality ⁹

$$P_{\downarrow}(r_{\text{th}}; \delta) = P_{\uparrow}(r_{\text{th}}; \beta, \delta), \quad (6.18)$$

with

$$P_{\downarrow}(x_0; \delta) = N_{\delta} \int_{0 < |\mathbf{x}| < x_0} d\mathbf{x} e^{-\frac{(\mathbf{x}-\mathbf{x}_0)^2}{2\delta^2}}, \quad P_{\uparrow}(x_0; \beta, \delta) = N_{\delta} \int_{x_0 < |\mathbf{x}| < 1} d\mathbf{x} e^{-\frac{(\mathbf{x}-\mathbf{x}_0)^2}{2\delta^2}} e^{-\beta \Delta E}, \quad (6.19)$$

where we defined the normalization constant

$$N_{\delta} = \left(\int_{0 \leq |\mathbf{x}| < 1} d\mathbf{x} e^{-\frac{(\mathbf{x}-\mathbf{n})^2}{2\delta^2}} \right)^{-1}, \quad |\mathbf{n}| = 1. \quad (6.20)$$

We can express the integrals in spherical coordinates as

$$\int d\mathbf{x} e^{-\frac{(\mathbf{x}-\mathbf{n})^2}{2\alpha^2}} = \int d|\mathbf{x}| |\mathbf{x}|^{N-1} d\theta \Omega^N e^{-\frac{(|\mathbf{x}|^2+1-2|\mathbf{x}|\cos\theta)}{2\alpha^2}}, \quad (6.21)$$

where θ is the angle between \mathbf{x} and \mathbf{n} and Ω^N is the solid angle in N dimensions. The θ -integration can be singled out,

$$\int d\theta \Omega^N e^{\frac{|\mathbf{x}|\cos\theta}{\alpha^2}} = \tilde{\Omega}^{N-1} \int_{-1}^1 e^{\frac{|\mathbf{x}|\cos\theta}{\alpha^2}} d\cos\theta = \frac{\tilde{\Omega}^{N-1}\alpha^2}{|\mathbf{x}|} \left(e^{\frac{|\mathbf{x}|}{\alpha^2}} - e^{-\frac{|\mathbf{x}|}{\alpha^2}} \right), \quad (6.22)$$

and the last result can be substituted back into Eq. (6.21) to obtain

$$\int d\mathbf{x} e^{-\frac{(\mathbf{x}-\mathbf{n})^2}{2\alpha^2}} = \tilde{\Omega}^{N-1}\alpha^2 \int d|\mathbf{x}| |\mathbf{x}|^{N-2} \left(e^{-\frac{(|\mathbf{x}|-1)^2}{2\alpha^2}} - e^{-\frac{(|\mathbf{x}|+1)^2}{2\alpha^2}} \right), \quad (6.23)$$

⁹In this appendix we set $\beta_c = 1$.

where $\tilde{\Omega}^{N-1}$ is the result of the integration of the residual angular coordinates, and it is independent of the radial position. Rescaling the integrand in Eq. (6.19) we obtain

$$P_{\downarrow}(x_0; \delta) = x_0^N \left(\frac{\int_{0 < |\mathbf{x}| < 1} d\mathbf{x} e^{-\frac{(\mathbf{x}-\mathbf{n})^2}{2\alpha^2}}}{\int_{0 < |\mathbf{x}| < 1} d\mathbf{x} e^{-\frac{(\mathbf{x}-\mathbf{n})^2}{2\delta^2}}} \right); \quad \alpha = \frac{\delta}{x_0}. \quad (6.24)$$

Denoting $P_{\downarrow}(x_0)$ the corresponding probability in the DSMC, we have

$$P_{\downarrow}(x_0; \delta) \geq P_{\downarrow}(x_0), \quad \lim_{x_0 \rightarrow 1} P_{\downarrow}(x_0; \delta) = P_{\downarrow}(x_0). \quad (6.25)$$

In addition we can use Eq. (6.23) to show that for $N \gg 1$ and $\delta \propto N^{-\xi}$, $0 < \xi < 1$ we have¹⁰

$$\lim_{x_0 \rightarrow 0} P_{\downarrow}(x_0; \delta) = \frac{2}{\delta^2} P_{\downarrow}(x_0) \left[1 + O(N^{-1}) \right]. \quad (6.26)$$

With similar manipulations we obtain an equivalent expression for P_{\uparrow} ,

$$P_{\uparrow}(x_0; \beta, \delta) = x_0^N \left(\frac{\int_{1 < |\mathbf{x}| < \frac{1}{x_0}} d\mathbf{x} |\mathbf{x}|^{-\beta N} e^{-\frac{(\mathbf{x}-\mathbf{n})^2}{2\alpha^2}}}{\int_{0 < |\mathbf{x}| < 1} d\mathbf{x} e^{-\frac{(\mathbf{x}-\mathbf{n})^2}{2\delta^2}}} \right). \quad (6.27)$$

or, integrating out the angular coordinate,

$$P_{\uparrow}(x_0; \beta, \delta) = x_0^{N-2} \left(\frac{\int_1^{1/x_0} dy y^{N-2-\beta N} \left(e^{-\frac{(y-1)^2}{2\alpha^2}} - e^{-\frac{(y+1)^2}{2\alpha^2}} \right)}{\int_0^1 dy y^{N-2} \left(e^{-\frac{(y-1)^2}{2\delta^2}} - e^{-\frac{(y+1)^2}{2\delta^2}} \right)} \right). \quad (6.28)$$

We can relate $P_{\uparrow}(x_0; \beta, \delta)$ with $P_{\uparrow}(x_0; \beta)$, the probability of increasing the radius in the DSMC. In fact, it is easy to show that

$$\lim_{x_0 \rightarrow 0} P_{\uparrow}(x_0; \beta, \delta) = \frac{2}{\delta^2} P_{\uparrow}(x_0; \beta) \left[1 + O(N^{-1}) \right]. \quad (6.29)$$

¹⁰We assume this scaling for δ and N to be valid through the remaining of the appendix.

6.B.1 Threshold Radius for MCMC steps

We are interested in the ratio between the two probabilities, which can be expressed as

$$\frac{P_{\uparrow}(x_0; \beta, \delta)}{P_{\downarrow}(x_0; \delta)} = R(x_0; \beta, \alpha) = \frac{\int_{1 < |\mathbf{x}| < \frac{1}{x_0}} d\mathbf{x} |\mathbf{x}|^{-\beta N} e^{-\frac{(\mathbf{x}-\mathbf{n})^2}{2\alpha^2}}}{\int_{0 < |\mathbf{x}| < 1} d\mathbf{x} e^{-\frac{(\mathbf{x}-\mathbf{n})^2}{2\alpha^2}}}, \quad (6.30)$$

or, keeping only the radial coordinates,

$$R(x_0; \beta, \alpha) = \frac{\int_1^{1/x_0} dy y^{N-2-\beta N} \left(e^{-\frac{(y-1)^2}{2\alpha^2}} - e^{-\frac{(y+1)^2}{2\alpha^2}} \right)}{\int_0^1 dy y^{N-2} \left(e^{-\frac{(y-1)^2}{2\alpha^2}} - e^{-\frac{(y+1)^2}{2\alpha^2}} \right)}. \quad (6.31)$$

In order to find the threshold radius, we need to solve

$$R(r_{\text{th}}; \beta, \alpha) = 1. \quad (6.32)$$

In the limit $x_0 \rightarrow 0$ we find the same value for the ratio as in the DSMC (in contrast with the single probabilities, the ratio has no N^{-1} corrections):

$$\lim_{x_0 \rightarrow 0} R(x_0; \beta, \alpha) = \frac{1}{\beta - 1}, \quad (6.33)$$

which gives a threshold radius $r_{\text{th}} = 0$ for $\beta = 2$. By imposing a null variation of $R(x_0; \beta, \alpha)$ with respect to x_0 and β it is easy to show that the threshold radius is a decreasing function of β . Since $r_{\text{th}} = 0$ at $\beta = 2$, there cannot be any entropically activated dynamics for $\beta > 2$. Using a saddle-point-like approximation the ratio $R(x_0; \beta, \alpha)$ reduces to

$$R(x_0; \beta, \alpha) = \frac{\mathcal{N}(x_0; \beta, \alpha)}{D(\alpha)}, \quad (6.34)$$

$$\begin{aligned} \mathcal{N}(x_0; \beta, \alpha) \approx & \frac{\left(1 - e^{-\frac{2}{\alpha^2}}\right)}{1 + N\Delta\beta} \left[1 - x_0^{1+N\Delta\beta}\right] + \frac{2}{\alpha^2} e^{-\frac{2}{\alpha^2}} \frac{1 - x_0^{N\Delta\beta} [1 + (1 - x_0) N\Delta\beta]}{N\Delta\beta [1 + N\Delta\beta]} + \\ & + \frac{1}{2\alpha^2} \left(\left(1 - \frac{4}{\alpha^2}\right) e^{-\frac{2}{\alpha^2}} - 1\right) \frac{2 - x_0^{N\Delta\beta-1} (1 - x_0)^2 N^2 \Delta\beta^2}{N\Delta\beta [N\Delta\beta + 1] [N\Delta\beta - 1]}, \end{aligned} \quad (6.35)$$

$$D(\alpha) \approx \frac{\left(1 - e^{-\frac{2}{\alpha^2}}\right)}{N - 1} - \frac{2}{\alpha^2} e^{-\frac{2}{\alpha^2}} \frac{1}{N(N - 1)} + \frac{1}{2\alpha^2} \left(\left(1 - \frac{4}{\alpha^2}\right) e^{-\frac{2}{\alpha^2}} - 1\right) \frac{2}{N(N + 1)(N - 1)} \quad (6.36)$$

where we defined $\Delta\beta = \beta - 1$. As long as $x_0 \gg \delta$ the last expressions simplify to

$$R(x_0; \beta, \alpha) \approx \frac{\left(1 - x_0^{1+N\Delta\beta}\right)}{\Delta\beta} \left[1 - \frac{1 + \Delta\beta}{N\Delta\beta} - \frac{x_0^2}{N^2 \delta^2} \left(\frac{2 - x_0^{N\Delta\beta} (1 - x_0)^2 N^2 \Delta\beta^2}{2\Delta\beta^2 (1 - x_0^{N\Delta\beta})} - 1\right)\right]. \quad (6.37)$$

Note that the condition $x_0 \gg \delta$, given the scaling of δ with N that we assumed, is true in the whole sphere barring a negligible volume around the origin which becomes important only when $\beta \rightarrow 2$.

In this regime however the MCMC dynamics is well approximated by the DSMC one.

In Eq. (6.37), the last term in square brackets is sub-leading in N^{-1} , so the equation $R(r_{\text{th}}; \beta, \alpha) = 1$ can be solved perturbatively. At the lowest order in N^{-1} we obtain a solution similar to the DSMC one,

$$r_{\text{th}}^{(0)}(\beta) = (1 - \Delta\beta)^{\frac{1}{N\Delta\beta+1}}. \quad (6.38)$$

Plugging this term in $R(x_0; \beta, \alpha)$ we obtain

$$\begin{aligned} R(x_0; \beta, \alpha) \approx & \frac{\left(1 - x_0^{1+N\Delta\beta}\right)}{\Delta\beta} \\ & \times \left[1 - \frac{1 + \Delta\beta}{N\Delta\beta} - \frac{(1 - \Delta\beta)^{\frac{2}{N\Delta\beta}}}{N^2 \delta^2} \left(\frac{2 - (1 - \Delta\beta) \left(1 - (1 - \Delta\beta)^{\frac{1}{N\Delta\beta}}\right)^2 N^2 \Delta\beta^2}{2\Delta\beta^3} - 1\right)\right], \end{aligned} \quad (6.39)$$

and

$$r_{\text{th}}(\beta) \approx (1 - \Delta\beta)^{\frac{1}{N\Delta\beta+1}} \\ \times \left[1 - \frac{1 + \Delta\beta}{N^2 \Delta\beta (1 - \Delta\beta)} - \frac{(1 - \Delta\beta)^{\frac{2}{N\Delta\beta}-1}}{N^3 \delta^2} \left(\frac{2 - (1 - \Delta\beta) \left(1 - (1 - \Delta\beta)^{\frac{1}{N\Delta\beta}} \right)^2 N^2 \Delta\beta^2}{2\Delta\beta^3} - 1 \right) \right]. \quad (6.40)$$

To leading order Eq. (6.40) coincides with the DSMC threshold radius, Eq. (6.10).

Part IV

Back matter

Bibliography

- [1] M. Born and J. R. Oppenheimer, Ann. Physik **84** (1927).
- [2] J. J. Sakurai and J. J. Napolitano, *Modern Quantum Mechanics* (Cambridge University Press, 2021).
- [3] R. J. Bartlett and M. Musiał, Rev. Mod. Phys. **79**, 291 (2007).
- [4] K. Burke, J. Chem. Phys. **136**, 150901 (2012).
- [5] G. D. Mahan, *Many-particle Physics* (Springer Science & Business Media, 2013).
- [6] G. L. Goodvin, M. Berciu, and G. A. Sawatzky, Phys. Rev. B **74**, 245104 (2006).
- [7] A. L. Fetter and J. D. Walecka, *Quantum theory of many-particle systems* (Courier Corporation, 2012).
- [8] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).
- [9] N. Artrith and A. Urban, Comput. Mater. Sci. **114**, 135 (2016).
- [10] N. Artrith, A. Urban, and G. Ceder, Phys. Rev. B. **96**, 014112 (2017).
- [11] A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A. W. Nelson, A. Bridgland, *et al.*, Nature **577**, 706 (2020).
- [12] M. Berciu, Phys. Rev. Lett. **97**, 036402 (2006).
- [13] L. D. Landau, Phys. Z. Sowjetunion **3**, 664 (1933).

- [14] S. Pekar, *Zh. Eksp. Teor. Fiz.* **16**, 341 (1946).
- [15] L. Landau and S. Pekar, *Zh. Eksp. Teor. Fiz.* **18**, 419 (1948).
- [16] T. Lee, F. Low, and D. Pines, *Phys. Rev.* **90**, 297 (1953).
- [17] H. Fröhlich, H. Pelzer, and S. Zienau, *Lond. Edinb. Dubl. Philos. Mag. J. Sci.* **41**, 221 (1950).
- [18] H. Fröhlich, *Adv. Phys.* **3**, 325 (1954).
- [19] R. P. Feynman, *Phys. Rev.* **97**, 660 (1955).
- [20] T. Holstein, *Ann. Phys. (NY)* **8**, 325 (1959).
- [21] T. Holstein, *Ann. Phys. (NY)* **8**, 343 (1959).
- [22] A. Alexandrov and N. Mott, *Rep. Prog. Phys.* **57**, 1197 (1994).
- [23] G. Scholes and G. Rumbles, *Nat. Mater.* **5**, 683–696 (2006).
- [24] S. H. Park, A. Roy, S. Beaupre, S. Cho, N. Coates, J. S. Moon, D. Moses, M. Leclerc, K. Lee, and A. J. Heeger, *Nat. Photonics* **3**, 297 (2009).
- [25] F. C. Spano, *Acc. Chem. Res.* **43**, 429 (2010).
- [26] K. Miyata, D. Meggiolaro, M. T. Trinh, P. P. Joshi, E. Mosconi, S. C. Jones, F. De Angelis, and X.-Y. Zhu, *Sci. Adv.* **3**, e1701217 (2017).
- [27] S. Trugman, *Phys. Rev. B* **37**, 1597 (1988).
- [28] D. Basov, M. Fogler, and F. G. De Abajo, *Science* **354** (2016).
- [29] J. Feist, J. Galego, and F. J. Garcia-Vidal, *ACS Photonics* **5**, 205 (2018).
- [30] B. Xiang, R. F. Ribeiro, M. Du, L. Chen, Z. Yang, J. Wang, J. Yuen-Zhou, and W. Xiong, *Science* **368**, 665 (2020).

- [31] A. Schirotzek, C.-H. Wu, A. Sommer, and M. W. Zwierlein, Phys. Rev. Lett. **102**, 230402 (2009).
- [32] M. Koschorreck, D. Pertot, E. Vogt, B. Fröhlich, M. Feld, and M. Köhl, Nature **485**, 619 (2012).
- [33] N. B. Jørgensen, L. Wacker, K. T. Skalmstang, M. M. Parish, J. Levinsen, R. S. Christensen, G. M. Bruun, and J. J. Arlt, Phys. Rev. Lett. **117**, 055302 (2016).
- [34] M.-G. Hu, M. J. Van de Graaff, D. Kedar, J. P. Corson, E. A. Cornell, and D. S. Jin, Phys. Rev. Lett. **117**, 055301 (2016).
- [35] M. Baggioli and O. Pujolas, Phys. Rev. Lett. **114**, 251602 (2015).
- [36] J. Sous and M. Pretko, npj Quantum Mater. **5**, 1 (2020).
- [37] J. Sous and M. Pretko, Phys. Rev. B **102**, 214437 (2020).
- [38] J. Bonča, S. Trugman, and I. Batistić, Phys. Rev. B **60**, 1633 (1999).
- [39] J. Bonča, S. Maekawa, and T. Tohyama, Phys. Rev. B **76**, 035121 (2007).
- [40] J. Bonča and S. A. Trugman, Phys. Rev. B **103**, 054304 (2021).
- [41] G. De Filippis, V. Cataudella, A. Mishchenko, C. A. Perroni, and N. Nagaosa, Phys. Rev. B **80**, 195104 (2009).
- [42] E. Jeckelmann and S. R. White, Phys. Rev. B **57**, 6376 (1998).
- [43] C. Zhang, E. Jeckelmann, and S. R. White, Phys. Rev. B **60**, 14092 (1999).
- [44] F. Dorfner, L. Vidmar, C. Brockt, E. Jeckelmann, and F. Heidrich-Meisner, Phys. Rev. B **91**, 104302 (2015).
- [45] B. Kloss, D. R. Reichman, and R. Tempelaar, Phys. Rev. Lett. **123**, 126601 (2019).

- [46] N. V. Prokof'ev and B. V. Svistunov, Phys. Rev. Lett. **81**, 2514 (1998).
- [47] N. Prokof'ev and B. Svistunov, Phys. Rev. B **77**, 020408 (2008).
- [48] A. Mishchenko, N. Prokof'ev, A. Sakamoto, and B. Svistunov, Phys. Rev. B **62**, 6317 (2000).
- [49] J. T. Titantah, C. Pierleoni, and S. Ciuchi, Phys. Rev. Lett. **87**, 206406 (2001).
- [50] P. Kornilovitch, Phys. Rev. Lett. **81**, 5382 (1998).
- [51] M. Berciu and G. L. Goodvin, Phys. Rev. B **76**, 165109 (2007).
- [52] L. Covaci and M. Berciu, EPL (Europhysics Letters) **80**, 67001 (2007).
- [53] G. L. Goodvin, L. Covaci, and M. Berciu, Phys. Rev. Lett. **103**, 176402 (2009).
- [54] M. Berciu, A. S. Mishchenko, and N. Nagaosa, EPL (Europhysics Letters) **89**, 37007 (2010).
- [55] D. Marchand, G. De Filippis, V. Cataudella, M. Berciu, N. Nagaosa, N. Prokof'ev, A. Mishchenko, and P. Stamp, Phys. Rev. Lett. **105**, 266605 (2010).
- [56] M. Berciu and H. Fehske, Phys. Rev. B **82**, 085116 (2010).
- [57] D. J. Marchand, P. C. Stamp, and M. Berciu, Phys. Rev. B **95**, 035117 (2017).
- [58] C. P. J. Adolphs and M. Berciu, Phys. Rev. B **90**, 085149 (2014).
- [59] J. Sous, M. Chakraborty, R. V. Krems, and M. Berciu, Phys. Rev. Lett. **121**, 247001 (2018).
- [60] L. Covaci and M. Berciu, Phys. Rev. Lett. **100**, 256405 (2008).
- [61] H. Ebrahimnejad, G. A. Sawatzky, and M. Berciu, Nat. Phys. **10**, 951 (2014).
- [62] M. M. Möller, G. A. Sawatzky, M. Franz, and M. Berciu, Nat. Commun. **8**, 1 (2017).

- [63] Y. Tanimura and R. Kubo, J. Phys. Soc. Jpn. **58**, 101 (1989).
- [64] L. Chen, Y. Zhao, and Y. Tanimura, J. Phys. Chem. Lett. **6**, 3110 (2015).
- [65] E. N. Economou, *Green's Functions in Quantum Physics* (Springer-Verlag, Berlin, 1983).
- [66] G. De Filippis, V. Cataudella, A. Mishchenko, and N. Nagaosa, Phys. Rev. B **85**, 094302 (2012).
- [67] A. Alvermann, H. Fehske, and S. A. Trugman, Phys. Rev. B **81**, 165113 (2010).
- [68] M. Chakraborty and B. I. Min, Phys. Rev. B **88**, 024302 (2013).
- [69] W. Su, J. Schrieffer, and A. J. Heeger, Phys. Rev. Lett. **42**, 1698 (1979).
- [70] A. Macridin, *Ph. D. Thesis* (Rijksuniversiteit Groningen, 2003).
- [71] S. Barišić, J. Labbé, and J. Friedel, Phys. Rev. Lett. **25**, 919 (1970).
- [72] S. Barišić, Phys. Rev. B **5**, 932 (1972).
- [73] S. Barišić, Phys. Rev. B **5**, 941 (1972).
- [74] J. Sous, M. Chakraborty, C. Adolphs, R. Krems, and M. Berciu, Sci. Rep. **7**, 1 (2017).
- [75] K. Hannewald, V. Stojanović, J. Schellekens, P. Bobbert, G. Kresse, and J. Hafner, Phys. Rev. B **69**, 075211 (2004).
- [76] T. C. Berkelbach, M. S. Hybertsen, and D. R. Reichman, J. Chem. Phys. **138**, 114102 (2013).
- [77] T. C. Berkelbach, M. S. Hybertsen, and D. R. Reichman, J. Chem. Phys. **138**, 114103 (2013).
- [78] J. H. Fetherolf, D. Golež, and T. C. Berkelbach, Phys. Rev. X **10**, 021062 (2020).
- [79] G. L. Goodvin, A. S. Mishchenko, and M. Berciu, Phys. Rev. Lett. **107**, 076403 (2011).

- [80] J. Sous, M. Berciu, and R. V. Krems, Phys. Rev. A **96**, 063619 (2017).
- [81] Y. Takahashi and H. Umezawa, Int. J. Mod. Phys. B **10**, 1755 (1996).
- [82] M. Suzuki, J. Phys. Soc. Japan **54**, 4483 (1985).
- [83] R. R. Puri, *Mathematical methods of quantum optics*, Vol. 79 (Springer Science & Business Media, 2001).
- [84] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Laser Phys. **16**, 1572 (2006).
- [85] L. Chen and Y. Zhao, J. Chem. Phys. **147**, 214102 (2017).
- [86] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, Phys. Rev. Lett. **105**, 136805 (2010).
- [87] A. Splendiani, L. Sun, Y. Zhang, T. Li, J. Kim, C.-Y. Chim, G. Galli, and F. Wang, Nano Lett. **10**, 1271 (2010).
- [88] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, Science **306**, 666 (2004).
- [89] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, M. I. Katsnelson, I. V. Grigorieva, D. S. V., and A. A. Firsov, Nature **438**, 197 (2005).
- [90] K. P. Loh, Q. Bao, G. Eda, and M. Chhowalla, Nat. Chem. **2**, 1015 (2010).
- [91] G. Wang, A. Chernikov, M. M. Glazov, T. F. Heinz, X. Marie, T. Amand, and B. Urbaszek, Rev. Mod. Phys **90**, 021001 (2018).
- [92] T. C. Berkelbach and D. R. Reichman, Annu. Rev. Condens. Matter Phys. **9**, 379 (2018).
- [93] Z. Cai, B. Liu, X. Zou, and H.-M. Cheng, Chem. Rev. **118**, 6091 (2018).
- [94] M. Chhowalla, H. S. Shin, G. Eda, L.-J. Li, K. P. Loh, and H. Zhang, Nat. Chem. **5**, 263 (2013).

- [95] N. Flöry, A. Jain, P. Bharadwaj, M. Parzefall, T. Taniguchi, K. Watanabe, and L. Novotny, Appl. Phys. Lett **107**, 123106 (2015).
- [96] S. Memaran, N. R. Pradhan, Z. Lu, D. Rhodes, J. Ludwig, Q. Zhou, O. Ogunsolu, P. M. Ajayan, D. Smirnov, A. I. Fernández-Domínguez, *et al.*, Nano Lett. **15**, 7532 (2015).
- [97] S. Wi, M. Chen, D. Li, H. Nam, E. Meyhofer, and X. Liang, Appl. Phys. Lett **107**, 062102 (2015).
- [98] Q. H. Wang, K. Kalantar-Zadeh, A. Kis, J. N. Coleman, and M. S. Strano, Nat. Nanotechnol. **7**, 699 (2012).
- [99] D. Ovchinnikov, A. Allain, Y.-S. Huang, D. Dumcenco, and A. Kis, ACS Nano **8**, 8174 (2014).
- [100] D. Lembke, S. Bertolazzi, and A. Kis, Acc. Chem. Res. **48**, 100 (2015).
- [101] D. Y. Qiu, H. Felipe, and S. G. Louie, Phys. Rev. Lett. **111**, 216805 (2013).
- [102] F. Wu, F. Qu, and A. H. MacDonald, Phys. Rev. B **91**, 075310 (2015).
- [103] T. C. Berkelbach, M. S. Hybertsen, and D. R. Reichman, Phys. Rev. B **88**, 045318 (2013).
- [104] K. F. Mak, K. He, C. Lee, G. H. Lee, J. Hone, T. F. Heinz, and J. Shan, Nat. Mater **12**, 207 (2013).
- [105] T. Cheiwchanchamnangij and W. R. Lambrecht, Phys. Rev. B **85**, 205302 (2012).
- [106] A. Chernikov, T. C. Berkelbach, H. M. Hill, A. Rigosi, Y. Li, O. B. Aslan, D. R. Reichman, M. S. Hybertsen, and T. F. Heinz, Phys. Rev. Lett. **113**, 076802 (2014).
- [107] K. He, N. Kumar, L. Zhao, Z. Wang, K. F. Mak, H. Zhao, and J. Shan, Phys. Rev. Lett. **113**, 026803 (2014).

- [108] G. Plechinger, P. Nagler, J. Kraus, N. Paradiso, C. Strunk, C. Schüller, and T. Korn, Phys. Status Solidi RRL **9**, 457 (2015).
- [109] M. Currie, A. Hanbicki, G. Kioseoglou, and B. Jonker, Appl. Phys. Lett. **106**, 201907 (2015).
- [110] A. Mitioglu, P. Plochocka, J. Jadczak, W. Escoffier, G. Rikken, L. Kulyuk, and D. Maude, Phys. Rev. B **88**, 245403 (2013).
- [111] G. Moody, C. K. Dass, K. Hao, C.-H. Chen, L.-J. Li, A. Singh, K. Tran, G. Clark, X. Xu, G. Berghäuser, *et al.*, Nat. Commun. **6**, 8315 (2015).
- [112] M. Selig, G. Berghäuser, A. Raja, P. Nagler, C. Schüller, T. F. Heinz, T. Korn, A. Chernikov, E. Malic, and A. Knorr, Nat. Commun. **7**, 13279 (2016).
- [113] O. A. Ajayi, J. V. Ardelean, G. D. Shepard, J. Wang, A. Antony, T. Taniguchi, K. Watanabe, T. F. Heinz, S. Strauf, X. Zhu, *et al.*, 2D Mater. **4**, 031011 (2017).
- [114] F. Cadiz, E. Courtade, C. Robert, G. Wang, Y. Shen, H. Cai, T. Taniguchi, K. Watanabe, H. Carrere, D. Lagarde, *et al.*, Phys. Rev. X **7**, 021026 (2017).
- [115] S. Shree, A. George, T. Lehnert, C. Neumann, M. Benelajla, C. Robert, X. Marie, K. Watanabe, T. Taniguchi, U. Kaiser, *et al.*, 2D Mater. **7**, 015011 (2019).
- [116] D. Christiansen, M. Selig, G. Berghäuser, R. Schmidt, I. Niehues, R. Schneider, A. Arora, S. M. de Vasconcellos, R. Bratschitsch, E. Malic, *et al.*, Phys. Rev. Lett. **119**, 187402 (2017).
- [117] K. Shinokita, X. Wang, Y. Miyauchi, K. Watanabe, T. Taniguchi, and K. Matsuda, Adv. Funct. Mater. , 1900260 (2019).
- [118] Y. Miyauchi, S. Konabe, F. Wang, W. Zhang, A. Hwang, Y. Hasegawa, L. Zhou, S. Mouri, M. Toh, G. Eda, *et al.*, Nat. Commun. **9**, 2598 (2018).

- [119] B. R. Carvalho, L. M. Malard, J. M. Alves, C. Fantini, and M. A. Pimenta, Phys. Rev. Lett. **114**, 136403 (2015).
- [120] M. Sidler, P. Back, O. Cotlet, A. Srivastava, T. Fink, M. Kroner, E. Demler, and A. Imamoglu, Nat. Phys **13**, 255 (2017).
- [121] P.-F. Li and Z.-W. Wang, J. Appl. Phys **123**, 204308 (2018).
- [122] Y.-W. Chang and D. R. Reichman, Phys. Rev. B **99**, 125421 (2019).
- [123] D. K. Efimkin and A. H. MacDonald, Phys. Rev. B **95**, 035417 (2017).
- [124] M. Klawunn and A. Recati, Phys. Rev. A **84**, 033607 (2011).
- [125] R. Schmidt, T. Enss, V. Pietilä, and E. Demler, Phys. Rev. A **85**, 021602 (2012).
- [126] M. M. Parish and J. Levinsen, Phys. Rev. A **87**, 033616 (2013).
- [127] A. Lherbier, X. Blase, Y.-M. Niquet, F. Triozon, and S. Roche, Phys. Rev. Lett. **101**, 036808 (2008).
- [128] G. Ramon, A. Mann, and E. Cohen, Phys. Rev. B **67**, 045323 (2003).
- [129] J. Henriques, G. Ventura, C. Fernandes, and N. Peres, J. Phys. Condens. Matter **32**, 025304 (2019).
- [130] N. S. Rytova, Vestn. Mosk. Univ. Fiz. Astron. **3**, 30 (1967).
- [131] L. Keldysh, J. Exp. theoret. Phys. Lett. **29**, 658 (1979).
- [132] S. Chandrasekhar, Astrophys. J **100**, 176 (1944).
- [133] N. P. Sandler and C. R. Proetto, Phys. Rev. B **46**, 7707 (1992).
- [134] F. Stern, Phys. Rev. Lett. **18**, 546 (1967).
- [135] T. C. Berkelbach, “Two-dimensional screening for doped semiconductors,” Unpublished.

- [136] G. F. Giuliani and J. J. Quinn, Phys. Rev. B **26**, 4421 (1982).
- [137] G. Strinati, La Rivista del Nuovo Cimento (1978-1999) **11**, 1 (1988).
- [138] M. Rohlffing and S. G. Louie, Phys. Rev. B **62**, 4927 (2000).
- [139] W. Hanke and L. Sham, Phys. Rev. Lett. **43**, 387 (1979).
- [140] L. X. Benedict, Phys. Rev. B **66**, 193105 (2002).
- [141] M. M. Glazov and A. Chernikov, Phys. Status Solidi B **255**, 1800216 (2018).
- [142] B. Scharf, D. Van Tuan, I. Žutić, and H. Dery, J. Phys. Condens. Matter **31**, 203001 (2019).
- [143] Y.-C. Chang, S.-Y. Shiau, and M. Combescot, arXiv preprint arXiv:1810.13061 (2018).
- [144] K. Andersen, S. Latini, and K. S. Thygesen, Nano Lett. **15**, 4616 (2015).
- [145] L. Cavalcante, A. Chaves, B. Van Duppen, F. Peeters, and D. Reichman, Phys. Rev. B **97**, 125427 (2018).
- [146] P. J. Gielisse, S. S. Mitra, J. N. Plendl, R. D. Griffis, L. C. Mansur, R. Marshall, and E. A. Pascoe, Phys. Rev. **155**, 1039 (1967).
- [147] F. A. Rasmussen and K. S. Thygesen, J. Phys. Chem. C **119**, 13169 (2015).
- [148] G. Astakhov, V. Kochereshko, D. Yakovlev, W. Ossau, J. Nürnberg, W. Faschinger, and G. Landwehr, Phys. Rev. B **62**, 10345 (2000).
- [149] T. Smołeński, O. Cotlet, A. Popert, P. Back, Y. Shimazaki, P. Knüppel, N. Dietler, T. Taniguchi, K. Watanabe, M. Kroner, *et al.*, arXiv preprint arXiv:1812.08772 (2018).
- [150] E. Courtade, M. Semina, M. Manca, M. Glazov, C. Robert, F. Cadiz, G. Wang, T. Taniguchi, K. Watanabe, M. Pierre, *et al.*, Phys. Rev. B **96**, 085302 (2017).
- [151] Z. Ye, L. Waldecker, E. Y. Ma, D. Rhodes, A. Antony, B. Kim, X.-X. Zhang, M. Deng, Y. Jiang, Z. Lu, *et al.*, Nat. Commun. **9**, 3718 (2018).

- [152] C. Fey, P. Schmelcher, A. Imamoglu, and R. Schmidt, arXiv preprint arXiv:1912.04873 (2019).
- [153] L. Huang and T. D. Krauss, Phys. Rev. Lett. **96**, 057407 (2006).
- [154] F. Wang, Y. Wu, M. S. Hybertsen, and T. F. Heinz, Phys. Rev. B **73**, 245424 (2006).
- [155] F. Hu, C. Yin, H. Zhang, C. Sun, W. W. Yu, C. Zhang, X. Wang, Y. Zhang, and M. Xiao, Nano Lett. **16**, 6425 (2016).
- [156] Steven G. Johnson, “Cubature (Multi-dimensional integration),” [http://ab-initio.mit.edu/wiki/index.php/Cubature_\(Multi-dimensional_integration\)](http://ab-initio.mit.edu/wiki/index.php/Cubature_(Multi-dimensional_integration)) (2017).
- [157] K. Wagner, E. Wietek, J. D. Ziegler, M. A. Semina, T. Taniguchi, K. Watanabe, J. Zipfel, M. M. Glazov, and A. Chernikov, Phys. Rev. Lett. **125**, 267401 (2020).
- [158] J. D. Jorgensen, M. A. Beno, D. G. Hinks, L. Soderholm, K. J. Volin, R. L. Hitterman, J. D. Grace, I. K. Schuller, C. U. Segre, K. Zhang, and M. S. Kleefisch, Phys. Rev. B **36**, 3608 (1987).
- [159] A. Cavalleri, C. Tóth, C. W. Siders, J. A. Squier, F. Ráksi, P. Forget, and J. C. Kieffer, Phys. Rev. Lett. **87**, 237401 (2001).
- [160] B. Kang and G. Ceder, Nature **458**, 190 (2009).
- [161] H. Peng, P. F. Ndione, D. S. Ginley, A. Zakutayev, and S. Lany, Phys. Rev. X **5**, 021016 (2015).
- [162] D. P. Tabor, L. M. Roch, S. K. Saikin, C. Kreisbeck, D. Sheberla, J. H. Montoya, S. Dwarak-nath, M. Aykol, C. Ortiz, H. Tribukait, *et al.*, Nat. Rev. Mater. **3**, 5 (2018).
- [163] J. E. Gubernatis and T. Lookman, Phys. Rev. Mater. **2**, 120301 (2018).
- [164] F. Meirer and B. M. Weckhuysen, Nat. Rev. Mater. **3**, 324 (2018).

- [165] A. Ankudinov, J. Rehr, J. J. Low, and S. R. Bare, *J. Chem. Phys.* **116**, 1911 (2002).
- [166] J. J. Rehr, J. J. Kas, M. P. Prange, A. P. Sorini, Y. Takimoto, and F. Vila, *C. R. Phys.* **10**, 548 (2009).
- [167] A. I. Frenkel, J. A. Rodriguez, and J. G. Chen, *ACS Catal.* **2**, 2269 (2012).
- [168] A. L. Ankudinov, B. Ravel, J. J. Rehr, and S. D. Conradson, *Phys. Rev. B* **58**, 7565 (1998).
- [169] D. Bazin and J. J. Rehr, *J. Phys. Chem. B* **107**, 12398 (2003).
- [170] G. Ciatto, A. Di Trolio, E. Fonda, P. Alippi, A. M. Testa, and A. A. Bonapasta, *Phys. Rev. Lett.* **107**, 127206 (2011).
- [171] Q. Ma, J. Prater, C. Sudakar, R. Rosenberg, and J. Narayan, *J. Phys. Condens. Matter* **24**, 306002 (2012).
- [172] A. Kuzmin and J. Chaboy, *IUCrJ* **1**, 571 (2014).
- [173] U. Srivastava and H. Nigam, *Coord. Chem. Rev.* **9**, 275 (1973).
- [174] T. Yamamoto, *X-Ray Spectrom.* **37**, 572 (2008).
- [175] H. Hanson and W. W. Beeman, *Phys. Rev.* **76**, 118 (1949).
- [176] J. Wong, F. W. Lytle, R. P. Messmer, and D. H. Maylotte, *Phys. Rev. B* **30**, 5596 (1984).
- [177] F. Farges, G. E. Brown Jr, and J. J. Rehr, *Geochim. Cosmochim. Acta* **60**, 3023 (1996).
- [178] F. Farges, G. E. Brown, and J. J. Rehr, *Phys. Rev. B* **56**, 1809 (1997).
- [179] F. Farges, G. E. Brown Jr, P.-E. Petit, and M. Munoz, *Geochim. Cosmochim. Acta* **65**, 1665 (2001).
- [180] W. E. Jackson, F. Farges, M. Yeager, P. A. Mabrouk, S. Rossano, G. A. Waychunas, E. I. Solomon, and G. E. Brown, *Geochim. Cosmochim. Acta* **69**, 4315 (2005).

- [181] F. A. Cotton and C. J. Ballhausen, *J. Chem. Phys.* **25**, 617 (1956).
- [182] F. A. Cotton and H. P. Hanson, *J. Chem. Phys.* **25**, 619 (1956).
- [183] G. Mountjoy, D. M. Pickup, G. Wallidge, R. Anderson, J. M. Cole, R. J. Newport, and M. E. Smith, *Chem. Mater.* **11**, 1253 (1999).
- [184] S. Bordiga, E. Groppo, G. Agostini, J. A. van Bokhoven, and C. Lamberti, *Chem. Rev.* **113**, 1736 (2013).
- [185] N. Jiang, D. Su, and J. C. H. Spence, *Phys. Rev. B* **76**, 214117 (2007).
- [186] S. Yoshida, T. Tanaka, T. Hanada, T. Hiraiwa, H. Kanai, and T. Funabiki, *Catal. Lett.* **12**, 277 (1992).
- [187] T. Tanaka, H. Yamashita, R. Tsuchitani, T. Funabiki, and S. Yoshida, *J. Chem. Soc., Faraday Trans. 1* **84**, 2987 (1988).
- [188] W. Zhang, M. Topsakal, C. Cama, C. J. Pelliccione, H. Zhao, S. Ehrlich, L. Wu, Y. Zhu, A. I. Frenkel, K. J. Takeuchi, *et al.*, *J. Am. Chem. Soc.* **139**, 16591 (2017).
- [189] G. Carleo and M. Troyer, *Science* **355**, 602 (2017).
- [190] E. P. L. van Nieuwenburg, Y.-H. Liu, and S. D. Huber, *Nat. Phys.* **13**, 435 (2017).
- [191] J. R. Kitchin, *Nat. Catal.* **1**, 230 (2018).
- [192] J. Timoshenko, D. Lu, Y. Lin, and A. I. Frenkel, *J. Phys. Chem. Lett.* **8**, 5091 (2017).
- [193] J. Timoshenko, A. Anspoks, A. Cintins, A. Kuzmin, J. Purans, and A. I. Frenkel, *Phys. Rev. Lett.* **120**, 225502 (2018).
- [194] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, *APL Mater.* **1**, 011002 (2013).

- [195] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, *Comput. Mater. Sci.* **68**, 314 (2013).
- [196] S. P. Ong, S. Cholia, A. Jain, M. Brafman, D. Gunter, G. Ceder, and K. A. Persson, *Comput. Mater. Sci.* **97**, 209 (2015).
- [197] J. J. Rehr and R. C. Albers, *Rev. Mod. Phys.* **72** (2000).
- [198] M. Taillefumier, D. Cabaret, A.-M. Flank, and F. Mauri, *Phys. Rev. B* **66**, 195107 (2002).
- [199] C. Gougoussis, M. Calandra, A. Seitsonen, C. Brouder, A. Shukla, and F. Mauri, *Phys. Rev. B* **79**, 045118 (2009).
- [200] D. Prendergast and G. Galli, *Phys. Rev. Lett.* **96**, 215502 (2006).
- [201] W. Chen, X. Wu, and R. Car, *Phys. Rev. Lett.* **105**, 017802 (2010).
- [202] J. Vinson, J. J. Rehr, J. J. Kas, and E. L. Shirley, *Phys. Rev. B* **83**, 115106 (2011).
- [203] J. Vinson, T. Jach, W. T. Elam, and J. D. Denlinger, *Phys. Rev. B* **90**, 205207 (2014).
- [204] A. Gulans, S. Kontur, C. Meisenbichler, D. Nabok, P. Pavone, S. Rigamonti, S. Sagmeister, U. Werner, and C. Draxl, *J. Phys. Condens. Matter* **26**, 363202 (2014).
- [205] C. Vorwerk, C. Cocchi, and C. Draxl, *Phys. Rev. B* **95**, 155121 (2017).
- [206] Y. Liang, J. Vinson, S. Pemmaraju, W. S. Drisdell, E. L. Shirley, and D. Prendergast, *Phys. Rev. Lett.* **118**, 096402 (2017).
- [207] Z. Sun, L. Zheng, M. Chen, M. L. Klein, F. Paesani, and X. Wu, *Phys. Rev. Lett.* **121**, 137401 (2018).
- [208] J. Timoshenko, C. J. Wräsmann, M. Luneau, T. Shirman, M. Cargnello, S. R. Bare, J. Aizenberg, C. M. Friend, and A. I. Frenkel, *Nano Lett.* **19**, 520 (2018).

- [209] M. W. Small, J. J. Kas, K. O. Kvashnina, J. J. Rehr, R. G. Nuzzo, M. Tromp, and A. I. Frenkel, *ChemPhysChem* **15**, 1569 (2014).
- [210] J. Timoshenko, A. Shihhare, R. W. J. Scott, D. Lu, and A. I. Frenkel, *Phys. Chem. Chem. Phys.* **18**, 19621 (2016).
- [211] H. Singh, M. Topsakal, K. Attenkofer, T. Wolf, M. Leskes, Y. Duan, F. Wang, J. Vinson, D. Lu, and A. I. Frenkel, *Phys. Rev. Mater.* **2**, 125403 (2018).
- [212] K. Mathew, C. Zheng, D. Winston, C. Chen, A. Dozier, J. J. Rehr, S. P. Ong, and K. A. Persson, *Sci. Data* **5**, 180151 (2018).
- [213] C. Zheng, K. Mathew, C. Chen, Y. Chen, H. Tang, A. Dozier, J. J. Kas, F. D. Vila, J. J. Rehr, L. F. Piper, *et al.*, *npj Comput. Mater.* **4**, 12 (2018).
- [214] D. Yan, M. Topsakal, S. Selcuk, J. Lyons, W. Zhang, Q. Wu, I. Waluyo, E. Stavitski, K. Attenkofer, S. Yoo, D. Lu, M. Hybertsen, D. Stacchiola, and M. Liu, unpublished .
- [215] J. Rehr, J. Kas, F. Vila, M. Prange, and K. Jorissen, *Phys. Chem. Chem. Phys.* **12**, 5503 (2010).
- [216] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, *Phys. Rev. Lett.* **108**, 058301 (2012).
- [217] J. S. Smith, O. Isayev, and A. E. Roitberg, *Chem. Sci.* **8**, 3192 (2017).
- [218] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, *Sci. Adv.* **3** (2017).
- [219] M. Pinsky and D. Avnir, *Inorg. Chem.* **37** (1998).
- [220] D. Waroquiers, X. Gonze, G.-M. Rignanese, C. Welker-Nieuwoudt, F. Rosowski, M. Göbel, S. Schenk, P. Degelmann, R. André, R. Glaum, and G. Hautier, *Chem. Mater.* **29**, 8346 (2017).

- [221] D. P. Kingma and J. Ba, arXiv preprint arXiv:1412.6980 (2014).
- [222] F. Chollet *et al.*, “Keras,” <https://keras.io> (2015).
- [223] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, (2015), software available from tensorflow.org.
- [224] K. Pearson, Philos. Mag. **2**, 559 (1901).
- [225] A. Manceau, A. I. Gorshkov, and V. A. Drits, Am. Mineral **77**, 1133 (1992).
- [226] H. Yoshitake, T. Sugihara, and T. Tatsumi, Phys. Chem. Chem. Phys. **5**, 767 (2003).
- [227] Y. LeCun, Y. Bengio, and G. Hinton, Nature **521**, 436 (2015).
- [228] K. G. Reyes and B. Maruyama, MRS Bulletin **44**, 530 (2019).
- [229] J. Schmidt, M. R. Marques, S. Botti, and M. A. Marques, npj Comput. Mater. **5**, 1 (2019).
- [230] M. Rupp, R. Ramakrishnan, and O. A. Von Lilienfeld, J. Phys. Chem. Lett **6**, 3309 (2015).
- [231] T. Xie and J. C. Grossman, Phys. Rev. Lett. **120**, 145301 (2018).
- [232] M. Gastegger, J. Behler, and P. Marquetand, Chem. Sci. **8**, 6924 (2017).
- [233] S. Ye, W. Hu, X. Li, J. Zhang, K. Zhong, G. Zhang, Y. Luo, S. Mukamel, and J. Jiang, PNAS **116**, 11612 (2019).
- [234] R. A. Vargas-Hernández, J. Sous, M. Berciu, and R. V. Krems, Phys. Rev. Lett. **121**, 255702 (2018).
- [235] J. F. Rodriguez-Nieva and M. S. Scheurer, Nat. Phys. **15**, 790 (2019).

- [236] B. Sanchez-Lengeling and A. Aspuru-Guzik, *Science* **361**, 360 (2018).
- [237] L. Zhang, J. Han, H. Wang, R. Car, and E. Weinan, *Phys. Rev. Lett.* **120**, 143001 (2018).
- [238] E. Alpaydin, *Introduction to machine learning* (MIT press, 2009).
- [239] A. Ankudinov, J. Rehr, J. J. Low, and S. R. Bare, *J. Chem. Phys.* **116**, 1911 (2002).
- [240] J. J. Rehr and R. C. Albers, *Rev. Mod. Phys.* **72**, 621 (2000).
- [241] A. Kuzmin and J. Chaboy, *IUCrJ* **1**, 571 (2014).
- [242] M. Taillefumier, D. Cabaret, A.-M. Flank, and F. Mauri, *Phys. Rev. B* **66**, 195107 (2002).
- [243] F. De Groot and A. Kotani, *Core level spectroscopy of solids* (CRC press, 2008).
- [244] J. Vinson, J. J. Rehr, J. J. Kas, and E. L. Shirley, *Phys. Rev. B* **83**, 115106 (2011).
- [245] A. Gulans, S. Kontur, C. Meisenbichler, D. Nabok, P. Pavone, S. Rigamonti, S. Sagmeister, U. Werner, and C. Draxl, *J. Phys. Condens. Matter* **26**, 363202 (2014).
- [246] M. R. Carbone, S. Yoo, M. Topsakal, and D. Lu, *Phys. Rev. Mater.* **3**, 033604 (2019).
- [247] J. Rehr, J. Kozdon, J. Kas, H. Krappe, and H. Rossner, *J. Synchrotron Radiat.* **12**, 70 (2005).
- [248] F. Farges, G. E. Brown, J. Rehr, *et al.*, *Phys. Rev. B* **56**, 1809 (1997).
- [249] D. Yan, M. Topsakal, S. Selcuk, J. L. Lyons, W. Zhang, Q. Wu, I. Waluyo, E. Stavitski, K. Attenkofer, S. Yoo, *et al.*, *Nano Lett.* (2019).
- [250] O. Trejo, A. L. Dadlani, F. De La Paz, S. Acharya, R. Kravec, D. Nordlund, R. Sarangi, F. B. Prinz, J. Torgersen, and N. P. Dasgupta, *Chem. Mater.* (2019).
- [251] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, in *Proceedings of the 34th International Conference on Machine Learning-Volume 70* (JMLR.org, 2017) pp. 1263–1272.

- [252] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, *Sci. Data* **1** (2014).
- [253] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, in *NIPS Autodiff Workshop* (2017).
- [254] A. Hagberg, P. Swart, and D. S Chult, *Exploring network structure, dynamics, and function using NetworkX*, Tech. Rep. (Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008).
- [255] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, arXiv preprint arXiv:1409.1259 (2014).
- [256] K. Kim, P. Zhu, N. Li, X. Ma, and Y. Chen, *Carbon* **49**, 1745 (2011).
- [257] O. Plekan, V. Feyer, R. Richter, M. Coreno, M. De Simone, K. Prince, and V. Caravetta, *J Electron Spectrosc* **155**, 47 (2007).
- [258] K. Pearson, *Philos. Mag.* **2**, 559 (1901).
- [259] RDKit, online, “RDKit: Open-source cheminformatics,” <http://www.rdkit.org>.
- [260] T. T. Tanimoto, Tech. Rep., IBM Internal Report (1958).
- [261] Q. Zhou, P. Tang, S. Liu, J. Pan, Q. Yan, and S.-C. Zhang, *Proc. Natl. Acad. Sci.* **115**, E6411 (2018).
- [262] P. G. Debenedetti, *Metastable Liquids* (Princeton University Press, Princeton, 1997).
- [263] A. Cavagna, *Phys. Rep.* **476**, 51 (2009).
- [264] J. Kurchan, *Glasses and grains: Poincaré seminar 2009* (Springer, Berlin, 2011), Chap. 1 (2011).
- [265] A. Crisanti and F. Ritort, *EPL (Europhysics Letters)* **52**, 640 (2000).
- [266] A. Crisanti and F. Ritort, *EPL (Europhysics Letters)* **51**, 147 (2000).

- [267] G. Ben-Arous, A. Bovier, and V. Gayrard, Phys. Rev. Lett. **88**, 087201 (2002).
- [268] V. Gayrard, Annales Henri Poincaré **17**, 537 (2016).
- [269] V. Gayrard, Probab. Theory Relat. Fields (2018).
- [270] A. Crisanti and F. Ritort, EPL (Europhysics Letters) **66**, 253 (2004).
- [271] A. Heuer, J. Phys. Condens. Matter **20**, 373101 (2008).
- [272] D. A. Stariolo and L. F. Cugliandolo, EPL (Europhysics Letters) **127**, 16002 (2019).
- [273] J. C. Dyre, Phys. Rev. Lett. **58**, 792 (1987).
- [274] J. Bouchaud, J. Phys. I France **2**, 1705 (1992).
- [275] R. A. Denny, D. R. Reichman, and J.-P. Bouchaud, Phys. Rev. Lett. **90**, 025503 (2003).
- [276] B. Doliwa and A. Heuer, Phys. Rev. Lett. **91**, 235501 (2003).
- [277] M. Baity-Jesi, G. Biroli, and C. Cammarota, J. Stat. Mech.: Theory Exp., 013301 (2018).
- [278] J. C. Dyre, Phys. Rev. B **51**, 12276 (1995).
- [279] C. Cammarota and E. Marinari, Phys. Rev. E **92**, 010301(R) (2015).
- [280] I. Junier and J. Kurchan, J. Phys. A **37**, 3945 (2004).
- [281] I. Junier and E. Bertin, Phys. Rev. E **70**, 066126 (2004).
- [282] M. Baity-Jesi, A. Achard-de Lustrac, and G. Biroli, Phys. Rev. E **98**, 012133 (2018).
- [283] M. Baity-Jesi and D. Reichman, (2020), in preparation.
- [284] G. Ben-Arous and J. Černý, in *Mathematical Statistical Physics*, Les Houches, Vol. 83, edited by A. Bovier, F. Dunlop, A. van Enter, F. den Hollander, and J. Dalibard (Elsevier, 2006) pp. 331 – 394.

- [285] V. Gayrard and L. Hartung, in *International Conference on Statistical Mechanics of Classical and Disordered Systems* (Springer, 2018) pp. 111–170.
- [286] R. G. Margiotta, R. Kühn, and P. Sollich, *J. Phys. A Math. Theor.* **51**, 294001 (2018).
- [287] R. G. Margiotta, R. Kühn, and P. Sollich, arXiv preprint arXiv:1906.07434 (2019).
- [288] A. Barrat and M. Mézard, *Journal de Physique I* **5**, 941 (1995).
- [289] E. Bertin, *J. Phys. A* **43**, 345002 (2010).
- [290] E. M. Bertin, *J. Phys. A: Math. Gen.* **36**, 10683 (2003).
- [291] E. Bertin, *Phys. Rev. E* **85**, 042104 (2012).
- [292] I. Hartarsky, M. Baity-Jesi, R. Ravasio, A. Billoire, and G. Biroli, *J. Stat. Mech.: Theory Exp* **2019**, 093302 (2019).
- [293] J.-P. Bouchaud and D.S. Dean, *J. Phys. I France* **5**, 265 (1995).
- [294] C. Cammarota and G. Biroli, *PNAS* **109**, 8850 (2012).
- [295] S. Karmakar and G. Parisi, *PNAS* **110**, 2752 (2013).
- [296] L. Angelani, M. Paoluzzi, G. Parisi, and G. Ruocco, *PNAS* **115**, 8700 (2018).
- [297] M. Baity-Jesi, V. Martín-Mayor, G. Parisi, and S. Perez-Gaviro, *Phys. Rev. Lett.* **115**, 267205 (2015).
- [298] J. A. R. Fris, G. A. Appignanesi, and E. R. Weeks, *Phys. Rev. Lett.* **107**, 065704 (2011).
- [299] M. Vogel and S. C. Glotzer, *Phys. Rev. Lett.* **92**, 255901 (2004).
- [300] C. Weiss, S. Page, and M. Holthaus, *Physica A: Statistical Mechanics and its Applications* **341**, 586 (2004).
- [301] F. Barthe, O. Guédon, S. Mendelson, A. Naor, *et al.*, *Ann. Probab.* **33**, 480 (2005).

List of publications & preprints as a graduate student

8. Bond-Peierls polaron: Non-exponential mass enhancement and current carrying ground state [[arXiv](#)]
M. R. Carbone,* Andrew J. Millis,* David R. Reichman* & John Sous*
7. Numerically Exact Generalized Green's Function Cluster Expansions for Electron-Phonon Problems [[arXiv](#)]
M. R. Carbone,* D. R. Reichman & J. Sous.* (under review, Phys. Rev. B.)
6. Predicting impurity spectral functions using machine learning [[arXiv](#)]
E. J. Sturm,^{†,*} M. R. Carbone,^{†,*} D. Lu, A. Weichselbaum & R. M. Konik. (accepted, Phys. Rev. B)
5. Random Forest machine learning models for interpretable x-ray absorption near-edge structure spectrum-property relationships [[ChemRxiv](#)]
S. B. Torrisi,* M. R. Carbone, B. A. Rohr, J. H. Montoya, Y. Ha, J. Yano, S. K. Suram* & L. Hung.* npj Comput Mater **6**, 109 (2020).
4. Microscopic model of the doping dependence of line widths in monolayer transition metal dichalcogenides [[arXiv](#)]
M. R. Carbone,* M. Z. Mayers & D. R. Reichman. J. Chem. Phys. **152**, 194705 (2020).
Part of a special issue on 2D materials
3. Effective trap-like activated dynamics in a continuous landscape [[arXiv](#)]
M. R. Carbone,* V. Astuti & M. Baity-Jesi. Phys. Rev. E **101**, 052304 (2020).

2. Machine-learning x-ray absorption spectra to quantitative accuracy [arXiv]

M. R. Carbone, M. Topsakal,* D. Lu* & S. Yoo.* Phys. Rev. Lett. **124**, 156401 (2020).

1. Classification of local chemical environments from x-ray absorption spectra using supervised machine learning [arXiv]

M. R. Carbone, S. Yoo, M. Topsakal* & D. Lu.* Phys. Rev. Mater **3**, 033604 (2019).

Editors' suggestion

† Equally contributing first authors

* Corresponding authors.