First-principles calculations of electron-defect interactions and defect-limited charge transport

Thesis by
I-Te Lu

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2020
Defended May 26, 2020

© 2020

I-Te Lu
ORCID: 0000-0003-1649-7931

All rights reserved

ACKNOWLEDGEMENTS

At the end of the summer in my first year at Caltech (2015), I thrillingly met my advisor, Marco Bernardi, who joined Caltech as a new faculty member.

“I want to surpass or reach your level to graduate,” I said to Marco, who seemed not offended.

“Yes, I believe you will, and you have to read a lot of articles and books day and night. I do not doubt that you will be better than me in your research area,” Marco answered with his warm smile. At that moment, I saw his passion in his eyes for helping students succeed and become independent in doing science.

Thanks to Marco’s professional and intense training, I have seen the boundary of my research field and had the opportunity to push its frontier forward. At the time of writing, though I am not confident to say that I have reached his level, I do understand that the purpose of doing research, at least for me, is not to surpass anyone, but to figure out what to contribute to science.

I am blessed and grateful to have Marco as my advisor, mentor, and friend. He always encourages and supports me and other group members to take opportunities to learn and explore. I have attended a few scientific workshops and given oral presentations at the APS March meetings since 2016 except for this year (2020) due to the coronavirus pandemic.

In addition to my advisor, I would like to thank for the help and advice from many professors at the Applied Physics and Materials Science department. I thank Prof. Brent Fultz for serving in my candidacy and thesis defense committee; he also helped us, materials science graduate students, a lot as a former option representative. I thank Prof. Austin Minnich and Prof. Stevan Nadj-Perge for being committee members of my thesis defense and sharing their wisdom about how to do research during the faculty-student coffee hours. I also appreciate Prof. Bill Johnson who served on my candidacy committee and gave me career advice. I thank Prof. Sandra Troian, Prof. Kerry Vahala, Prof. Andrei Faraon, Prof. Harry Atwater, and Prof. Julia Greer for sharing their stories in their faculty-student coffee hours.

I thank the administrative staff members, Christy Jenstad, Jennifer Blankenship, and Celene Gates for dealing with administrative stuff for us such that we can totally focus on doing good science.

It is my pleasure to see the growth of Marco's group from a small three-person group (Marco, Vatsal, and I) to a medium-size group of around 10 persons. I feel blessed to have my current and past colleagues and friends in the group: Dr. Luis Agapito, Dr. Jin-Jian Zhou, Dr. Vatsal Jhalani, Dr. Raffaello Bianoc, Dr. Shiyuan Gao, Dr. Ivan Maliyov, Vsevolod Ivanov, Nien-En Lee, Austin Liu, Megan Schill, Jinsoo Park, Hsiao-Yi Chen, Xiao Tong, Chao-Jung Lee, Benjamin K. Chang, Dhruv Desai, and Ina M. Sørensen. From them, I have learned many languages, computer programming skills, and different cultures. I also had the pleasure to work with some promising undergraduate students: Kyle Frohna, Anita S. Kulkarni, Deepak Dhariwal, Aditya Sriram, Shreshth Malik, Bogdan Zviazhynski, and Vishal Hariharan.

I am grateful to have my wonderful classmates and friends in materials science: Neal Brodnik, Wei-Lin Tan, Jin Qian, Max Lifson, Mike Citrin, Sophia Cheng, Jong Hun Kang, Daryl Yee, Harpreet Arora, Xiaolin Xu, and Danielle Duggins. We had worked together during our first and second years and had a great time ever since then.

I thank Prof. Pu-Wei Wu, Dr. Lee-Jene Lai, among others at my previous university for encouraging me to study abroad. I thank also all my friends for spiritual support.

I thank for the financial support they provided: the Ministry of Education of Taiwan, the Air Force Office of Scientific Research, and the National Science Foundation.

Most importantly, I thank my wife, who supports me in every aspect of my life from physical to emotional to mental well-being. As a result, I am always healthy and energetic to start my research journey every day and finally almost reach a life milestone. Lastly, I thank my parents, younger brother, and relatives for spiritual support.

I-Te Lu
Pasadena, CA, USA
May 13, 2020

ABSTRACT

Crystallographic defects and impurities govern charge transport at low temperature, where the electron-defect (e-d) interactions limit the carrier mobility and manifest themselves in a wide range of phenomena of broad relevance in condensed matter physics. Theoretical treatments of e-d interactions have so far relied on heuristic approaches and analytic models. However, the band structure, electronic wave functions, and defect perturbation potential are far more complex in real materials than in these simplified models. First-principles calculations can provide atomistic details of the atomic and electronic structures of the material and make accurate predictions of their properties. Yet, *ab initio* calculations of e-d interactions are still in their infancy, mainly because they require large simulation cells and computationally expensive workflows. This thesis aims to overcome the open challenge of computing the e-d interactions and the associated e-d matrix elements, e-d relaxation times, and defect-limited mobility using first-principles methods. We develop an efficient first-principles method to compute the e-d matrix elements and apply it to neutral vacancy and interstitial defects in silicon. Using the new approach, we demonstrate systematic convergence of the e-d relaxation times with respect to supercell size, defect position, and Brillouin zone sampling. To speed up the e-d calculations, we formulate and implement an interpolation scheme to compute the e-d matrix elements using maximally-localized Wannier functions. We show for the first time fully *ab initio* calculations of the temperature dependent defect-limited carrier mobility and investigate its numerical convergence. To treat charged defects, we develop a different interpolation method and apply it to a charged point defect in silicon. We use this approach together with importance sampling integration to effectively compute the e-d relaxation times for charged defects. Finally, we provide technical details of the e-d routines and discuss their integration in the open source code PERTURBO developed in the Bernardi group. In summary, the methods developed in this thesis have laid a solid foundation for future ab initio e-d interaction calculations, which can be applied broadly to address materials design challenges in electronics, energy, and quantum technologies.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] I.-T. Lu, J.-J. Zhou, and M. Bernardi, Phys. Rev. Mater. **3**, 033804 (2019), DOI: https://doi.org/10.1103/PhysRevMaterials.3.033804,
I.-T. Lu derived the formula, wrote the electron-defect routines, carried out the calculations, analyzed the data, and wrote the manuscript.

[2] I.-T. Lu, J. Park, J.-J. Zhou, and M. Bernardi, Npj Comput. Mater. **6**, 17 (2020), DOI: https://doi.org/10.1038/s41524-020-0284-y,
I.-T. Lu derived the formula, implemented the formula in the electron-defect routines, analyzed and prepared the data, and wrote the manuscript.

[3] J.-J. Zhou, J. Park, I.-T. Lu, I. Maliyov, X. Tong, and M. Bernardi, arXiv:2002.02045 (2020), DOI: https://arxiv.org/abs/2002.02045,
I.-T. Lu participated in developing the PERTURBO code and prepared the user manual and part of the data.

[4] I.-T. Lu and M. Bernardi, Sci. Rep. **7**, 3403 (2017), DOI: https://doi.org/10.1038/s41598-017-01434-8,
I.-T. Lu carried out the calculations and wrote the manuscript.

TABLE OF CONTENTS

Acknowledgements iii
Abstract v
Published Content and Contributions vi
Table of Contents vii
List of Illustrations ix
List of Tables x
Chapter I: Introduction 1
1.1 Scope of the Thesis 1
1.2 Theoretical Tools for Electron-Defect Interaction 2
1.3 First-Principles Methods for Electron-Defect Interaction 5
1.4 Outline of the Thesis 9
Chapter II: Efficient First-Principles Method for Computing Electron-Defect Interactions 14
2.1 Introduction 14
2.2 Derivation of an Efficient Formalism for Electron-Defect Interaction 15
2.3 Semiclassical Boltzmann Equation within the Relaxation Time Approximation 18
2.4 Electron-Defect Interaction Workflow and Computational Details . . 18
2.5 Comparison between the All-Supercell Method and Our Approach . 20
2.6 *Ab Initio* Electron-Defect Relaxation Times 22
2.7 Rigorous Strategy for Converging Electron-Defect Relaxation Times 26
2.8 Carrier Mobility due to Neutral Defects in Silicon 28
2.9 Summary 30
Chapter III: *Ab Initio* Electron-Defect Interactions Using Wannier Functions . 33
3.1 Introduction 33
3.2 Derivation of a Fourier-Wannier Interpolation Scheme for Electron-Defect Matrix Elements 34
3.3 Wannier Interpolation Workflow and Computational Details 36
3.4 Validation of the Wannier-Fourier Interpolation Method for Electron-Defect Matrix Elements 38
3.5 Carrier Relaxation Time and Mobility Using the Wannier Interpolation Method 40
3.6 Speedup of the Wannier Interpolation Method Compared to the Direct Computation Method 43
3.7 Wannier Interpolation Method for the Brillouin Zone Summation in a Metal 43
3.8 Summary 46
Chapter IV: *Ab Initio* Electron-Defect Interactions for Ionized Impurities in Semiconductors 49

4.1 Introduction 49
4.2 Electron-Defect Matrix Elements for Ionized Impurities 50
4.3 Wannier Interpolation Method for Charged Defects 52
4.4 Workflow and Computation Details 55
4.5 Electron-Defect Matrix Elements for Charged Defects 55
4.6 Electron-Defect Relaxation Times due to Charged Defects 58
4.7 Electron Carrier Mobility as a Function of Doping Concentration in Silicon 60
4.8 Summary 62
Chapter V: The PERTURBO Open Source Code and its Electron-Defect Routines 65
5.1 Introduction 65
5.2 Boltzmann Transport Equation for Electron-Defect Scattering 66
5.3 Software to Compute Electron-Defect Matrix Elements for Neutral and Charged Defects 71
5.4 Parallelization of Electron-Defect Scattering Calculations 72
5.5 Summary 77
Chapter VI: Future Extension of *ab initio* electron-defect interactions 80
Appendix A: Elastic and Incoherent Electron-Defect Scattering Rate 85
Appendix B: Derivation of Local Part of Electron-Defect Matrix Elements . . . 87
Appendix C: Derivation of Nonlocal Part of Electron-Defect Matrix Elements 89
Appendix D: Relationship between DFT and WS supercells 91

LIST OF ILLUSTRATIONS

Number *Page*

1.1 Schematic of the Wannier interpolation procedure 8
2.1 Workflow for computing the electron-defect matrix elements, relaxation times, and defect-limited mobility 19
2.2 Absolute value of the local electron-defect matrix elements obtained from our approach and the all-supercell method 22
2.3 *Ab initio* electron-defect relaxation time for neutral vacancy 23
2.4 Convergence of the electron-defect RTs for vacancy defects in silicon 25
2.5 Defect-limited mobilities for neutral vacancy and interstitial defects in silicon 29
3.1 Workflow for interpolating the electron-defect matrix elements using Wannier functions 37
3.2 Absolute value of the electron-defect matrix elements along high-symmetry lines 39
3.3 Carrier relaxation times and hole mobility for vacancy defects in silicon 41
3.4 Relaxation times and the defect-limited resistivity for vacancy defects in copper 44
3.5 Scattering strength in copper and silicon for vacancy defects 45
4.1 Workflow for interpolating electron-defect matrix elements for ionized impurity scattering 54
4.2 Electron-defect matrix elements for the Yukawa potential 56
4.3 Relaxation times due to ionized impurity scattering 59
4.4 Electron mobility in Phosphorus-doped silicon as a function of doping concentration 61
5.1 Workflow for electron-defect calculations in PERTURBO 70
5.2 The simple concept of hybrid parallel programming 73
5.3 A portion of the electron-defect routine to read wave functions 75
D.1 Supercells and matrix elements 91

LIST OF TABLES

Number *Page*

3.1 The mean and maximum difference between the interpolated and directly computed electron-defect matrix elements, given for several coarse $\boldsymbol{k}_{\mathrm{c}}$-grids 40

Chapter 1

INTRODUCTION

1.1 Scope of the Thesis

Crystallographic defects and impurities in solid-state materials typically control charge transport at low temperatures and in highly disordered materials as a result of the interactions between defects and charge carriers [1, 2]. Defects can be neutral or charged, nonmagnetic or magnetic, point-like or extended [3]. Charge carriers can carry positive or negative charges and possess additional degrees of freedom such as spin and valley in semiconductors [4]. The interaction among these two parties, here and below referred to as electron-defect (e-d) interaction, manifests itself in a wide range of physical phenomena [1, 2] and technological applications. For example, charged defects can scatter charge carriers and reduce carrier mobility, a key figure of merit of materials for renewable energy technologies, high-power devices, and light-emitting diodes [5–12]. In addition, magnetic impurities or defects in the presence of spin-orbit coupling can rotate or flip the spin of the charge carriers, thus governing their spin relaxation, a quantity of key interest in spintronics and spin qubits [13–15], while line defects can filter electrons with opposite valleys in valleytronics [16]. Given the broad implications, understanding e-d interactions can provide physical insights and design rules for new materials for a wide range of technologies as well as help address open problems of fundamental relevance in condensed matter physics, such as localization and decoherence. Among relevant physical properties, the carrier mobility is actively investigated in computational materials theory, and this work focuses on predicting the defect-limited carrier mobility using first-principles computational methods. The e-d interactions are described and quantified by the e-d matrix elements, which encode the transition probability amplitude from an electronic state to another state due to scattering by the perturbation induced by the defects in the material. Computing the e-d matrix elements for real materials is a difficult task that requires knowledge of the electronic wave functions in the entire Brillouin zone (BZ) and the detailed value of the spatially dependent defect perturbation potential.

1.2 Theoretical Tools for Electron-Defect Interaction

The carrier mobility in semiconductors can be limited by ionized and neutral impurities. Ionized impurity scattering occurs when electrons are scattered by charged defects or ionized external impurities, a typical example of which are dopants in semiconductors. The scattering potential associated with the ionized defects in semiconductors can be approximated as a Yukawa potential,

$$V(\boldsymbol{r}) = \frac{Ze}{\epsilon_s r} e^{-q_0 r}, \tag{1.1}$$

where Z is the number of charges carried by the defect, e is the electron charge, ϵ_s is the dielectric constant of the material, and q_0 is the inverse of the screening length, which depends on temperature and free carrier concentration. For the last few decades, computing *e*-d scattering rates has relied heavily on simplified analytic approaches such as the Conwell-Weisskipf (CW) [17] and the Brooks-Herring (BH) formulas [18]. Both approaches use the Yukawa potential together with simplified band structure and wave functions (usually, plane waves) to compute the *e*-d scattering matrix elements. Using these *e*-d matrix elements, electron scattering rates or scattering cross sections can be computed within the lowest-order of perturbation theory (so-called Born approximation) by integrating over final electronic states with the same energy as the initial state, assuming an elastic scattering process and a parabolic band structure with an empirical value of the effective mass. The CW approach employs the mean distance between impurities as a cutoff for the Yukawa potential, which implies that the electrons can only be scattered over this distance. In contrast, in the BH approach, electrons can be scattered by impurities over any distance, and the scattering events are assumed to be two-body collisions. Both approaches are widely used for modeling ionized impurity scattering and interpreting low-temperature mobility measurements. However, due to their severe approximations, several corrections have been proposed and developed for ionized impurities in semiconductors [19], although approaching the problem from the viewpoint of atomistic *ab initio* calculations is still an open research frontier.

Neutral impurities can also be scattering sources for charge carriers. The scattering cross sections of neutral impurities are typically computed using a hydrogenic or a square-well potential model [20]. The scattering cross section for an electron scattered by a neutral hydrogen atom is typically computed using a variational method [21]; Erginsoy applied the approach to neutral impurity scattering in semiconductors [22]. The square-well potential model first solves the quantum mechanical problem of electrons scattering off a square-well potential, whose solutions can

be categorized into three cases. The first is called hard-sphere scattering, whereby electrons are scattered by a short-range hard sphere potential, leading to a negligibly small scattering cross section. The second case is associated with the so-called Ramsauer-Townsend effect, in which a minimum appears in the scattering cross section at a particular electron kinetic energy; its scattering cross section, similar to the hard-sphere scattering case, is weak. The last case is resonant scattering, which has a larger scattering cross section peaked at the resonance energy. If the potential of the neural impurity is strong enough, bound states may form so the resonant scattering cross section further increases. This scenario is described by the Sclar formula [23], which predicts a scattering cross section with a similar order of magnitude as the Erginsoy formula for a hydrogenic model, but with a different energy dependence.

The Yukawa potential for ionized impurity scattering cannot extend to the impurity core. To deal with the potential at the core, the simplest approach is to include a square-well potential, called central-cell contribution, in the Yukawa potential [24]. The scattering cross sections in this model can be separated into three terms: one is due to ionized impurity scattering, the second to square-well potential scattering, and the other to the interference from the first two. For the second term, only resonant scattering is important as in the neutral impurity scattering case. Using this model, one can predict the curious result that negatively charged impurities have a weaker electron scattering strength than positively charged impurities because no resonance occurs [24].

Impurities in metals experience strong screening from the valence electrons; a simple Yukawa potential is typically inadequate to describe the defect perturbation potential in metals [25]. The carrier mobility or electrical resistivity in metals are typically computed using the phase shifts of the partial waves around the impurities [25]. The phase shifts are determined self-consistently, which we discuss below, to satisfy the Friedel sum rule to conserve the extra charge around the impurity. The starting point of the self-consistent calculation is a given potential, whose detailed form is unimportant, with a fitting parameter – e.g., the potential depth in the square-well model. This parametrized potential is used to compute the phase shifts, which in turn are used to check if the Friedel sum is satisfied; the fitting parameter is thus adjusted until the Friedel sum is satisfied. The resulting potential is used to obtain the phase-shifts and from them to compute the residual low-temperature resistivity in metals.

A more advanced approach to treat impurity scattering is using Green's functions in the many-body perturbation theory framework [26–29]. The Green function of an electron can be thought of as the transition probability amplitude for an electron to propagate from one state to another in the presence of an external perturbation. The Green function of a pristine system, usually a non-interacting electron system, can be analytically computed using plane waves. This bare Green function serves as a starting point to obtain an interacting (or "dressed") Green function that includes the interaction (called perturbation here) among electrons or between electrons and defects or phonons. The new Green function can be computed using a series of Feynman diagrams. Many approximations have been devised, such as resumming part of the diagrams or truncating higher orders; as a result, the Green function approach has become widely popular for treating interacting systems in condensed matter physics. The interacting Green functions can be used to compute a wide range of material properties such as the total energy, density of states, spectral function, among others [26, 28, 30]. For the e-d interactions, the defect perturbation potential determines the interacting Green function, for which several approximations and diagrams have been explored and computed, among which the lowest-order Born approximation and the T-matrix formalism have been widely employed [26–29]. The Green function for e-d interactions can be used to compute the carrier mobility. In the lowest order theory, the width of the spectral function (imaginary part of the Green function) is the scattering rate in energy units, and has an expression equivalent to Fermi's golden rule. The inverse of this scattering rate is the e-d relaxation time, which plays a key role in understanding electron dynamics in materials. Within the semiclassical treatment of transport of the Boltzmann transport equation (BTE), one can use the e-d scattering rates to compute the collision term in the BTE, and from it the mobility either in the relaxation time approximation or from a full solution of the linearized BTE [26, 31]. Another approach to compute the carrier mobility is to use Kubo formula, which uses directly the Green's function [29, 32]. Also noteworthy is the fact that Feynman diagrams expressing the correlation of two current operators for e-d interactions have enabled studies of quantum interference effects such as weak localization and universal fluctuations [1, 2, 32].

Though we primarily focus on atomic defects, similar approaches for computing scattering rates and carrier mobility can be applied to dislocations and interface scattering or grain boundary scattering [25]. While many theoretical approaches have been formulated for e-d interactions, the vast majority of the work done in the field uses semiempirical approaches and simplified analytic models. However,

the defect perturbation potential, band structure, and electronic functions in real materials are far more complex than the simplified description in these models. The atomistic details are very important because the same defect − for example, a vacancy or dislocation − in two different materials could possess entirely different properties and interactions with the charge carriers. The atomistic details of the short-range potential and the atomic relaxation around the defect play a key role and cannot be captured with simplified models. These observations motivate employing atomistic first-principles calculations, such as those based on density functional theory and related methods, to treat the e-d interactions and remove the ambiguities and crude approximations of approximate models.

1.3 First-Principles Methods for Electron-Defect Interaction

Density functional theory (DFT) is a mean-field theory that describes the many-electron system and its interactions using the electron density. It has become the starting point for nearly all computational theories of materials due to its wide availability and favorable trade-off between computational cost and accuracy. Central to DFT is the single-electron Kohn-Sham (KS) Hamiltonian, which depends on the electron density and can be solved self-consistently. The KS Hamiltonian $\hat{H}_{\mathrm{KS}}$ employed in DFT with the pseudopotential method includes the electronic kinetic energy, the Hartree potential $V_{\mathrm{H}}(\boldsymbol{r})$ accounting for the repulsion between the electron and the electron density distribution, the exchange-correlation potential $V_{\mathrm{XC}}(\boldsymbol{r})$ encoding electron exchange and correlation interactions, and the pseudopotentials $\hat{V}_{\mathrm{pp}}$ for the core-valence electron interactions:

$$\hat{H}_{\mathrm{KS}} = -\frac{1}{2}\nabla^2 + V_{\mathrm{H}}(\boldsymbol{r}) + V_{\mathrm{XC}}(\boldsymbol{r}) + \hat{V}_{\mathrm{pp}}. \tag{1.2}$$

After solving the KS equation self-consistently, the electronic energies and KS wavefunctions can be employed to compute the total energy or, as is relevant in this thesis, be used as a starting point for developing perturbation theories to investigate electron dynamics. DFT and related approaches are known as first-principles or *ab initio* methods because no experimental input is required other than knowledge of the crystal structure and atomic positions. Due to its computational efficiency and accuracy, DFT has been applied to a wide range of materials such as metals, semiconductors, and insulators, and recently to correlated and topological materials [33–37].

The DFT electronic ground state is also the starting point for lattice dynamics and phonon calculations. DFT can treat not only bulk materials but also heterogeneous

systems that contain interfaces, defects, or impurities. The defect-related physical quantities that can be computed using DFT include the defect relaxed atomic structure, formation energy, single-particle defect energy level, among others [15, 38–41]. Even though DFT has become a standard tool to explore systems containing defects, many defect-related physical properties and processes remain challenging to compute from first-principles, both due to computational cost and to lack of a theoretical formulation amenable to computation. Computational cost is a key barrier for defect calculations in DFT. When using a plane-wave basis, as is common in many DFT codes, the defect needs to be placed in a supercell large enough to avoid spurious interactions with the image cells due to the periodic boundary conditions. As a result, defect calculations may need simulation cells with hundreds of atoms treated explicitly, making the calculations very computationally demanding. Examples include carrier capture at defects [42–44], charged impurities in two-dimensional materials [45], or extended defects such as dislocations and grain boundaries [46, 47], which require a large number of atoms in the simulation cell.

Over the last few years, it has become possible to compute transport properties in materials entirely from first principles [48–54]. A lead example of these advances is the calculation of electron-phonon (e-ph) interactions and scattering processes, and from them calculations of the mobility, which is typically controlled by e-ph scattering near room temperature. The phonon-limited carrier mobility as a function of temperature has been computed for a wide range of semiconductors and insulators, providing predictions in excellent agreement with experiments and providing unprecedented microscopic insight into the charge transport mechanisms [54–58]. An important technical advance that has made these calculations possible are approaches to interpolate the e-ph matrix elements using a localized basis set such as Wannier functions [48, 59–61] or atomic orbitals [62]. Because transport calculations require integrating over small energy windows and correspondingly small regions in the BZ, the ability to compute scattering processes on arbitrarily fine grids has been instrumental for advances in *ab initio* transport calculations [48, 53].

While near room temperature the carrier mobility is controlled by phonons, at low temperature the mobility is typically controlled by defect scattering. An example of a first-principles method that can compute the defect-limited carrier mobility is the Korringa-Kohn-Rostoker (KKR) Green function method [63], which is particularly suitable for metals and alloys. Several properties have been computed with this approach, including the magnetoresistance due to impurity scattering [64], the

residual resistivity of metals and alloys [65], spin relaxation [66–69], and the spin Hall effect [70–72]. However, the KKR formalism requires all-electron codes, which is highly inconvenient. In addition, *e*-ph interactions cannot be easily incorporated in the KKR framework. To combine the current first-principles framework for *e*-ph interactions with *e*-d interactions to study charge transport over a wide temperature range, a key requirement is to treat *e*-d and *e*-ph interactions on the same footing [73], computing matrix elements for the interaction, and then using perturbation theory to obtain relaxation times, carrier mobilities from the BTE, or even spectral function and quantum transport in the Green-Kubo formalism [52].

The approach taken in this thesis is to compute the matrix elements for the *ab initio* *e*-d interaction using KS electronic wave functions and obtain the defect perturbation potential as the difference between the KS Hamiltonians [see Eq. (1.2)] of the pristine system and the system containing a defect. While the brute-force approach to computing *e*-d matrix elements would entail using large supercells to obtain the wave functions in the presence of a defect [46, 74, 75], this work develops an elegant approach to circumvent this issue and use only the wave functions of a unit cell of the material hosting the defect. Also challenging is computing the *e*-d relaxation times and defect-limited carrier mobility, each of which requires a summation over ultra-fine BZ grids. These noteworthy challenges have hampered progress on *ab initio* *e*-d interactions and the related carrier dynamics calculations. This thesis successfully overcomes these main technical challenges.

Although Bloch states are a suitable representation and the eigenstates of the Hamiltonian used in electronic structure calculations, other representations are available such as Wannier functions (WFs). WFs can be constructed and transformed from Bloch states via a unitary transformation; however, the implementation of WFs in *ab initio* electronic calculations had once progressed slowly because WFs are strongly nonunique, which results from the phase indeterminacy of the Bloch states at every wave vector $\boldsymbol{k}$. To solve the phase indeterminacy, maximal-localization criteria were proposed and developed to identify a unique set of WFs by minimizing the second-moment spread of the WFs [76]. The resulting WFs are called maximally localized WFs referred to as WFs here and below. The applications of WFs range from the description of chemical and dielectric properties of complex materials, to the construction of effective Hamiltonians for strongly correlated systems, to other applications that are out of the traditional realm of electronic structure calculations such as phonons and photonic crystals [77].

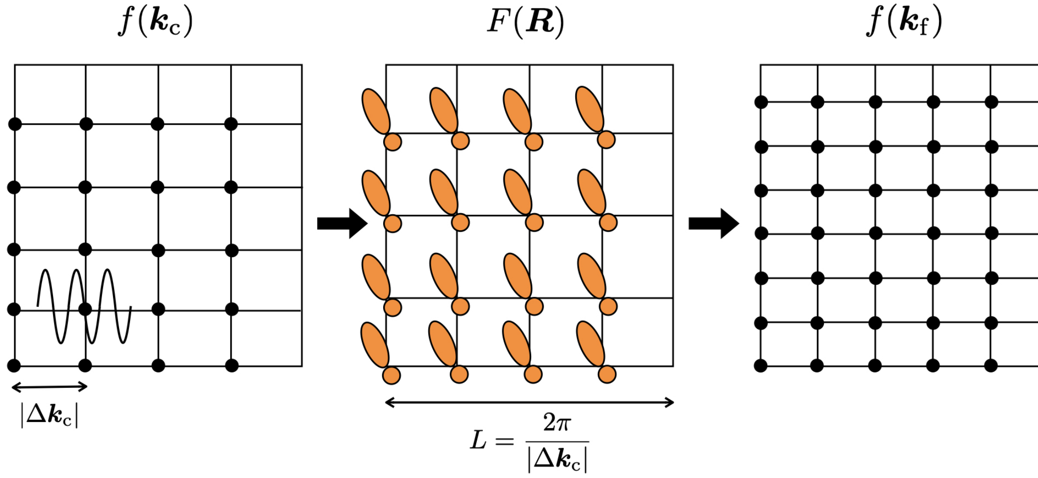

Figure 1.1: Schematic of the Wannier interpolation procedure.

Among the WF techniques used in first principle calculations is the Wannier interpolation procedure, which is relevant to this work and illustrated in Figure 1.1. A physical quantity $f(\boldsymbol{k}_\mathrm{c})$ is computed from the Bloch states on a uniform coarse BZ $\boldsymbol{k}_\mathrm{c}$-grid using first-principles methods. The Bloch states are then transformed into the WFs, and the physical quantity $f(\boldsymbol{k}_\mathrm{c})$ is transformed into $F(\boldsymbol{R})$ in the Wannier representation accordingly. Using the Wannier representation $F(\boldsymbol{R})$, we can accurately interpolate the $f(\boldsymbol{k}_\mathrm{f})$ for any arbitrary wave vector $\boldsymbol{k}_\mathrm{f}$ on a fine BZ $\boldsymbol{k}_\mathrm{f}$-grid via the inverse transformation. The success of the Wannier interpolation procedure lies in the spatial decay of $F(\boldsymbol{R})$ within the linear dimension $L = 2\pi/|\Delta\boldsymbol{k}_\mathrm{c}|$ of the equivalent supercell where $\Delta\boldsymbol{k}_\mathrm{c}$ is the mesh spacing of the coarse BZ $\boldsymbol{k}_\mathrm{c}$-grid. The denser the coarse BZ $\boldsymbol{k}_\mathrm{c}$-grid is, the more accurate the interpolated results are. However, a coarser BZ $\boldsymbol{k}_\mathrm{c}$-grid is desirable to save computational costs; the interpolated results should thus be checked and converged with respect to the coarse BZ $\boldsymbol{k}_\mathrm{c}$-grid size and compared with the directly computed results. Due to its efficiency and accuracy, the Wannier interpolation procedure has been used in many applications such as interpolating band structures and other physical quantities (e.g., e-ph interactions) that involve demanding BZ integrals [59, 78]. In this thesis, we will also develop a Wannier interpolation method to deal with computationally costly calculations for e-d interactions.

1.4 Outline of the Thesis

In this thesis, we aim to address the open challenge of computing the e-d interactions and the associated e-d matrix elements, relaxation times, and defect-limited mobility using first-principles methods. Our goal is to lay a solid foundation for future *ab initio* e-d interaction calculations.

In Chapter 2, we develop a new efficient method to compute the e-d matrix elements. The new method employs quantities such as electronic wave functions, defect perturbation potentials, and electronic band structure obtained entirely from first principles. This approach is a major departure from the empirical methods described above. It also dramatically improves over early attempts to compute e-d interactions from first principles using electronic wave functions from supercell calculations. By using only the wave functions of the primitive cell, we speed up the matrix element calculations by orders of magnitude while also reducing memory usage. We apply our approach to neutral vacancy and interstitial defects in silicon. The new approach can reproduce the e-d matrix elements computed from the all-supercell method. We further demonstrate how to converge the e-d RTs with respect to supercell size, defect position, and the number of $\boldsymbol{k}$-points in the BZ, which have not been rigorously shown before due to the prohibitive computational cost.

In Chapter 3, we develop an interpolation scheme to compute e-d matrix elements using Wannier functions to speed up the e-d calculations. We apply the interpolation method to a neutral vacancy in silicon. The interpolated e-d matrix elements using the interpolation method can reproduce the directly computed results. We also demonstrate for the first time how to systematically converge the temperature dependent defect-limited carrier mobility as a function of BZ grid size. Moreover, we apply the interpolation method to a metal, copper, which requires denser BZ grids, to directly map the e-d relaxation times on its Fermi surface.

In Chapter 4, we develop a different interpolation method for charged defects, and apply it to a point-charged defect in silicon. We apply the method to interpolate the e-d matrix elements for an ionized impurity modeled as a source of Yukawa potential. To converge the e-d relaxation times, we use a random BZ grid generated according to a Cauchy distribution. Using the interpolation method and importance sampling integration, the e-d relaxation times for charged defects are shown to converge when about 1 million scattering channels are included for each electronic state.

In Chapter 5, we provide the technical details of the e-d routines developed in

this work, as well as their integration in the open source software, PERTURBO developed by the Bernardi group, which has so far focused mainly on *ab initio* *e*-ph calculations [54]. In Chapter 6, we summarize the key results and achievements and outlines future extensions of our work and new research directions enabled by the *e*-d methods developed in this thesis.

References

[1] P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 287 (1985).

[2] S. Datta, *Electronic transport in mesoscopic systems* (Cambridge University Press, 1997).

[3] R. J. Tilley, *Defects in solids*, Vol. 4 (John Wiley & Sons, 2008).

[4] J. R. Schaibley, H. Yu, G. Clark, P. Rivera, J. S. Ross, K. L. Seyler, W. Yao, and X. Xu, Nat. Rev. Mater. **1**, 16055 (2016).

[5] S. Kim, A. Konar, W.-S. Hwang, J. H. Lee, J. Lee, J. Yang, C. Jung, H. Kim, J.-B. Yoo, J.-Y. Choi, Y. W. Jin, S. Y. Lee, D. Jena, W. Choi, and K. Kim, Nat. Commun. **3**, 1011 (2012).

[6] B. Radisavljevic and A. Kis, Nat. Mater. **12**, 815 (2013).

[7] P. Sun, B. Wei, J. Zhang, J. M. Tomczak, A. M. Strydom, M. Søndergaard, B. B. Iversen, and F. Steglich, Nat. Commun. **6**, 7475 (2015).

[8] T. M. Brenner, D. A. Egger, L. Kronik, G. Hodes, and D. Cahen, Nat. Rev. Mater. **1**, 15007 (2016).

[9] S.-L. Li, K. Tsukagoshi, E. Orgiu, and P. Samorì, Chem. Soc. Rev. **45**, 118 (2016).

[10] L. M. Herz, ACS Energy Lett. **2**, 1539 (2017).

[11] J. Mao, J. Shuai, S. Song, Y. Wu, R. Dally, J. Zhou, Z. Liu, J. Sun, Q. Zhang, C. dela Cruz, S. Wilson, Y. Pei, D. J. Singh, G. Chen, C.-W. Chu, and Z. Ren, Proc. Natl. Acad. Sci. U.S.A. **114**, 10548 (2017).

[12] H. Peelaers and C. G. Van de Walle, Phys. Rev. B **100**, 081202 (R) (2019).

[13] I. Z. Žutić, J. Fabian, and S. Das Sarma, Rev. Mod. Phys. **76**, 323 (2004).

[14] R. Jansen, Nat. Mater. **11**, 400 (2012).

[15] V. Ivády, I. A. Abrikosov, and A. Gali, Npj Comput. Mater. **4**, 76 (2018).

[16] D. Gunlycke and C. T. White, Phys. Rev. Lett. **106**, 136806 (2011).

[17] E. Conwell and V. F. Weisskopf, Phys. Rev. **77**, 388 (1950).

[18] H. Brooks, in *Advances in electronics and electron physics*, Vol. 7, edited by L. Marton, 1st ed. (Academic Press, New York, 1955), pp. 85–182.

[19] D. Chattopadhyay and H. J. Queisser, Rev. Mod. Phys. **53**, 745 (1981).

[20] B. K. Ridley, *Quantum processes in semiconductors* (Oxford University Press, 2013).

[21] H. S. W. Massey and B. L. Moiseiwitsch, Phys. Rev. **78**, 180 (1950).

[22] C. Erginsoy, Phys. Rev. **79**, 1013 (1950).

[23] N. Sclar, Phys. Rev. **104**, 1559 (1956).

[24] H. M. El-Ghanem and B. K. Ridley, J. Phys. C **13**, 2041 (1980).

[25] V. F. Gantmakher and Y. B. Levinson, *Carrier scattering in metals and semiconductors* (Elsevier, 2012).

[26] G. D. Mahan, *Many-particle physics* (Springer US, 2000).

[27] H. Bruus and K. Flensberg, *Many-body quantum theory in condensed matter physics: An introduction* (Oxford University Press, 2004).

[28] E. N. Economou, *Green's functions in quantum physics* (Springer, 1983).

[29] J. Rammer, *Quantum transport theory* (CRC Press, 2018).

[30] A. L. Fetter and J. D. Walecka, *Quantum theory of many-particle systems* (Courier Corporation, 2012).

[31] C. Jacoboni, *Theory of electron transport in semiconductors: A pathway from elementary physics to nonequilibrium green functions*, Vol. 165 (Springer Science & Business Media, 2010).

[32] E. Akkermans and G. Montambaux, *Mesoscopic physics of electrons and photons* (Cambridge University Press, 2007).

[33] R. M. Martin, *Electronic structure: Basic theory and practical methods* (Cambridge University Press, 2004).

[34] W. Zhang, R. Yu, H.-J. Zhang, X. Dai, and Z. Fang, New J. Phys. **12**, 065013 (2010).

[35] A. Jain, Y. Shin, and K. A. Persson, Nat. Rev. Mater. **1**, 15004 (2016).

[36] H. Weng, X. Dai, and Z. Fang, J. Phys. Condens. Matter **28**, 303001 (2016).

[37] J. Zou, Z. He, and G. Xu, Npj Comput. Mater. **5**, 96 (2019).

[38] C. G. Van de Walle and J. Neugebauer, J. Appl. Phys. **95**, 3851 (2004).

[39] A. Alkauskas, P. Deák, J. Neugebauer, A. Pasquarello, and C. G. Van de Walle, *Advanced calculations for defects in materials: Electronic structure methods* (John Wiley & Sons, 2011).

[40] C. Freysoldt, B. Grabowski, T. Hickel, J. Neugebauer, G. Kresse, A. Janotti, and C. G. Van de Walle, Rev. Mod. Phys. **86**, 253 (2014).

[41] J. Spitaler and S. K. Estreicher, Front. Mater. **5**, 70 (2018).

[42] L. Shi and L.-W. Wang, Phys. Rev. Lett. **109**, 245501 (2012).

[43] A. Alkauskas, Q. Yan, and C. G. Van de Walle, Phys. Rev. B **90**, 075202 (2014).

[44] S. Kim, S. N. Hood, and A. Walsh, Phys. Rev. B **100**, 041202 (R) (2019).

[45] J. Xiao, K. Yang, D. Guo, T. Shen, H.-X. Deng, S.-S. Li, J.-W. Luo, and S.-H. Wei, Phys. Rev. B **101**, 165306 (2020).

[46] M. H. Evans, X.-G. Zhang, J. D. Joannopoulos, and S. T. Pantelides, Phys. Rev. Lett. **95**, 106802 (2005).

[47] J. J. Kuo, S. D. Kang, K. Imasato, H. Tamaki, S. Ohno, T. Kanno, and G. J. Snyder, Energy Environ. Sci. **11**, 429 (2018).

[48] J. J. Zhou and M. Bernardi, Phys. Rev. B **94**, 201201(R) (2016).

[49] V. A. Jhalani, J.-J. Zhou, and M. Bernardi, Nano Lett. **17**, 5012 (2017).

[50] J.-J. Zhou, O. Hellman, and M. Bernardi, Phys. Rev. Lett. **121**, 226603 (2018).

[51] N.-E. Lee, J.-J. Zhou, L. A. Agapito, and M. Bernardi, Phys. Rev. B **97**, 115203 (2018).

[52] J.-J. Zhou and M. Bernardi, Phys. Rev. Res. **1**, 033138 (2019).

[53] J. Park, J.-J. Zhou, and M. Bernardi, Phys. Rev. B **101**, 045202 (2020).

[54] J.-J. Zhou, J. Park, I.-T. Lu, I. Maliyov, X. Tong, and M. Bernardi, arXiv:2002.02045 (2020).

[55] K. Kaasbjerg, K. S. Thygesen, and K. W. Jacobsen, Phys. Rev. B **85**, 115317 (2012).

[56] W. Li, Phys. Rev. B **92**, 75405 (2015).

[57] J. I. Mustafa, M. Bernardi, J. B. Neaton, and S. G. Louie, Phys. Rev. B **94**, 155105 (2016).

[58] S. Poncé, W. Li, S. Reichardt, and F. Giustino, Rep. Prog. Phys. **83**, 036501 (2020).

[59] F. Giustino, M. Cohen, and S. Louie, Phys. Rev. B **76**, 165108 (2007).

[60] C. Verdi and F. Giustino, Phys. Rev. Lett. **115**, 176401 (2015).

[61] J. Sjakste, N. Vast, M. Calandra, and F. Mauri, Phys. Rev. B **92**, 054307 (2015).

[62] L. A. Agapito and M. Bernardi, Phys. Rev. B **97**, 235146 (2018).

[63] H. Ebert, D. Köedderitzsch, and J. Minár, Rep. Prog. Phys. **74**, 096501 (2011).

[64] P. Zahn, I. Mertig, M. Richter, and H. Eschrig, Phys. Rev. Lett. **75**, 2996 (1995).

[65] I. Mertig, Rep. Prog. Phys. **62**, 237 (1999).

[66] D. V. Fedorov, P. Zahn, M. Gradhand, and I. Mertig, Phys. Rev. B **77**, 092406 (2008).

[67] M. Gradhand, D. V. Fedorov, P. Zahn, and I. Mertig, Phys. Rev. B **81**, 020403(R) (2010).

[68] D. V. Fedorov, M. Gradhand, S. Ostanin, I. V. Maznichenko, A. Ernst, J. Fabian, and I. Mertig, Phys. Rev. Lett. **110**, 156602 (2013).

[69] N. H. Long, P. Mavropoulos, B. Zimmermann, S. Heers, D. S. Bauer, S. Blügel, and Y. Mokrousov, Phys. Rev. B **87**, 224420 (2013).

[70] M. Gradhand, D. V. Fedorov, P. Zahn, and I. Mertig, Phys. Rev. B **81**, 245109 (2010).

[71] M. Gradhand, D. V. Fedorov, P. Zahn, and I. Mertig, Phys. Rev. Lett. **104**, 186403 (2010).

[72] A. Hönemann, C. Herschbach, D. V. Fedorov, M. Gradhand, and I. Mertig, Phys. Rev. B **99**, 024420 (2019).

[73] M. Bernardi, Eur. Phys. J. B **89**, 239 (2016).

[74] O. D. Restrepo, K. Varga, and S. T. Pantelides, Appl. Phys. Lett. **94**, 212103 (2009).

[75] V. Lordi, P. Erhart, and D. Åberg, Phys. Rev. B **81**, 235204 (2010).

[76] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997).

[77] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, and D. Vanderbilt, Rev. Mod. Phys **84**, 1419 (2012).

[78] I. Souza, N. Marzari, and D. Vanderbilt, Phys. Rev. B **65**, 035109 (2001).

Chapter 2

EFFICIENT FIRST-PRINCIPLES METHOD FOR COMPUTING ELECTRON-DEFECT INTERACTIONS

2.1 Introduction

Defects in materials can scatter or capture charge carriers. They control carrier dynamics at low temperature, where phonons are frozen out, or even at room temperature in the presence of strong disorder. Calculations of *e*-d interactions have so far relied almost exclusively on empirical models [1]. Due to their severe approximations, empirical models are not reliable to compute *e*-d interactions, and their predictions can be qualitatively and quantitatively incorrect. First-principles approaches are therefore desirable to treat *e*-d interactions. *Ab initio* *e*-d calculations can take advantage of existing tools developed for electron-phonon (*e*-ph) interactions, for which accurate calculations of relaxation times (RTs) [2–5], matrix elements and their interpolation [6], and phonon-limited carrier mobility [7–11] have been recently developed. These and other workflows developed for *e*-ph calculations can be generalized to treat *e*-d interactions.

However, there are open challenges specific to *e*-d calculations that currently prevent their broad applicability. First-principles *e*-d calculations need large supercells to obtain the electron wavefunctions and defect perturbation potentials; they additionally involve computationally costly *e*-d matrix elements, and require systematic convergence of the RTs with respect to supercell size and Brillouin zone (BZ) grids. The cost of computing *e*-d matrix elements using the current existent all-supercell method is prohibitively large due to the use of electron wavefunctions from large supercells.

This chapter aims to address these open challenges by showing a method to carry out *ab initio* *e*-d calculations that are more affordable and broadly applicable. We develop a new efficient formalism for *e*-d matrix elements to overcome the challenges that have so far limited the broad application of *ab initio* *e*-d interactions. Our approach does not require the wavefunctions of large supercells, which dramatically speeds up the calculations. Using our approach, we can compute and converge the *e*-d matrix elements and the associated RTs and defect-limited carrier mobility. We apply our method to study *e*-d interactions in silicon in two cases, neutral

vacancy and interstitial defects, for which we compute and converge the e-d RTs as a function of energy and the carrier mobility as a function of temperature below 150 K. The results show that, contrary to conventional wisdom, the RTs depend strongly on carrier energy and defect type, and the defect-limited mobility for neutral defects depends on temperature. Our results provide new microscopic insight into e-d scattering, and our approach, together with its future extensions, can uncover new defect physics in materials and devices for electronics, energy, and quantum technologies.

2.2 Derivation of an Efficient Formalism for Electron-Defect Interaction

In the following, we work under the assumption that the defects are neutral (non-charged) and that the e-d scattering events are independent, uncorrelated, and elastic. The e-d scattering rate $\Gamma_{n\boldsymbol{k}}$ (and its inverse, the RT, $\tau_{n\boldsymbol{k}} = \Gamma_{n\boldsymbol{k}}^{-1}$) for a Bloch state $|n\boldsymbol{k}\rangle$, where n is the band index and $\boldsymbol{k}$ the crystal momentum, is computed using the lowest-order Born approximation (see Appendix A):

$$\Gamma_{n\boldsymbol{k}} = \frac{2\pi}{\hbar}\frac{n_{\mathrm{at}}C_{\mathrm{d}}}{N_{\boldsymbol{k}'}}\sum_{m\boldsymbol{k}'}|M_{mn}(\boldsymbol{k}',\boldsymbol{k})|^2\,\delta\,(\varepsilon_{m\boldsymbol{k}'}-\varepsilon_{n\boldsymbol{k}}), \tag{2.1}$$

where n_{at} is the number of atoms in a primitive cell, C_{d} the (dimensionless) defect atomic concentration (the number of defects divided by total number of atoms), $N_{\boldsymbol{k}'}$ the number of BZ $\boldsymbol{k}'$-points used in the sum, and $\varepsilon_{n\boldsymbol{k}}$ the unperturbed energy of the state $|n\boldsymbol{k}\rangle$ from the primitive cell. The e-d matrix elements $M_{mn}(\boldsymbol{k}',\boldsymbol{k})$ encode the probability amplitude for scattering from the Bloch state $|n\boldsymbol{k}\rangle$ to $|m\boldsymbol{k}'\rangle$ due to the perturbation potential $\Delta V_{\mathrm{e-d}}$ from a defect:

$$M_{mn}(\boldsymbol{k}',\boldsymbol{k}) = \langle m\boldsymbol{k}'|\Delta V_{\mathrm{e-d}}|n\boldsymbol{k}\rangle\,. \tag{2.2}$$

Within DFT [12], the e-d perturbation can be computed as the difference between the Kohn-Sham (KS) potentials V_{KS} of a defect-containing supercell and a pristine supercell with no defect, namely, $\Delta V_{\mathrm{e-d}} = V_{\mathrm{KS}}^{(\mathrm{d})} - V_{\mathrm{KS}}^{(\mathrm{p})}$. Here and below, we use superscripts (d) and (p) to denote the defect-containing and pristine systems, respectively. The KS potential can be written as a sum of local and nonlocal parts [13], when using norm-conserving pseudopotentials in the Kleinman-Bylander (KB) form [14] for the contribution from the nuclei and the core electrons,

$$V_{\mathrm{KS}} = V_{\mathrm{L}}(\boldsymbol{r}) + \hat{V}_{\mathrm{NL}}. \tag{2.3}$$

The local potential $V_\mathrm{L}(\boldsymbol{r})$ comprises the Hartree and the exchange-correlation potentials plus the local part of the pseudopotentials,

$$V_\mathrm{L}(\boldsymbol{r}) = V_\mathrm{H}(\boldsymbol{r}) + V_\mathrm{XC}(\boldsymbol{r}) + V_\mathrm{pp}(\boldsymbol{r}). \tag{2.4}$$

The nonlocal potential $\hat{V}_\mathrm{NL}$ is defined as a sum over all atoms in the supercell of KB projectors $|\beta_i^{(s)}\rangle$, each localized in the core region of the s-th atom:

$$\hat{V}_\mathrm{NL} = \sum_{s=1} \sum_{ij} D_{ij}^{(s)} \, |\beta_i^{(s)}\rangle \, \langle\beta_j^{(s)}| \,, \tag{2.5}$$

where i and j are orbital angular momentum quantum numbers, and $D_{ij}^{(s)}$ are KB coefficients [13]. The local and the nonlocal potentials of the KS potentials can be computed in the pristine and the defect-containing supercells using standard DFT calculations. Accordingly, we can separate the e-d matrix elements into a local and a nonlocal part,

$$M_{mn}(\boldsymbol{k}',\boldsymbol{k}) = M_{mn}^\mathrm{L}(\boldsymbol{k}',\boldsymbol{k}) + M_{mn}^\mathrm{NL}(\boldsymbol{k}',\boldsymbol{k})\,, \tag{2.6}$$

each due to the respective defect perturbation, $\Delta V_\mathrm{L}(\boldsymbol{r}) = V_L^{(\mathrm{d})}(\boldsymbol{r}) - V_L^{(\mathrm{p})}(\boldsymbol{r})$ for the local and $\Delta\hat{V}_\mathrm{NL} = \hat{V}_\mathrm{NL}^{(\mathrm{d})} - \hat{V}_\mathrm{NL}^{(\mathrm{p})}$ for the nonlocal part:

$$\begin{aligned} M_{mn}^\mathrm{L}(\boldsymbol{k}',\boldsymbol{k}) &= \langle m\boldsymbol{k}'|\Delta V_\mathrm{L}(\boldsymbol{r})|n\boldsymbol{k}\rangle \\ M_{mn}^\mathrm{NL}(\boldsymbol{k}',\boldsymbol{k}) &= \langle m\boldsymbol{k}'|\Delta\hat{V}_\mathrm{NL}|n\boldsymbol{k}\rangle\,. \end{aligned} \tag{2.7}$$

In the all-supercell method, the defect perturbation potential and the unperturbed Bloch states are all obtained from the supercells, which are computationally costly.

Here we develop a new approach to reduce the computational cost of the local and nonlocal matrix elements. We Fourier-transform the local perturbation potential $\Delta V_\mathrm{L}(\boldsymbol{r})$ and compute the local matrix elements as (see Appendix B)

$$M_{mn}^\mathrm{L}(\boldsymbol{k}',\boldsymbol{k}) = \sum_{\boldsymbol{G}} \Delta\widetilde{V}_\mathrm{L}(\boldsymbol{k}'-\boldsymbol{k}-\boldsymbol{G})\,\langle u_{m\boldsymbol{k}'}|e^{-i\boldsymbol{G}\cdot\boldsymbol{r}}|u_{n\boldsymbol{k}}\rangle_\mathrm{uc}\,, \tag{2.8}$$

where $u_{n\boldsymbol{k}}(\boldsymbol{r})$ is the periodic part of the Bloch wavefunction (normalized in a primitive cell with volume Ω_uc), $\boldsymbol{G}$ are reciprocal lattice vectors of the primitive cell, and $\Delta\widetilde{V}_\mathrm{L}$ are Fourier coefficients of the local defect perturbation potential (computed in a supercell with volume Ω_sup):

$$\Delta\widetilde{V}_\mathrm{L}(\boldsymbol{q}) = \frac{1}{\Omega_\mathrm{uc}} \int_{\Omega_\mathrm{sup}} d\boldsymbol{r}\;\Delta V_\mathrm{L}(\boldsymbol{r})\, e^{-i\boldsymbol{q}\cdot\boldsymbol{r}}, \tag{2.9}$$

where $\boldsymbol{q} = \boldsymbol{k}' - \boldsymbol{k} - \boldsymbol{G}$ is the transferred momentum in the $|n\boldsymbol{k}\rangle \rightarrow |m\boldsymbol{k}'\rangle$ scattering process. Through Eq. (2.8), we effectively separate the local matrix element calculation into two independent parts, the plane-wave matrix elements $\langle u_{m\boldsymbol{k}'}|e^{-i\boldsymbol{G}\cdot\boldsymbol{r}}|u_{n\boldsymbol{k}}\rangle_{\text{uc}}$, which are easily computed by integrating over the primitive cell (denoted as uc at the subscript), and the Fourier coefficients $\Delta\widetilde{V}_{\text{L}}(\boldsymbol{q})$, which are computed in the supercell. In addition, since the local perturbation potential due to a neutral defect is smooth and decays in real space over a few angstroms, its Fourier coefficients $\Delta\widetilde{V}_{\text{L}}(\boldsymbol{q})$ decay rapidly in reciprocal space, and the summation over $\boldsymbol{G}$ in Eq. (2.8) can be truncated to a small cutoff – typically, just a few reciprocal lattice vectors – to reduce the computational cost. For the same reason, the Fourier coefficients $\Delta\widetilde{V}_{\text{L}}(\boldsymbol{q})$ can be efficiently interpolated (in our case, with a B-spline interpolation method) at arbitrary $\boldsymbol{q}$ starting from their calculation at a few thousand $\boldsymbol{q}$-points in a cubic box centered at $\boldsymbol{q} = 0$. A great advantage of this formulation is that one can compute the local matrix elements using only the wave functions of the primitive cell (as opposed to those of the supercell), and effectively interpolate the perturbation potential to fine BZ grids.

The nonlocal matrix elements $M^{\text{NL}}_{mn}(\boldsymbol{k}',\boldsymbol{k})$ are computed as the difference between the nonlocal potentials of a defect-containing and a pristine supercell:

$$\begin{aligned} M^{\text{NL}}_{mn}(\boldsymbol{k}',\boldsymbol{k}) &= \langle m\boldsymbol{k}'|\hat{V}^{(\text{d})}_{\text{NL}} - \hat{V}^{(\text{p})}_{\text{NL}}|n\boldsymbol{k}\rangle \\ &= \langle m\boldsymbol{k}'|\hat{V}^{(\text{d})}_{\text{NL}}|n\boldsymbol{k}\rangle - \langle m\boldsymbol{k}'|\hat{V}^{(\text{p})}_{\text{NL}}|n\boldsymbol{k}\rangle\,, \end{aligned} \tag{2.10}$$

where the matrix elements of $\hat{V}_{\text{NL}}$ for each of the two supercells (labelled by α = d, p), using Eq. (2.5), read:

$$\langle m\boldsymbol{k}'|\hat{V}^{(\alpha)}_{\text{NL}}|n\boldsymbol{k}\rangle = \sum_{s_\alpha=1}\sum_{ij} D^{(s_\alpha)}_{ij}\,\langle n'\boldsymbol{k}'|\beta^{(s_\alpha)}_i\rangle\,\langle\beta^{(s_\alpha)}_j|n\boldsymbol{k}\rangle\,. \tag{2.11}$$

Similar to the local matrix elements, the computation in Eq. (2.11) is split into a primitive cell and a supercell calculation, by expressing the scalar products $\langle\beta^{(s)}_j|n\boldsymbol{k}\rangle$ as (See Appendix C)

$$\langle\beta^{(s)}_j|n\boldsymbol{k}\rangle = \frac{1}{\sqrt{\Omega_{\text{uc}}}}\sum_{\boldsymbol{G}} B^{(s)*}_{j\boldsymbol{k}}(\boldsymbol{G})\,\langle e^{i\boldsymbol{G}\cdot\boldsymbol{r}}|u_{n\boldsymbol{k}}\rangle_{\text{uc}}\,, \tag{2.12}$$

where $B^{(s)}_{j\boldsymbol{k}}(\boldsymbol{G})$ is the Fourier coefficient of the KB projector $\beta^{(s)}_j$ (multiplied by the phase factor $e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}$) at the primitive-cell reciprocal lattice vector $\boldsymbol{G}$:

$$B^{(s)}_{j\boldsymbol{k}}(\boldsymbol{G}) = \frac{1}{\sqrt{\Omega_{\text{uc}}}}\int_{\Omega_{\text{sup}}} d\boldsymbol{r}\;\left[\beta^{(s)}_j(\boldsymbol{r})e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\right]e^{-i\boldsymbol{G}\cdot\boldsymbol{r}}. \tag{2.13}$$

Note that the nonlocal matrix elements, which are computed as the difference in Eq. (2.10), are nonzero both because the atomic positions change upon relaxing the structure in the defect-containing supercell and because the number and type of atoms, in general, differ in the two supercells, as is the case when considering a vacancy or impurity. The nonlocal matrix elements, which involve the KB projects and are thus more complicated than the local matrix elements, are often ignored in the all-supercell methods. Here we include the nonlocal matrix elements because they can affect the phase and the magnitude of the total matrix elements $M_{mn}(\boldsymbol{k}', \boldsymbol{k})$.

2.3 Semiclassical Boltzmann Equation within the Relaxation Time Approximation

We compute the defect-limited carrier mobility μ at temperature T within the RT approximation of the linearized Boltzmann transport equation [15]:

$$\mu_{\alpha\beta}(T) = \frac{e}{n_c} \int_{-\infty}^{+\infty} dE \; [-\partial f(T,E)/\partial E\,] \times \Sigma_{\alpha\beta}\,(E)\,, \tag{2.14}$$

where e is the electron charge, n_c the carrier concentration, $f(T, E)$ the Fermi-Dirac distribution, and $\Sigma\,(E)$ the transport distribution function (TDF) at energy E,

$$\Sigma_{\alpha\beta}\,(E) = \frac{2}{\Omega_{\mathrm{uc}}} \sum_{n\boldsymbol{k}} \tau_{n\boldsymbol{k}} \mathbf{v}^{\alpha}_{n\boldsymbol{k}} \mathbf{v}^{\beta}_{n\boldsymbol{k}} \delta\,(E - \varepsilon_{n\boldsymbol{k}})\,, \tag{2.15}$$

where α and β are Cartesian directions. The TDF is computed with a tetrahedron integration method [9], using converged e-d RTs and Wannier-interpolated band velocities $\mathbf{v}_{n\boldsymbol{k}}$ [16, 17].

2.4 Electron-Defect Interaction Workflow and Computational Details

Figure 2.1 shows our workflow for computing properties related to e-d interactions from first principles as follows: (1) several inputs are computed with DFT, including the KS wave functions and eigenvalues of a primitive cell, and the local and nonlocal parts of the KS potential, separately in a pristine and a defect-containing supercell. (2) The local and nonlocal matrix elements are then computed by splitting the calculation into a primitive cell and a supercell part, an approach that dramatically reduces the computational cost. Importantly, only the KS wave functions, band structure, and $\boldsymbol{k}$-points of the primitive cell are used, while the supercells are employed only to obtain the perturbation potential due to the defect. Computing local matrix elements is the most expensive step, while the nonlocal matrix elements only involve relatively inexpensive reciprocal space sums. (3) The local matrix elements are then computed by Fourier transforming and interpolating the local perturbation

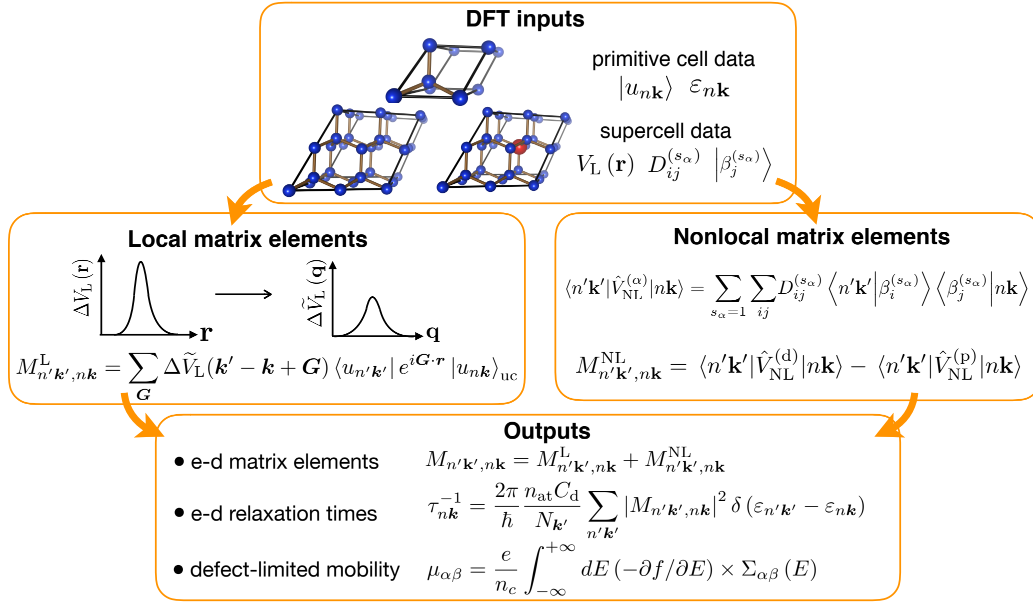

Figure 2.1: Workflow for computing the e-d matrix elements, relaxation times, and defect-limited mobility. The workflow is described in the main text.

potential, $\Delta V_{\mathrm{L}}(\boldsymbol{r})$, and combining it with the plane-wave matrix elements of the primitive cell. (4) The nonlocal matrix elements are similarly computed by splitting the calculation into a primitive cell and a supercell part, using only the KS wave functions of the primitive cell. (5) The total e-d matrix elements are then formed by adding the local and nonlocal parts. Once computed, the e-d matrix elements are employed to calculate the e-d RTs and the defect-limited mobility, among other quantities of interest. This approach allows us to systematically converge the RTs and other properties related to e-d interactions with respect to supercell size and BZ grids.

We apply our approach to compute the e-d RTs and defect-limited mobility in silicon, separately for vacancy and (tetrahedral) interstitial defects. A defect concentration of 1 ppm (one defect in 10^6 atoms) is assumed in both cases. The ground states of the primitive cell and of supercells with size $N \times N \times N$ (where N is the number of primitive cells along each lattice vector) are computed using DFT within the local density approximation [18], using a plane wave basis and norm-conserving pseudopotentials [14] with the Quantum ESPRESSO package [19]. Briefly, for the primitive cell we use a lattice constant of 5.43 Å, a 40 Ry kinetic energy cut-off and a $12 \times 12 \times 12$ Monkhorst-Pack $\boldsymbol{k}$-point grid [20], converging the total energy to within 10 meV/atom; a consistent lattice constant and total energy conver-

gence criterion is employed for the supercells. In the defect-containing supercells, the atomic forces are relaxed to within 25 meV/Å to account for the structural changes induced by the defect, and the resulting KS potentials are used to compute the e-d matrix elements. Due to the different reference potentials in the pristine and defect-containing supercells, we employ the core-average potential alignment method [21] to align the local potentials of the two supercells when computing the local perturbation potential; the reference potential is taken as the average of the local potential at the atom that is farthest from the defect site. In the e-d RT and mobility calculations, we select the electronic states of relevance in a small (~100 meV) energy window near the band edges, and interpolate the band structure using maximally localized Wannier functions [22] with the Wannier90 code [16, 17]. All e-d calculations have been implemented in our PERTURBO code [23] following the workflow in Fig. 2.1. The PERTURBO open source code can be downloaded at `https://perturbo-code.github.io/`.

2.5 Comparison between the All-Supercell Method and Our Approach

In the all-supercell method [24], one uses the pristine and the defect-containing supercells to provide all the necessary quantities for computing the e-d matrix elements, including the wave functions, band structure, perturbation potentials, and BZ grids. However, using supercell wave functions makes it challenging to compute and converge the e-d matrix elements and RTs, and ultimately to carry out accurate e-d calculations, since unconverged e-d RTs and transport properties can differ widely from the converged results.

All e-d calculations need to be converged with respect to supercell size; as we discuss below, converging the RTs for a neutral defect typically requires very large supercells with hundreds of atoms. In our approach, this convergence does not constitute a challenge since the same (primitive cell) wave functions are employed, regardless of supercell size. Conversely, in the all-supercell method, one uses wave functions from the pristine supercell, and the computational cost to compute and store the wave functions and obtain the matrix elements increases dramatically with supercell size, making accurate convergence tests too computationally demanding.

Let us analyze the cost of the most computationally intensive step, namely obtaining the local e-d matrix elements, $M^{\mathrm{L}}_{mn}(\boldsymbol{k}', \boldsymbol{k})$. Using a uniform BZ grid with $N_{\boldsymbol{k}}$ points, one obtains $\mathcal{O}(N_{\boldsymbol{k}}^2)$ matrix elements, each for a distinct $|n\boldsymbol{k}\rangle \rightarrow |m\boldsymbol{k}'\rangle$ e-d scattering process. In a typical calculation, a uniform grid with at least $N_{\boldsymbol{k}} \approx 10^6$ points is

needed to converge the RTs in the entire BZ. In a carrier mobility calculation, one typically selects a small energy window of ∼100 meV near the band edges (in a semiconductor, or near the Fermi energy in a metal), which reduces the required number of $\boldsymbol{k}$-points to $N_{\boldsymbol{k}} \approx 10^4$.

In the all-supercell method, the local matrix elements are computed as:

$$M^{\mathrm{L}}_{mn}(\boldsymbol{k}',\boldsymbol{k}) = \langle m\boldsymbol{k}'|\Delta V_{\mathrm{L}}(\boldsymbol{r})|n\boldsymbol{k}\rangle_{\mathrm{sup}}, \tag{2.16}$$

where the subscript (sup) denotes that both the local defect perturbation potential $\Delta V_{\mathrm{L}}(\boldsymbol{r})$ and the wave functions are obtained from a DFT calculation on a supercell. Since the cost of the DFT calculations scales with system size as roughly N_{at}^3, where N_{at} is the number of atoms in the supercell, computing $N_{\boldsymbol{k}}$ supercell wave functions, from which the local matrix elements are computed on the uniform grid, costs $N_{\boldsymbol{k}} \times N_{\mathrm{at}}^3$ in the all-supercell method.

By contrast, in our method only the primitive cell wave functions are used, and thus the computational cost of the matrix elements does not depend on N_{at} through the wave functions. To obtain the local matrix elements on the uniform grid with our method [see Eq. (2.8)], the only supercell data one needs are the Fourier coefficients $\Delta\widetilde{V}_{\mathrm{L}}(\boldsymbol{q})$ of the local defect perturbation potential. Obtaining these coefficients at a few thousand $\boldsymbol{q}$ points − from which an interpolation table can be constructed − has a cost that scales as N_{at}^3, but this step is required only once for a given supercell size. Therefore, computing the local matrix elements on a uniform BZ grid with $N_{\boldsymbol{k}}$ points has a cost of order N_{at}^3 in our method, versus a cost of $N_{\boldsymbol{k}} \times N_{\mathrm{at}}^3$ in the all-supercell method. For the typical mobility calculation mentioned above, this represents a speed-up by a factor of $N_{\boldsymbol{k}} \approx 10,000$ over the all-supercell method. Note that our carrier mobility calculations are already expensive (tens of thousands of CPU hours), so approaches that are thousands of times more expensive are clearly impractical. An additional benefit is that in our approach, the large supercell wave functions are never stored or loaded into memory, so the speed up is significant even for computing a single e-d RT.

Finally, one would like to map the e-d scattering processes onto the band structure of the primitive cell, as is done for e-ph scattering processes. This is possible in our approach due to our use of primitive cell band structures and $\boldsymbol{k}$-point grids, but impractical in the all-supercell method, where one uses the supercell band structures and $\boldsymbol{k}$-point grids, which depend on the choice of a supercell and differ from those of the primitive cell due to nontrivial BZ folding effects. Due to its computational

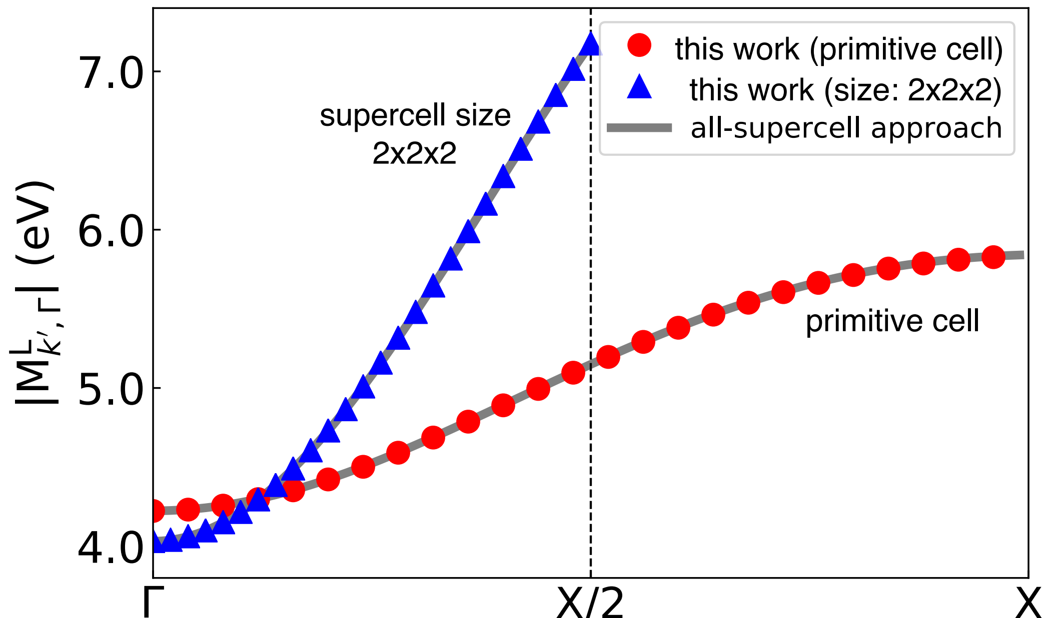

Figure 2.2: Absolute value of the local e-d matrix elements, obtained from our approach and, for comparison, with the all-supercell method, for two simulation cell sizes, a primitive cell and a $2 \times 2 \times 2$ supercell.

efficiency and convenience, we thus believe that our approach solves key technical challenges that have so far prevented efficient and accurate *ab initio* calculations of e-d interactions.

Figure 2.2 validates our approach by comparing the local e-d matrix elements computed with our method [using Eq. (2.8)] with those obtained with the all-supercell method using Eq. (2.16) for neutral vacancy defects in silicon. The initial state for the local matrix elements is in the lowest valence band at Γ, and the final states are in the same band with crystal momenta $\boldsymbol{k}'$ along the $\Gamma - X$ high-symmetry line. It is seen that for a test case of a primitive cell and a $2 \times 2 \times 2$ supercell, the results obtained with the two methods are in perfect agreement. For the supercell size $2 \times 2 \times 2$, the crystal momenta for the final states are along $\Gamma - X/2$ due to the folding of the BZ. This is but one of many benchmark tests we have performed.

2.6 *Ab Initio* Electron-Defect Relaxation Times

We first analyze the e-d RTs for neutral vacancy defects in silicon, and later study their convergence with supercell size and number of $\boldsymbol{k}$-points in the BZ. Figure 2.3(a) gives the converged RTs (and their inverse, the scattering rates) of electrons and

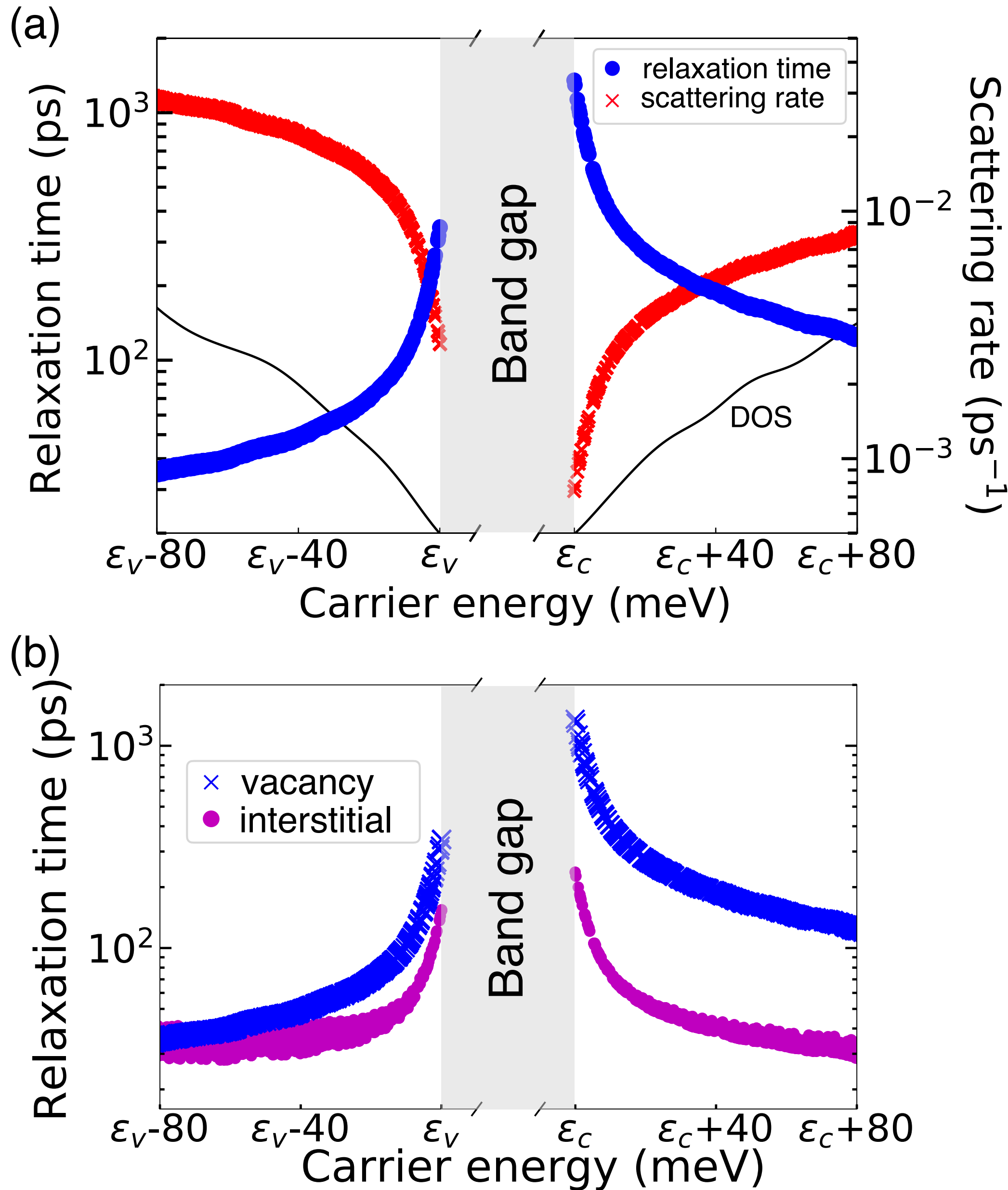

Figure 2.3: *Ab initio* electron-defect relaxation time. (a) The relaxation times and their inverse, the scattering rates, for electrons and holes due to *e*-d scattering with neutral vacancy defects in silicon. (b) Comparison between the relaxation times due to *e*-d scattering with neutral vacancy and interstitial defects in silicon.

holes near the band edges (the valence band maximum ε_v and the conduction band minimum ε_c) due to scattering with vacancy defects with a 1 ppm concentration as a reference point. For other defect concentration, the *e*-d scattering rates can be scaled linearly with the concentration as long as the lowest-order Born approximation holds. As the electron and hole energy increases away from the respective band edges, the

scattering rates increase (and the RTs decrease) due to the increased phase space for scattering − that is, the larger number of final states that can be accessed in the scattering process. The latter can be quantified by the density of states (DOS plotted in arbitrary units), which indeed shows a trend similar to the scattering rates. Overall, the *e*-d RTs are in the ps to ns range, and thus much longer than the typical *e*-ph RTs near room temperature, which are in the fs to ps range [15]. This result is consistent with the fact that *e*-ph scattering dominates near room temperature, while *e*-d scattering becomes important at low temperatures, where phonons are frozen out.

We find that the RTs depend on carrier energy, type of carrier, and type of neutral defect, at odds with simplified empirical models employed for decades to model *e*-d scattering [25]. Figure 2.3(b) shows the RTs as a function of carrier energy in silicon for neutral vacancy and interstitial defects, both with a 1 ppm concentration. For the vacancy defects, the electron RT is roughly 1 ns at the conduction band minimum and the hole RT is roughly 300 ps at the valence band maximum; both these RTs decrease by an order of magnitude ~100 meV away from the band edges. Overall, the holes relax significantly faster than the electrons, both near the band edge and at higher energies, a result we attribute to the higher density of states near the three-fold degenerate valence band maximum [26]. A similar asymmetry in the electron and hole RTs has been predicted in *e*-ph scattering in GaN [26], where it was also attributed to the valence band degeneracy. For interstitials, in contrast, we find that the RTs are similar for electrons and holes. They are 150 ps for electrons and 100 ps for holes at the respective band edges, and for both carriers, the RTs approach a value of 30 ps roughly 100 meV away from the band edges. The unexpected energy, carrier type, and defect type dependence of the RTs cannot be explained by the widely used Erginsoy formula [25], which predicts an energy-independent RT for *e*-d scattering due to neutral defects. Different from the Erginsoy model, our *ab initio* calculations take the atomic and electronic structure into account, providing accurate results that are material and defect specific. All the above calculations are converged using a 200^3 BZ $\boldsymbol{k}$-grid (the number of points for a $\boldsymbol{k}$-grid that is not achievable by the all-supercell method) with a 5 meV broadening, and a supercell size of $6 \times 6 \times 6$ (432 atoms) for vacancy and $8 \times 8 \times 8$ (1024 atoms) for interstitial defects. More details about convergence test, which has not been demonstarted in the literature, are discussed in the next session.

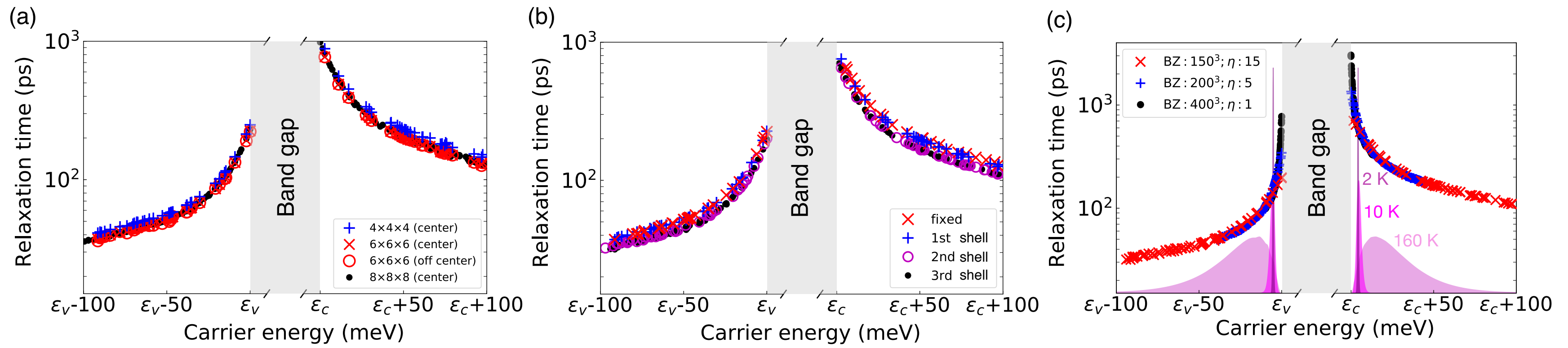

Figure 2.4: Convergence of the e-d RTs, shown here for vacancy defects in silicon. We consider the effect of (a) supercell size, (b) structural relaxation, and (c) BZ grid and energy broadening in Eq. (2.1), where for each broadening value, we use a converged BZ grid. Also shown in (c) is the function $(-\partial f/\partial E) \times \Sigma(E)$ in Eq. (3.11), at several temperatures in arbitrary units.

2.7 Rigorous Strategy for Converging Electron-Defect Relaxation Times

In *ab initio* e-d calculations, there are significant challenges with converging the RTs, which have so far not been examined in detail. This convergence is crucial since many transport properties and physical observables associated with e-d interactions depend sensitively on the RTs. Figure 2.4 shows how to systematically converge the e-d RTs with respect to three key factors – the supercell size, structural relaxation, and BZ grid used in the sum over final states in Eq. (2.1). This convergence study is discussed here for the neutral vacancy case, although we find similar results for interstitial defects.

Figure 2.4(a) shows the convergence of the RTs with respect to supercell size. Results are given for supercell sizes ranging from $4 \times 4 \times 4$ to $8 \times 8 \times 8$, in each case containing one vacancy at the center of the supercell. To isolate the role of supercell size, the atomic structure is not relaxed in these calculations. The RTs in the $4 \times 4 \times 4$ and $6 \times 6 \times 6$ supercells are within 20% and 5%, respectively, of the $8 \times 8 \times 8$ supercell results, which can be considered fully converged. To verify that the RTs do not depend on defect position, we compute the RTs for an off-center vacancy that is placed away by one primitive lattice vector from the center of a $6 \times 6 \times 6$ supercell. For the same supercell size, the RTs of the off-center and centered vacancy match exactly, as they should – as long as the defect perturbation potential of the off-center vacancy is still within the supercell.

Structural relaxation can be extensive around a defect and is expected to play an important role in accurately computing e-d interactions. Since it is costly to relax the structure in large supercells, an approximate scheme that retains accurate RTs is desirable. To this end, we compute the RTs in a supercell of a fixed size (here, $6 \times 6 \times 6$) in which only the atoms up to i-th nearest neighbor shell of the vacancy defect are relaxed, and those that are farther away are kept fixed because in real materials we expect the atoms that are far away from the defect to remain fixed. Figure 2.4(b) shows the RTs for structural relaxation up to the 1st, 2nd, and 3rd nearest neighbor shell. We find that the RTs are almost converged for structural relaxation within the 2nd shell, and nearly identical to those for relaxation up to the 3rd shell, which can be considered converged since the atomic forces are negligible outside the 3rd nearest neighbor shell. The conclusion is that one needs to relax only a small portion of the atoms around the defect to accurately compute the e-d RTs.

Most critical when computing the e-d RTs is converging the $\boldsymbol{k}'$-point grid in the sum

over final states in Eq. (2.1), which is equivalent to converging the grid of transferred momenta, $\boldsymbol{q} = \boldsymbol{k}' - \boldsymbol{k} - \boldsymbol{G}$. There is a cross-convergence effect between this grid and the energy broadening η employed to represent the delta function in Eq. (2.1), which is implemented as a normalized Gaussian with broadening, $\delta_\eta(x) = \frac{1}{\sqrt{2\pi}\eta} e^{-x^2/2\eta^2}$. The situation is fully analogous to converging the e-ph scattering rates [9]. Briefly, the broadening has to be small enough to not alter the final result, but the smaller the broadening, the denser the $\boldsymbol{k}'$-point BZ grid needed to converge the sum in Eq. (2.1). Systematic convergence is achieved by starting with a small broadening (say, $\eta \approx 10$ meV) and converging the $\boldsymbol{k}'$-point BZ grid, and then decreasing the broadening to a smaller value and converging the BZ sum again. At convergence, the RTs do not change upon decreasing the broadening and converging the BZ sum. Note that the $\boldsymbol{k}'$-point grid can in principle be distinct from the $\boldsymbol{k}$-point grid at which the RTs are computed, but this is feasible in practice only if one has a mechanism to effectively interpolate the matrix elements. When this is possible, using random or importance sampling $\boldsymbol{k}'$-point grids can significantly speed up the calculations [15]. Here, in each calculation, we use the same uniform BZ grid for $\boldsymbol{k}$- and $\boldsymbol{k}'$-points, and refer to it below as the BZ grid (a uniform $M \times M \times M$ grid will be denoted as an M^3 grid).

Figure 2.4(c) shows the RTs for several values of the energy broadening η and gives the corresponding BZ grid at convergence. The BZ grid required to converge the RTs are denser for smaller values of the broadening; for η values of 1, 5, and 15 meV, uniform BZ grids with 400^3, 200^3, and 150^3 points are needed, respectively. For electron energies higher than 25 meV above the conduction band minimum, a 15 meV broadening and a 150^3 BZ grid are sufficient to converge the RTs. For electron energies within 25 meV of the band edge, a 5 meV broadening with a 200^3 BZ grid gives the same RTs as a smaller 1 meV broadening with a 400^3 BZ grid. Converging the RTs within a few meV of the band edges may require even smaller broadenings and denser BZ grids, but it is not necessary for later carrier mobility calculations, even at very low temperatures, because the carrier velocity vanishes at the band edges. The broadening and BZ grid values at convergence are similar for electrons and holes, and for vacancies and interstitials. Importantly, the RTs computed with unconverged grids can differ widely from the converged values, especially at energies near the band edges [see Fig. 2.4(c)], which critically contribute to charge transport. It is therefore essential to have an efficient method for computing and converging the e-d RTs on fine BZ grids to accurately compute charge transport at low temperature.

2.8 Carrier Mobility due to Neutral Defects in Silicon

At room temperature, where e-ph interactions typically dominate, charge transport can be accurately predicted from first principles in several families of materials [7–11]. However, many devices and experiments operate at low temperature, where charge transport is governed by e-d interactions. It is thus critically important to develop *ab initio* calculations that can predict carrier dynamics at low temperature in the presence of e-d scattering.

To estimate the carrier energy range contributing significantly to the mobility, we plot the integrand of the mobility formula in Eq. (2.14), the function $(-\partial f/\partial E) \times \Sigma(E)$, in Fig. 2.4(c) for temperatures of 2, 10 and 160 K. As the temperature increases, the peak of the function broadens and moves away in energy from the band edges, indicating that the energy region contributing to the mobility shifts to higher carrier energies. The most stringent conditions for computing the RTs are below 10 K, where the contribution to the mobility peaks 5 meV away from the band edge; in this regime, BZ grids as dense as 200^3 $\boldsymbol{k}'$-points and a broadening of 5 meV are needed to accurately compute the mobility.

We compute the mobility of electrons and holes in silicon, considering separately neutral vacancy and interstitial defects. Figure 2.5 shows the computed mobility curves at temperatures below 150 K. Shown are the results for electron-vacancy (e−V), hole-vacancy (h−V), electron-interstitial (e−I), and hole-interstitial (h−I) interactions. The electron (solid circles) and hole (empty squares) mobilities are given for vacancy (dotted line) and interstitial (solid line) defects. A defect concentration of 1 defect in 10^6 atoms is assumed. The electron and hole mobilities for vacancy defects are higher than the corresponding mobilities for interstitial defects due to the longer e-d RTs for vacancies. We find that, in all cases, the defect-limited mobility is roughly constant below 10 K, and decreases at higher temperatures. Note that e-ph interactions are not included here, so these trends are due solely to the e-d interactions. Above 50 K, the temperature dependence of the mobility is well approximated by a power law, $\mu \propto T^{-\alpha}$, with coefficients α of order 0.5. For vacancy defects, the best-fit values of α are 0.53 for electrons and 0.55 for holes, and for interstitials, 0.46 for electrons and 0.38 for holes.

An interesting interpretation, which is particularly apt for vacancies, is that the defects can be regarded as a substance added to the pure crystal to make an "alloy." One thus expects that the temperature dependence of the carrier mobility for defect scattering is similar to that of alloy scattering, for which a power law with $\alpha = 0.5$ is

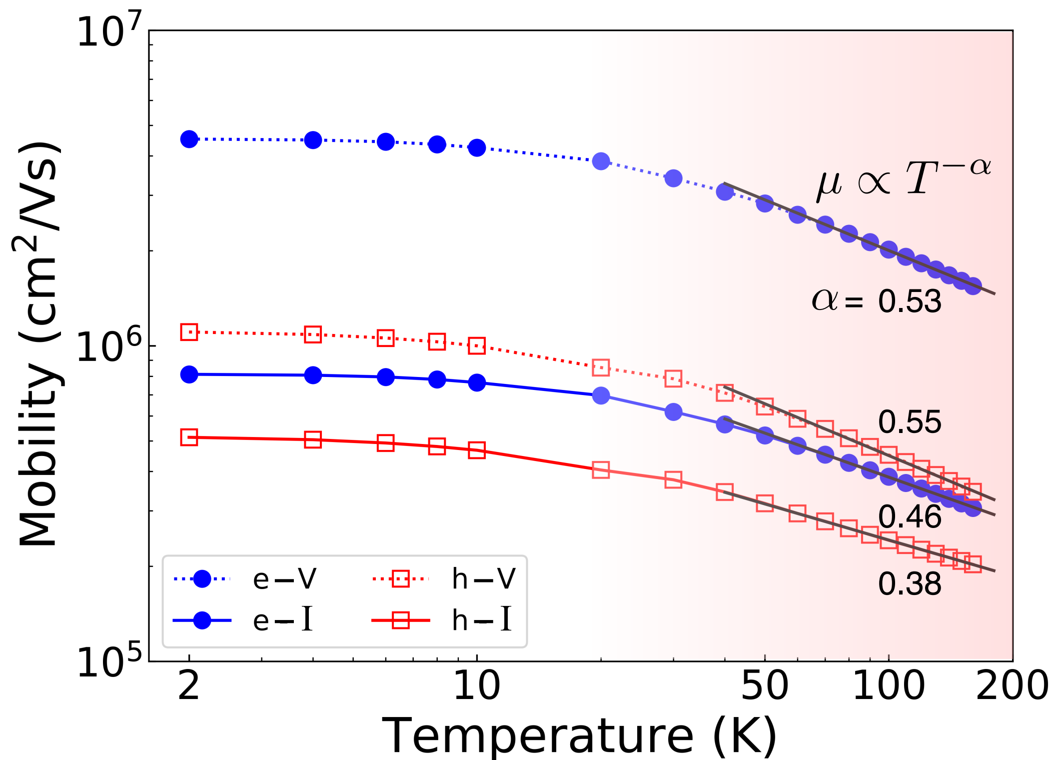

Figure 2.5: Defect-limited mobilities in silicon, as a function of temperature below 150 K.

expected based on existing models [27, 28]. Our results above 50 K are consistent with this interpretation, but we additionally find that the value of the exponent α depends on carrier and defect type. By contrast, the Erginsoy formula [25] predicts a temperature independent mobility for neutral defect scattering, which is clearly inconsistent with our results, and also with experiment.

An early experiment [29] on *n*-type doped silicon obtained the mobility due to neutral impurity scattering by subtracting the lattice and ionized impurity contributions using emperical models for *e*-ph and *e*-d scatterings. The resulting mobility, which is limited by neutral defect scattering alone, decreases above 50 K, in agreement with our results. While this trend has been attributed by the authors to inelastic *e*-d scattering [29], our results show that it can be explained by accurately computing elastic *e*-d scattering. The mobility decrease is due to the energy dependence of the RTs – as the temperature increases, so does the average energy of the electrons contributing to the mobility [see Fig. 2.4(c)], and their average RTs decrease as we have shown, causing the mobility decrease with temperature.

The mobility computed here using e-d interactions with only neutral point defects can be seen as an upper bound corresponding to an ideally pure material. In silicon, the mobility in the purest crystals (with impurity concentrations of $\sim 10^{12}$ cm^{-3}) is roughly 10^6 cm^2/Vs at 10 K, a value that can be explained by ionized impurity scattering [30]. In these pure silicon samples, the concentration of neutral defects can also be as low as 10^{12} cm^{-3} ($C_\mathrm{d} \approx 10^{-10}$ in dimensionless units) [30]. We extend our results to this lower neutral defect concentration, and estimate a mobility limit in silicon of $\sim 10^{10}$ cm^2/Vs at 10 K for an ideal scenario in which ionized impurity scattering is absent and only neutral point defects scatter the carriers. This mobility limit is higher than the value measured in samples with ionized impurity scattering, since the latter is much stronger than neutral defect scattering due to its long-range character. More extensive quantitative comparisons between computed and measured low-temperature charge transport data will be the subject of future work. We conclude that our approach is a powerful tool to compute charge transport at low temperature and estimate mobility limit values in materials.

We close by discussing technical remarks. Since the broadening needed to converge the RTs increases with increasing carrier energy, using an adaptive broadening scheme could significantly speed up the RT calculations. One could use relatively coarse BZ grids and larger broadening values at higher carrier energies to save computational time, while using finer grids (and a smaller broadening) only at low carrier energy. The proposed e-d calculations are general since they take into account the atomic structure of the material (including important structural relaxation effects around the defect) and its electronic structure. Unlike empirical models, there is no particular extension needed to treat different types of point and extended defects or different materials, provided one can afford the large DFT calculations needed to obtain the perturbation potentials. For example, our method is suitable for extended defects such as dislocations or grain boundaries, but to study them one may need supercells with thousands of atoms.

2.9 Summary

We have presented an efficient approach that overcomes the main technical challenges for *ab initio* calculations of e-d interactions. The method is applied to compute and systematically converge the elastic e-d RTs and the associated defect-limited carrier mobility below 150 K for vacancy and interstitial defects in silicon. The RTs exhibit a pronounced dependence on energy, carrier type, defect type, and the defect-limited mobility is temperature dependent. These results cannot be

explained using widely used empirical models of e-d interactions. Our approach can provide new microscopic insight into e-d scattering processes. It is broadly applicable and can be generalized to treat charged defects, magnetic impurities, and extended defects. We expect that this work will lay a solid foundation for efficient *ab initio* calculations of e-d interactions.

To further speed up and overcome the demanding e-d matrix element calculations and the BZ summation to obtain the e-d RTs, a promising strategy is interpolating the e-d matrix elements. Similar to e-ph calculations, interpolating the e-d matrix elements, for example using Wannier functions or atomic orbitals [6], would be highly desirable as one could compute the primitive cell wave functions only on coarse grids rather than on the fine grids needed to converge the RTs. Developing such an interpolation method for e-d matrix elements is the primary goal of the next chapter.

References

[1] D. Chattopadhyay and H. J. Queisser, Rev. Mod. Phys. **53**, 745 (1981).

[2] M. Bernardi, D. Vigil-Fowler, J. Lischner, J. B. Neaton, and S. G. Louie, Phys. Rev. Lett. **112**, 257402 (2014).

[3] M. Bernardi, D. Vigil-Fowler, C. S. Ong, J. B. Neaton, and S. G. Louie, Proc. Natl. Acad. Sci. U.S.A. **112**, 5291 (2015).

[4] M. Bernardi, J. Mustafa, J. B. Neaton, and S. G. Louie, Nat. Commun. **6**, 7044 (2015).

[5] H. Tanimura, J. Kanasaki, K. Tanimura, J. Sjakste, N. Vast, M. Calandra, and F. Mauri, Phys. Rev. B **93**, 161203 (2016).

[6] L. A. Agapito and M. Bernardi, Phys. Rev. B **97**, 235146 (2018).

[7] J. I. Mustafa, M. Bernardi, J. B. Neaton, and S. G. Louie, Phys. Rev. B **94**, 155105 (2016).

[8] C. H. Park, N. Bonini, T. Sohier, G. Samsonidze, B. Kozinsky, M. Calandra, F. Mauri, and N. Marzari, Nano Lett. **14**, 1113 (2014).

[9] J. J. Zhou and M. Bernardi, Phys. Rev. B **94**, 201201(R) (2016).

[10] N.-E. Lee, J.-J. Zhou, L. A. Agapito, and M. Bernardi, Phys. Rev. B **97**, 115203 (2018).

[11] J.-J. Zhou, O. Hellman, and M. Bernardi, Phys. Rev. Lett. **121**, 226603 (2018).

[12] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[13] R. M. Martin, *Electronic structure: Basic theory and practical methods* (Cambridge University Press, 2004).

[14] L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[15] M. Bernardi, Eur. Phys. J. B **89**, 239 (2016).

[16] J. R. Yates, X. Wang, D. Vanderbilt, and I. Souza, Phys. Rev. B **75**, 195121 (2007).

[17] A. A. Mostofi, J. R. Yates, G. Pizzi, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, Comput. Phys. Commun. **185**, 2309 (2014).

[18] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[19] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, J. Phys.: Condens. Matter **21**, 395502 (2009).

[20] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[21] Y. Kumagai and F. Oba, Phys. Rev. B **89**, 195205 (2014).

[22] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997).

[23] J.-J. Zhou, J. Park, I.-T. Lu, I. Maliyov, X. Tong, and M. Bernardi, arXiv:2002.02045 (2020).

[24] O. D. Restrepo, K. Varga, and S. T. Pantelides, Appl. Phys. Lett. **94**, 212103 (2009).

[25] C. Erginsoy, Phys. Rev. **79**, 1013 (1950).

[26] V. A. Jhalani, J.-J. Zhou, and M. Bernardi, Nano Lett. **17**, 5012 (2017).

[27] M. P. Vaughan, F. Murphy-Armando, and S. Fahy, Phys. Rev. B **85**, 165209 (2012).

[28] C. Hamaguchi, *Basic semiconductor physics* (Springer-Verlag Berlin Heidelberg, 2010).

[29] P. Norton, T. Braggins, and H. Levinstein, Phys. Rev. B **8**, 5632 (1973).

[30] C. Canali, C. Jacoboni, F. Nava, G. Ottaviani, and A. Alberigi-Quaranta, Phys. Rev. B **12**, 2265 (1975).

Chapter 3

AB INITIO ELECTRON-DEFECT INTERACTIONS USING WANNIER FUNCTIONS

3.1 Introduction

Ab initio e-d calculations using pseudopotentials or projector augmented waves [1–3] have progressed slowly in the last decade, mainly due to the high computational cost of obtaining the e-d interaction matrix elements needed for perturbative calculations when supercell wavefunctions are used [4]. We recently developed an *ab initio* method [3], as shown in Chapter 2, to compute efficiently the e-d interactions and the associated matrix elements. Our approach uses only the wave functions of the primitive cell, thus significantly reducing computational cost compared to e-d calculations that use supercell wave functions [1, 2]. Since our method uses a plane wave (PW) basis set and pseudopotentials, it is compatible with widely used density functional theory (DFT) codes. A different method developed by Kaasbjerg *et al.* [5] uses an atomic orbital (AO) basis to compute the e-d matrix elements; the advantage of this approach is that one can compute the e-d matrix elements using only a small set of AOs, although one is limited by the quality and completeness of the AO basis set [6].

To benefit from both the completeness and accuracy of the PW basis and the versatility of a small localized basis set, approaches combining PWs and AOs [7] or Wannier functions (WFs) [8, 9] have been developed for electron-phonon (e-ph) interactions. They have enabled efficient interpolation of the e-ph matrix elements and have been instrumental to advancing carrier dynamics calculations [10–12]. To date, such an interpolation scheme does not exist for e-d interactions to our knowledge, so performing demanding Brillouin zone (BZ) integrals needed to compute e-d relaxation times (RTs) and defect-limited charge transport remains an open problem. Interpolating the interaction matrix elements to uniform, random, or importance-sampling fine BZ grids is key to systematically converging the RTs and transport properties [10, 13], and it has been an important development in first-principles calculations of e-ph interactions and phonon-limited charge transport.

In this chapter, we develop a method for interpolating the e-d interaction matrix elements using WFs. Through a generalized double-Fourier transform, our approach

can efficiently transform the matrix elements from a Bloch representation on a coarse BZ grid to a localized WF representation and ultimately to a Bloch representation on an arbitrary fine BZ grid. We show the rapid spatial decay of the e-d interactions in the WF basis, which is crucial to the accuracy and efficiency of the method. Using our approach, we investigate e-d interactions due to charge-neutral vacancies in silicon and copper. In both cases, we can accurately interpolate the e-d matrix elements and converge the e-d scattering rates and defect-limited carrier mobility or resistivity. In copper, we map the e-d RTs directly on the Fermi surface, and show their peculiar dependence on electronic state. We demonstrate computations of e-d matrix elements on random and uniform BZ grids as dense as $600\times600\times600$ points, whose computational cost would be prohibitive for direct computation.

3.2 Derivation of a Fourier-Wannier Interpolation Scheme for Electron-Defect Matrix Elements

The perturbation potential $\Delta V_{\mathrm{e}-d}$ introduced by a point defect in a crystal couples different Bloch eigenstates of the unperturbed (defect-free) crystal. The matrix elements associated with this e-d interaction are defined as

$$M_{mn}(\boldsymbol{k}',\boldsymbol{k}) = \langle m\boldsymbol{k}'|\Delta V_{\mathrm{e}-d}|n\boldsymbol{k}\rangle\,, \tag{3.1}$$

where $|n\boldsymbol{k}\rangle$ is the Bloch state with band index n and crystal momentum $\boldsymbol{k}$. To handle these e-d interactions, one needs to store and manipulate a matrix M_{mn} of size N_b^2 (N_b is the number of bands) for each pair of crystal momenta $\boldsymbol{k}'$ and $\boldsymbol{k}$ in the BZ. Within DFT, the perturbation potential $\Delta V_{\mathrm{e}-d}$ can be computed as the difference between the Kohn-Sham potential of a defect-containing supercell and that of a pristine supercell with no defect [3].

We compute the e-d matrix elements in Eq. (3.1) using the method we developed in Chapter 1, which uses only the Bloch wave functions of the primitive cell and does not require computing or manipulating the wave functions of the supercell, thus significantly reducing computational cost. The Bloch states can be expressed in terms of maximally localized WFs using

$$|n\boldsymbol{k}\rangle = \sum_{j\boldsymbol{R}} e^{i\boldsymbol{k}\cdot\boldsymbol{R}} U^{\dagger}_{jn,\boldsymbol{k}}\,|j\boldsymbol{R}\rangle\,, \tag{3.2}$$

where $|j\boldsymbol{R}\rangle$ is the WF with index j centered at the Bravais lattice vector $\boldsymbol{R}$ in the Wigner-Seitz (WS) supercell where WFs are located. The unitary matrices U in

Eq. (3.2) maximize the spatial localization of the WFs [14]

$$|j\boldsymbol{R}\rangle = \frac{1}{N_{\boldsymbol{k}}}\sum_{n\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} U_{nj,\boldsymbol{k}} \, |n\boldsymbol{k}\rangle \,, \tag{3.3}$$

where $N_{\boldsymbol{k}}$ is the number of $\boldsymbol{k}$-points in the BZ. The e-d matrix element $M_{ij}(\boldsymbol{R}',\boldsymbol{R})$ between two WFs centered at the lattice vectors $\boldsymbol{R}'$ and $\boldsymbol{R}$ in the Wannier representation is defined as

$$M_{ij}(\boldsymbol{R}',\boldsymbol{R}) = \langle i\boldsymbol{R}'|\Delta V_{\mathrm{e}-d}|j\boldsymbol{R}\rangle \,. \tag{3.4}$$

If the center of the perturbation potential lies at the origin of the WS supercell, the absolute value of the e-d matrix elements, $|M_{ij}(\boldsymbol{R}',\boldsymbol{R})|$, decays rapidly (within a few lattice constants) for increasing values of the lattice vectors $\boldsymbol{R}'$ and $\boldsymbol{R}$ due to the short-range nature of the perturbation potential from the defect, which is assumed to be charge-neutral here; for charged defects, we need another approach as described in the next Chapter. As a result, only a small number of lattice vectors $\boldsymbol{R}$, which we arrange in a WS supercell centered at the origin, is needed to compute the e-d matrix elements in the Wannier representation.

Using Eqs. (3.1)-(3.4), the e-d matrix elements in the Wannier representation can be written as a generalized double Fourier transform of the matrix elements in the Bloch representation, which are first computed on a coarse BZ grid with points $\boldsymbol{k}_{\mathrm{c}}$:

$$M(\boldsymbol{R}',\boldsymbol{R}) = \left(\frac{1}{N_{\boldsymbol{k}_{\mathrm{c}}}}\right)^2 \sum_{\boldsymbol{k}'_{\mathrm{c}}\boldsymbol{k}_{\mathrm{c}}} e^{i(\boldsymbol{k}'_{\mathrm{c}}\cdot\boldsymbol{R}'-\boldsymbol{k}_{\mathrm{c}}\cdot\boldsymbol{R})} U^{\dagger}_{\boldsymbol{k}'_{\mathrm{c}}} M(\boldsymbol{k}'_{\mathrm{c}},\boldsymbol{k}_{\mathrm{c}}) U_{\boldsymbol{k}_{\mathrm{c}}}. \tag{3.5}$$

Here and below, we omit all band indices for clarity. Through the inverse transform, we can interpolate the e-d matrix elements to any desired pair of fine BZ grid points $\boldsymbol{k}'_{\mathrm{f}}$ and $\boldsymbol{k}_{\mathrm{f}}$, using

$$M(\boldsymbol{k}'_{\mathrm{f}},\boldsymbol{k}_{\mathrm{f}}) = \sum_{\boldsymbol{R}'\boldsymbol{R}} e^{-i(\boldsymbol{k}'_{\mathrm{f}}\cdot\boldsymbol{R}'-\boldsymbol{k}_{\mathrm{f}}\cdot\boldsymbol{R})} U_{\boldsymbol{k}'_{\mathrm{f}}} M(\boldsymbol{R}',\boldsymbol{R}) U^{\dagger}_{\boldsymbol{k}_{\mathrm{f}}}. \tag{3.6}$$

While $U_{\boldsymbol{k}_{\mathrm{c}}}$ in Eq. (3.5) is the coarse-grid unitary matrix used to construct the WFs [see Eq. (3.3)], the unitary matrix on the fine grid, $U_{\boldsymbol{k}_{\mathrm{f}}}$ in Eq. (3.6) is obtained by diagonalizing the fine-grid Hamiltonian,

$$H(\boldsymbol{k}_{\mathrm{f}}) = \sum_{\boldsymbol{R}} e^{i\boldsymbol{k}_{\mathrm{f}}\cdot\boldsymbol{R}} H(\boldsymbol{R}), \tag{3.7}$$

where $H(\boldsymbol{R})$ is the electronic Hamiltonian in the WF basis. These equations are analogous to those used for interpolating the e-ph matrix elements [7–9], except that here the lattice vectors $\boldsymbol{R}'$ and $\boldsymbol{R}$ are both associated with electronic states. The

lattice vectors $\boldsymbol{R}'$ and $\boldsymbol{R}$ in the WS supercell are determined−through the periodic boundary conditions−by the $\boldsymbol{k}'_\mathrm{c}$ and $\boldsymbol{k}_\mathrm{c}$ coarse grids, respectively. In practice, we choose a uniform coarse BZ grid, and the size of the WS supercell is equal to the size of this coarse grid. Note that our e-d matrix elements in the Wannier representation require WFs only for the primitive cell, while a method developed in Ref. [15] requires WFs for the defect-containing supercells.

Similar to the e-ph case [7], the rapid spatial decay of the e-d matrix elements in the WF basis is crucial to reducing the computational cost since it puts an upper bound to the number of lattice sites $\boldsymbol{R}'$ and $\boldsymbol{R}$ at which $M(\boldsymbol{R}', \boldsymbol{R})$ needs to be computed. In particular, while computing $M(\boldsymbol{k}'_\mathrm{f}, \boldsymbol{k}_\mathrm{f})$ at small $\boldsymbol{k}'_\mathrm{f}$ and $\boldsymbol{k}_\mathrm{f}$ vectors would in principle require summing the Fourier transform in Eq. (3.6) up to correspondingly large lattice vectors of length $|\boldsymbol{R}'| = 2\pi/|\boldsymbol{k}'_\mathrm{f}|$ and $|\boldsymbol{R}| = 2\pi/|\boldsymbol{k}_\mathrm{f}|$, in practice this is not needed due to the rapid spatial decay of the e-d matrix elements $M(\boldsymbol{R}', \boldsymbol{R})$. The choice of a WS supercell and its relation to the DFT supercell are discussed in the Appendix D.

3.3 Wannier Interpolation Workflow and Computational Details

The workflow for interpolating the e-d matrix elements to a fine grid with points $\boldsymbol{k}_\mathrm{f}$ consists of several steps (see Fig. 3.1): (i) Compute the e-d matrix elements in the Bloch representation on a coarse BZ grid with points $\boldsymbol{k}_\mathrm{c}$ using Eq. (3.1); (ii) obtain the e-d matrix elements in the Wannier representation using Eq. (3.5); (iii) interpolate the Hamiltonian using Eq. (3.7) and diagonalize it to obtain the fine-grid unitary matrices $U_{\boldsymbol{k}_\mathrm{f}}$; (iv) interpolate the e-d matrix elements to any desired pair of fine-grid points $\boldsymbol{k}'_\mathrm{f}$ and $\boldsymbol{k}_\mathrm{f}$ using the matrix elements in the Wannier representation and the fine-grid unitary matrices [see Eq. (3.6)].

The ground state of a primitive cell and of supercells with size $N \times N \times N$ (where N is the number of primitive cells along each lattice vector) are computed using DFT within the local density approximation. We use a plane-wave basis set and norm-conserving pseudopotentials [16] with the Quantum Espresso code [17]. The total energy is converged to within 10 meV/atom in all structures. In the defect-containing supercells, the atomic forces are relaxed to within 25 meV/Å to account for structural changes induced by the defect. For silicon, we use an experimental lattice constant of 5.43 Å and a plane-wave kinetic energy cutoff of 40 Ry. For copper, we use an experimental lattice constant of 3.61 Å and a plane-wave kinetic energy cutoff of 90 Ry. We use a 12 × 12 × 12 $\boldsymbol{k}$-point grid [18] for the primitive cells of

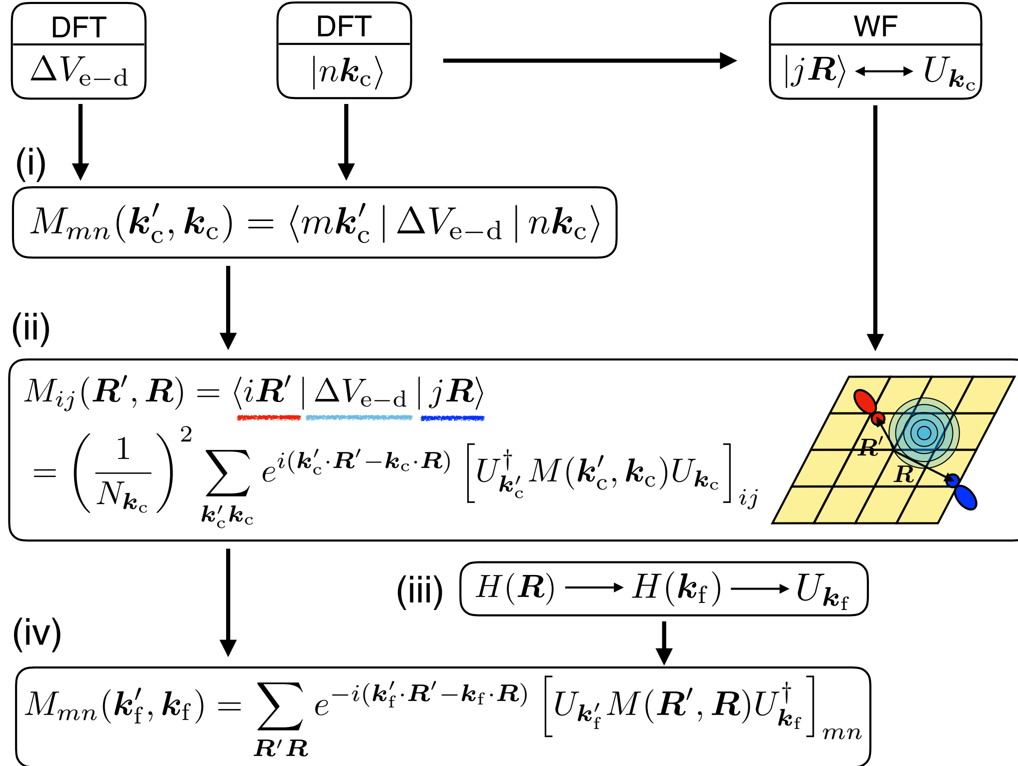

Figure 3.1: Workflow for interpolating the *e*-d matrix elements using WFs. The steps are numbered as in the text.

both materials to converge the charge density and total energy, and interpolate their band structures using maximally localized WFs [14] with the Wannier90 code [19, 20]. Coarse $\boldsymbol{k}_{\mathrm{c}}$-grids between $4 \times 4 \times 4$ and $10 \times 10 \times 10$ are used in the non-self-consistent DFT calculations and in the wannierization procedure. The dense $\boldsymbol{k}_{\mathrm{f}}$-grid used in the matrix element interpolation or RT calculations is unrelated to the WF generation. In silicon, we wannierize the four highest valence and four lowest conduction bands together, using sp^3 orbitals centered on the silicon atoms as the initial guess, to compute the *e*-d matrix elements used for electron and hole RTs. A different wannierization is employed to compute the *e*-d matrix elements along high-symmetry BZ lines in Fig. 3.2; in this case, to reduce the computational cost, we wannierize only the four highest valence bands using four *s* orbitals centered at the fractional coordinates (-1/8, 3/8, -1/8), (-1/8, -1/8, -1/8), (3/8, -1/8, -1/8), and (-1/8, -1/8, 3/8). In copper, we exclude the four lowest (core) 3*sp*-bands and wannierize the next seven bands using five *d* orbitals centered at the copper atom and two *s* orbitals centered at the fractional coordinates (1/4, 1/4, 1/4) and (-1/4, -1/4, -1/4) as the initial guess.

The methods to directly compute the e-d matrix elements and from them obtain the RTs, mobility, and conductivity (or resistivity) are described in detail in Chapter 2. Briefly, we compute the coarse-grid e-d matrix elements using the wave functions of the primitive cell, and obtain the perturbation potential due to a vacancy defect using a $6 \times 6 \times 6$ supercell with a $2 \times 2 \times 2$ $\boldsymbol{k}$-point grid. The atomic positions around the vacancy are relaxed up to the third nearest-neighbor shell in both silicon and copper. The potential alignment for the supercell containing the relaxed vacancy is chosen as the core-averaged potential of the farthest atom from the vacancy in the same (but unrelaxed) supercell.

In the e-d RT and defect-limited mobility calculations in silicon, we use only electronic states in a small (~100 meV) energy window near the band edges since these are the only states contributing to the mobility [10]; similarly, in copper we use only states within 100 meV of the Fermi energy. In silicon, we use a broadening value η = 5 meV to compute the delta function in Eq. (3.9), and a uniform BZ grid with 300^3 points for the RTs; for the mobility, we use a 1 meV broadening and e-d matrix elements interpolated from a 10^3 coarse BZ grid. In copper, we compute the RTs and resistivity on a fine BZ grid with 240^3 points, using a 1 meV broadening and e-d matrix elements interpolated from a coarse 8^3 BZ grid. The equilibrium vacancy concentration at temperature T in copper is estimated using [21]

$$C_\mathrm{v}(T) = e^{-(\Delta H_\mathrm{v} - T\Delta S_\mathrm{v})/k_\mathrm{B}T}, \tag{3.8}$$

where ΔH_v and ΔS_v are the vacancy formation enthalpy and entropy, respectively, and k_B is the Boltzmann constant. In copper, $\Delta S_\mathrm{v} = 3.0\ k_\mathrm{B}$ and $\Delta H_\mathrm{v} = 1.19$ eV [21].

3.4 Validation of the Wannier-Fourier Interpolation Method for Electron-Defect Matrix Elements

We validate our WF-based interpolation method using vacancy defects in silicon as an example. The relaxed symmetry of the vacancy in our calculation is T_d, while previous work and experiment find a D_{2d} symmetry [22]. The reason for this inconsistency is that we do not randomly displace the atoms before relaxation, which is needed to break the symmetry and obtain the lower-energy D_{2d} vacancy structure. Although our calculations use a vacancy with T_d rather than D_{2d} symmetry, the results we present are not affected by this choice. Figure 3.2(a) compares the e-d matrix elements calculated directly using Eq. (3.1) with the same matrix elements obtained by interpolation starting from two different coarse BZ grids with respec-

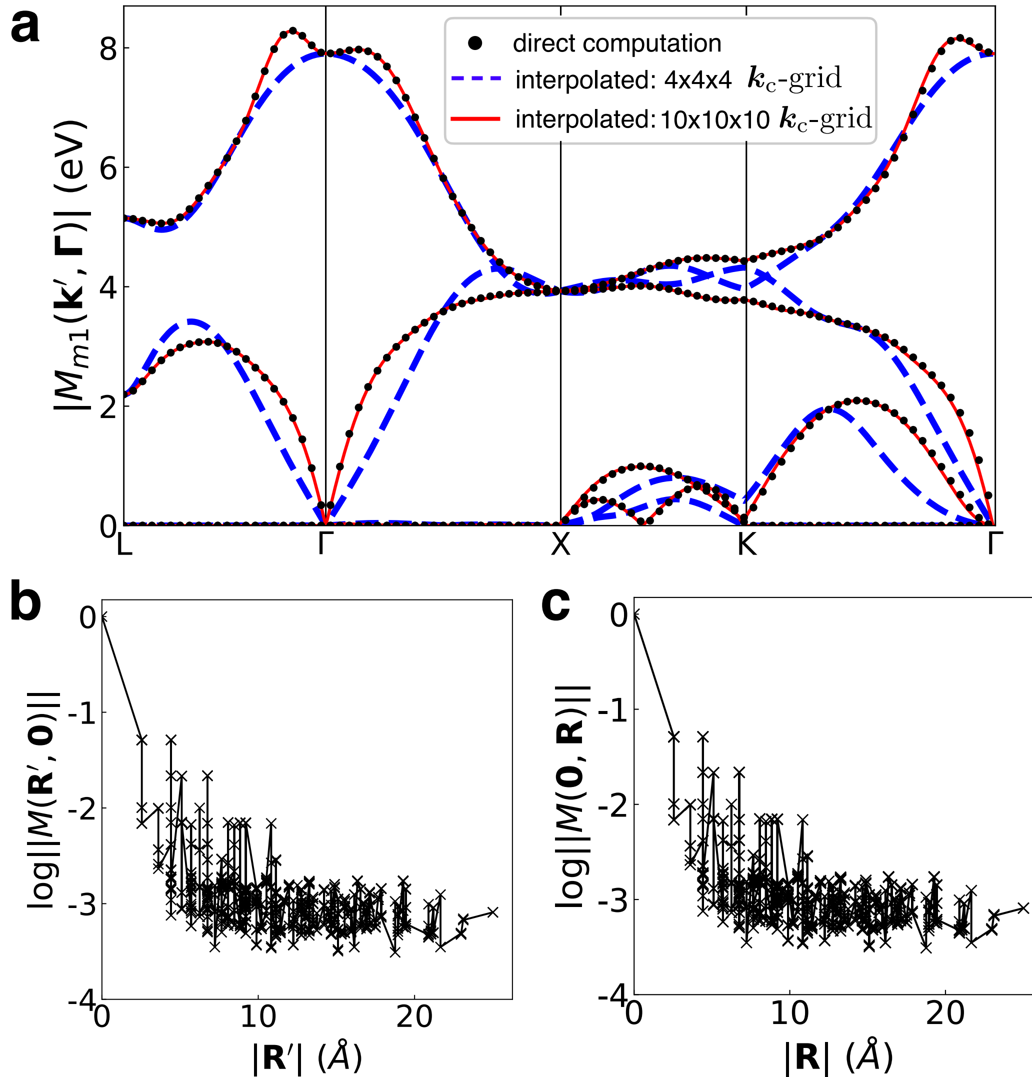

Figure 3.2: Absolute value of the electron-defect matrix elements along high-symmetry lines. (a) Absolute value of the e-d matrix elements, computed along high-symmetry BZ lines. The initial state is set to the lowest valence band at $\mathbf{\Gamma}$, while the final states span all four valence bands and possess crystal momenta chosen along the high-symmetry lines shown in figure. Panels (b) and (c) show the spatial decay of the e-d matrix elements in the Wannier basis, $||M(\boldsymbol{R}', \boldsymbol{R})||$, which are plotted in (b)) as a function of $|\boldsymbol{R}'|$ for $\boldsymbol{R} = \mathbf{0}$ and in (c) as a function of $|\boldsymbol{R}|$ for $\boldsymbol{R}' = \mathbf{0}$. The highest value of $||M(\boldsymbol{R}', \boldsymbol{R})||$ is normalized to 1 in both cases, and the plots use a logarithmic scale.

tively 4^3 and 10^3 points $\boldsymbol{k}_c$ (here and below, we denote an $N \times N \times N$ uniform grid as N^3). The interpolated results can qualitatively reproduce the direct computation for both coarse grids, but the results from the 10^3 coarse grid achieve a superior quantitative accuracy as the interpolated matrix elements agree with the directly

Table 3.1: The mean and maximum difference between the interpolated and directly computed e-d matrix elements, given for several coarse $\boldsymbol{k}_c$-grids. The data are for a neutral vacancy in silicon, with e-d matrix elements computed for the four lowest valence bands and along high-symmetry BZ lines L–Γ–X–K–Γ shown in Fig. 3.2.

$\boldsymbol{k}_c$-grids	Mean deviation (eV)	Max deviation (eV)
4 × 4 × 4	0.2074	1.2414
6 × 6 × 6	0.0820	0.7184
8 × 8 × 8	0.0300	0.3531
10 × 10 × 10	0.0095	0.1521

computed ones within 1% over the entire BZ. The accuracy can be systematically improved by increasing the size of the coarse $\boldsymbol{k}_c$-grid (see Table 3.1). This trend implies that the e-d perturbation potential decays to a negligible value over more than 4 but less than 10 lattice constants.

The spatial decay of the matrix elements in the WF basis is essential for the accuracy of our approach. To analyze the spatial behavior of the matrix elements in the WF basis, we define for each pair of lattice vectors $\boldsymbol{R}'$ and $\boldsymbol{R}$ the maximum absolute value of the e-d matrix elements as $||M(\boldsymbol{R}',\boldsymbol{R})|| = \max_{ij}|M_{ij}(\boldsymbol{R}',\boldsymbol{R})|$. Figures 3.2(b) and 3.2(c) show the spatial behavior of $||M(\boldsymbol{R}',\boldsymbol{R})||$ for a neutral vacancy defect in silicon as a function of $|\boldsymbol{R}'|$ while keeping $\boldsymbol{R} = \boldsymbol{0}$ and as a function of $|\boldsymbol{R}|$ while keeping $\boldsymbol{R}' = \boldsymbol{0}$, respectively. Note that $|M(\boldsymbol{R}',0)|$ and $|M(0,\boldsymbol{R})|$ in Fig. 3.2(b) and 3.2(c) are identical because the defect perturbation potential is centered at the origin of the WS supercell. This result does not hold in general for an arbitrary position of the defect in the WS supercell. We find that the matrix elements in the WF basis decay exponentially over a few unit cells, thus confirming that the WFs are a suitable basis set for interpolating the e-d matrix elements.

3.5 Carrier Relaxation Time and Mobility Using the Wannier Interpolation Method

The e-d RTs and the defect-limited carrier mobility are key to characterizing carrier dynamics at low temperatures, and also near room temperature in highly doped or disordered materials. Here we reproduce the equations for e-d RTs and defect-limited carrier mobility, which are shown in Chapter 1, for convenience. We compute the e-d RTs, $\tau_{n\boldsymbol{k}}$, associated with elastic carrier-defect scattering using the lowest-order Born approximation [3, 23]:

$$\tau_{n\boldsymbol{k}}^{-1} = \frac{2\pi}{\hbar}\frac{n_{\mathrm{at}}C_{\mathrm{d}}}{N_{\boldsymbol{k}'}}\sum_{m\boldsymbol{k}'}|M_{mn}(\boldsymbol{k}',\boldsymbol{k})|^2\,\delta(\varepsilon_{m\boldsymbol{k}'}-\varepsilon_{n\boldsymbol{k}}), \tag{3.9}$$

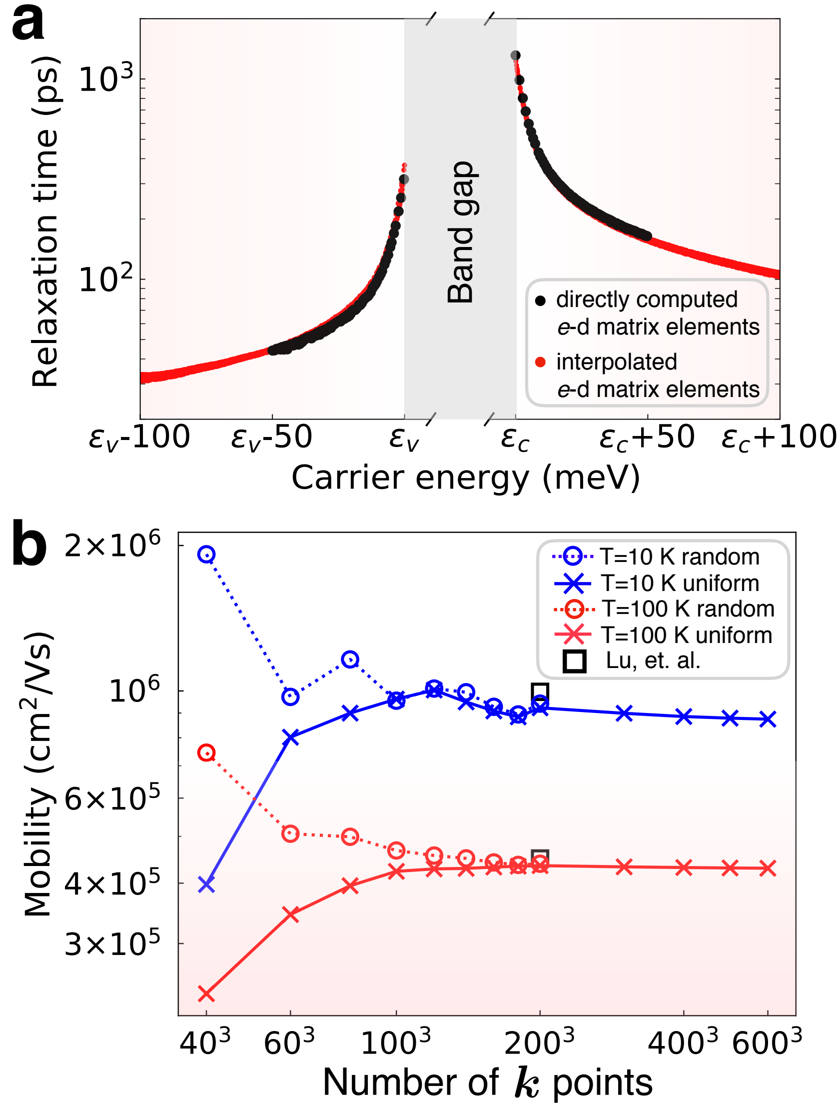

Figure 3.3: Carrier relaxation times and the hole mobility for vacancy defects in silicon. (a) Electron-defect RTs obtained from directly computed and interpolated e-d matrix elements. Here, ε_{c} is the conduction band minimum and ε_{v} the valence band maximum. (b) Convergence of the hole mobility with respect to the size of the fine BZ grid used for interpolation, shown for both uniform and random grids. The computed hole mobilities at 10 and 100 K from Ref. [3] are also shown. A reference vacancy concentration of 1 ppm is used in all calculations.

where $\hbar$ is the reduced Planck constant, n_{at} the number of atoms in a primitive cell, C_{d} the (dimensionless) defect concentration (defined as the number of defects divided by the number of atoms), $N_{\boldsymbol{k}'}$ the number of $\boldsymbol{k}$-points used in the summation, and $\varepsilon_{n\boldsymbol{k}}$ the unperturbed energy of the Bloch state $|n\boldsymbol{k}\rangle$ in the primitive cell. The delta function is implemented as a normalized Gaussian with a small broadening

η, $\delta_\eta(x) = e^{-x^2/2\eta^2}/\sqrt{2\pi}\eta$. Note that the e-d RTs are proportional to the defect concentration since our approach assumes that the scattering events are independent and uncorrelated [3]. For defect-limited carrier transport, we first compute the conductivity tensor $\sigma(T)$ at temperature T [10] using

$$\sigma_{\alpha\beta}(T) = e^2 \int_{-\infty}^{+\infty} dE\ [-\partial f(T,E)/\partial E\,] \times \Sigma_{\alpha\beta}(E), \tag{3.10}$$

where e is the electron charge and E the electron energy, $f(T,E)$ the Fermi-Dirac distribution, and $\Sigma(E)$ the transport distribution function (TDF) at energy E, defined as

$$\Sigma_{\alpha\beta}(E) = \frac{2}{\Omega_{\mathrm{uc}}} \sum_{n\boldsymbol{k}} \tau_{n\boldsymbol{k}} \mathbf{v}^{\alpha}_{n\boldsymbol{k}} \mathbf{v}^{\beta}_{n\boldsymbol{k}} \delta(E - \varepsilon_{n\boldsymbol{k}}), \tag{3.11}$$

where α and β are Cartesian directions, and Ω_{uc} the volume of the primitive cell. The TDF is computed with a tetrahedron integration method [10], using our calculated e-d RTs and Wannier-interpolated band velocities $\mathbf{v}_{n\boldsymbol{k}}$ [19, 20]. The mobility is obtained as $\mu = \sigma/n_c e$, where n_c is the carrier concentration, while the resistivity is obtained by inverting the conductivity tensor.

We first study the e-d RTs and hole carrier mobility in silicon with neutral vacancy defects. We use interpolated e-d matrix elements and focus on the accuracy and convergence of our interpolation method. The results given here assume a defect concentration of 1 vacancy in 10^6 silicon atoms (1 ppm concentration), but results for different defect concentrations can be obtained by rescaling these reference RTs to a different defect concentration [using Eq. (3.9)]; this approach is valid only within the concentration range in which the Born approximation holds. Figure 3.3(a) compares the RTs obtained from directly computed and interpolated e-d matrix elements. The two sets of RTs are in close agreement with each other for both electrons and holes, confirming the accuracy of our interpolation method. The directly computed matrix elements are calculated using the wave functions obtained from non-self-consistent calculations on a dense grid with 300^3 $\boldsymbol{k}_{\mathrm{f}}$-points, while the interpolated matrix elements are calculated on the same dense grid starting from a coarse grid with 10^3 $\boldsymbol{k}_{\mathrm{c}}$-points. Figure 3.3(b) shows the convergence of the defect-limited hole mobility with respect to the size of the fine BZ grids, for BZ grids ranging from 40^3 to 600^3 points; the convergence is studied at two temperatures (10 and 100 K) and for two types of grids, random and uniform. At 10 K, the mobilities are fully converged for fine grids with 200^3 points, for both random and uniform grids. We observe a similar trend at 100 K. The converged values of the mobility are consistent with

our previous calculations using directly computed (rather than interpolated) matrix elements [3].

3.6 Speedup of the Wannier Interpolation Method Compared to the Direct Computation Method

The interpolation method allows us to use extremely dense BZ grids with up to 600^3 points due to its superior computational efficiency. Let us briefly analyze the overall speed-up of the interpolation method for a carrier mobility calculation. To converge the mobility at low temperature, one needs to consider only fine BZ grid points in a small energy window, roughly within 100 meV of the band edges in semiconductors [3, 10] or of the Fermi energy in metals; these are the only states contributing to the conductivity in Eq. (3.10). In this small energy window, the number of $\boldsymbol{k}$-points is a small fraction α of the total number of points $N_{\boldsymbol{k}_\mathrm{f}}$ in the entire fine BZ grid. In the direct computation, one computes the e-d matrix elements $M(\boldsymbol{k}'_\mathrm{f}, \boldsymbol{k}_\mathrm{f})$ between all crystal momentum pairs, and thus a number of matrix elements of order $(\alpha N_{\boldsymbol{k}_\mathrm{f}})^2$. In the interpolation approach, the most time-consuming step is directly computing the $N^2_{\boldsymbol{k}_\mathrm{c}}$ e-d matrix elements on the coarse grid [step (i) in Fig. 3.1], whereas interpolating the matrix elements per se is orders of magnitude less computationally expensive. [In our machine, the average CPU time to directly compute one matrix element is ~ 0.2 s for silicon (with an electronic kinetic energy cutoff of 40 Ry), while the same calculation done with our interpolation method requires only $\sim 80\,\mu$s.] As a result, the overall speed-up of the interpolation approach over the direct computation is $\sim (\alpha N_{\boldsymbol{k}_\mathrm{f}})^2/N^2_{\boldsymbol{k}_\mathrm{c}}$. The typical value of $N_{\boldsymbol{k}_\mathrm{f}}$ is around 10^3–$10^4 N_{\boldsymbol{k}_\mathrm{c}}$. For our silicon calculations, the value of α is of order 10^{-2}, so the interpolation approach speeds up the mobility calculation by at least 2−4 orders of magnitude.

3.7 Wannier Interpolation Method for the Brillouin Zone Summation in a Metal

The method for directly computing the e-d matrix elements we developed in chapter 1 and the interpolation method shown here are general, and can be applied to metals, semiconductors and insulators. As an example, we show a calculation on a metal, copper, containing vacancy defects. In metals, the fine grids required to compute the e-d RTs near the Fermi energy and the resistivity are a major challenge for direct e-d calculations without interpolation. Figure 3.4(a) shows the e-d RTs computed at $\boldsymbol{k}$-points on the Fermi surface of copper, using interpolated e-d matrix elements

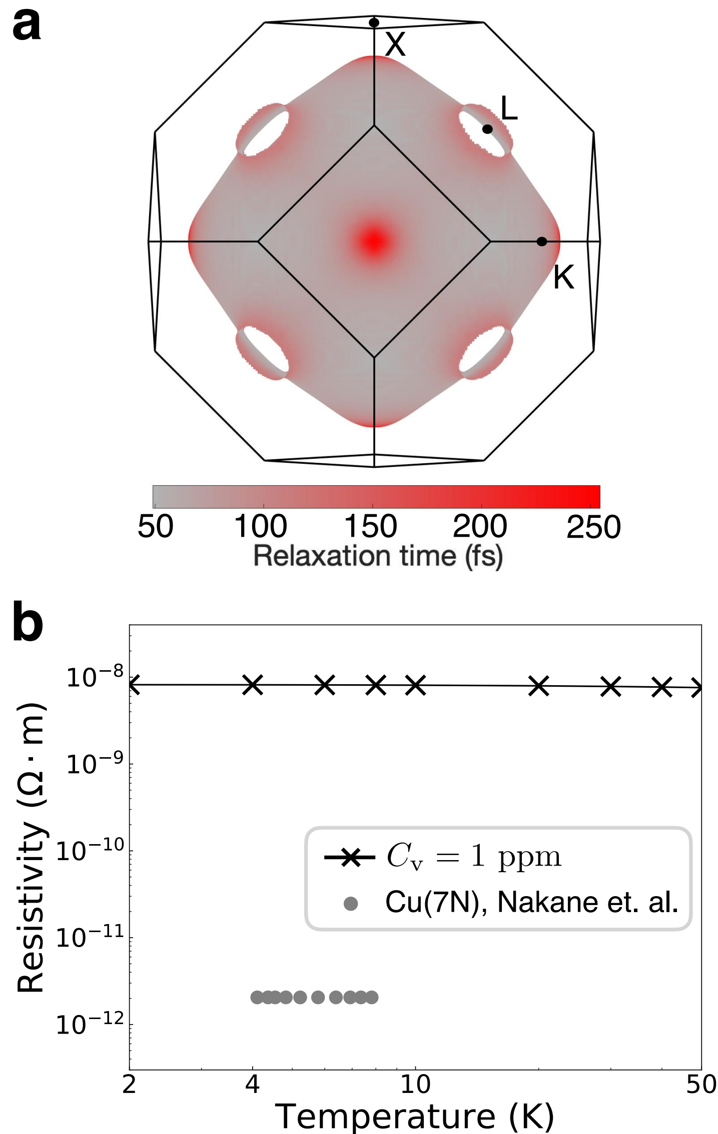

Figure 3.4: Relaxation times and the defect-limited resistivity for vacancy defects in copper. (a) RTs mapped on the Fermi surface, obtained using interpolated e-d matrix elements for a reference vacancy concentration of 1 ppm. (b) Defect-limited resistivity as a function of temperature for an assumed reference vacancy concentration of 1 ppm, compared with experimental data from Ref. [24].

obtained from a moderate size ($8 \times 8 \times 8$) coarse $\boldsymbol{k}_{\mathrm{c}}$-grid. The e-d RTs for a reference vacancy concentration of 1 ppm are between 0.3−1.4 ps. Interestingly, these e-d RTs are orders of magnitude shorter than the electron RTs in silicon for the same vacancy concentration, which suggests that scattering due to vacancy defects in copper is significantly stronger than in silicon. We compute the scattering strength $|M(\varepsilon_{n\boldsymbol{k}})|$

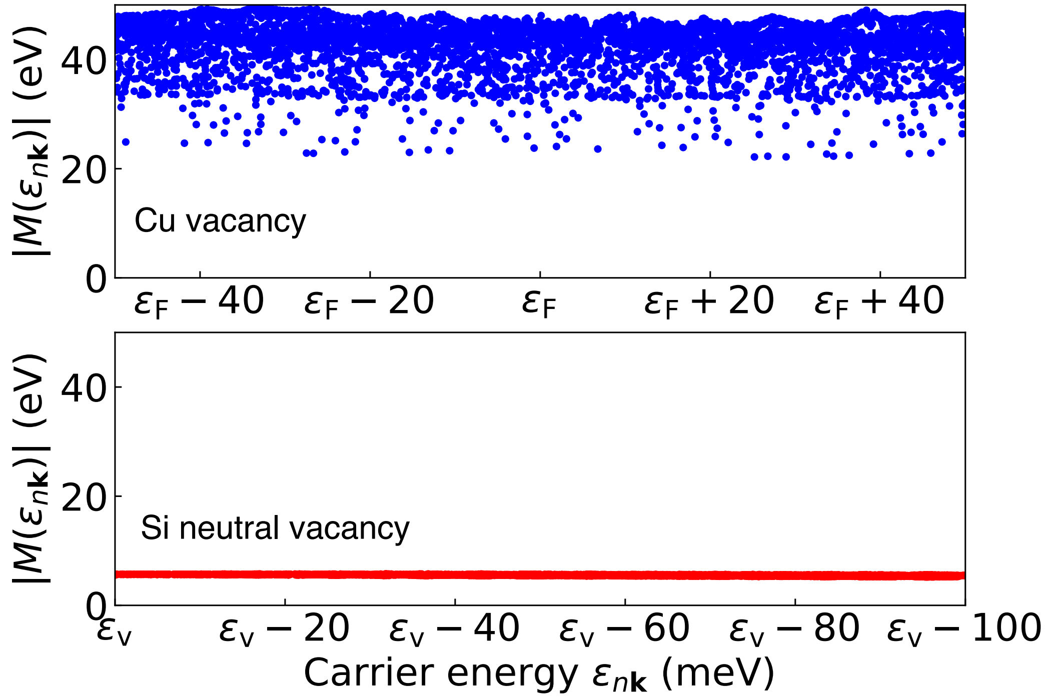

Figure 3.5: Scattering strength in copper (upper panel) and silicon (lower panel) for vacancy defects as a function of carrier energy. For copper, we show the scattering strength for electronic states within 50 meV of the Fermi energy ε_{F}, and in silicon for states within 100 meV of the valence band maximum ε_{v}. To compute the scattering strength in Eq. (3.12), we use a fine BZ grid with 240^3 $\boldsymbol{k}_{\mathrm{f}}$-points, starting from coarse BZ grids with 10^3 $\boldsymbol{k}_{\mathrm{c}}$-points for silicon and 8^3 $\boldsymbol{k}_{\mathrm{c}}$-points for copper.

for carrier energy $\varepsilon_{n\boldsymbol{k}}$ as

$$|M(\varepsilon_{n\boldsymbol{k}})| = \left(\frac{\sum_{m\boldsymbol{k}'} |M_{mn}(\boldsymbol{k}', \boldsymbol{k})|^2 \delta(\varepsilon_{m\boldsymbol{k}'} - \varepsilon_{n\boldsymbol{k}})}{\sum_{m\boldsymbol{k}'} \delta(\varepsilon_{m\boldsymbol{k}'} - \varepsilon_{n\boldsymbol{k}})} \right)^{1/2}, \tag{3.12}$$

where $M_{mn}(\boldsymbol{k}', \boldsymbol{k})$ is the e-d matrix element coupling Bloch state $|n\boldsymbol{k}\rangle$ to $|m\boldsymbol{k}'\rangle$. This quantity has contributions from all the final scattering states $|m\boldsymbol{k}'\rangle$ and can quantify the scattering strength for a carrier with energy $\varepsilon_{n\boldsymbol{k}}$. Figure 3.5 shows this e-d scattering strength for vacancy defects in copper (upper panel) and silicon (lower panel). In copper, the scattering strength for a 1 ppm vacancy concentration is of order 30-50 eV for states near the Fermi energy, versus a much lower value of about 5 eV in silicon for states near the valence band edge. This result shows quantitatively that electron-vacancy scattering is stronger in copper than in silicon. The e-d RTs due to vacancy scattering in copper are strongly state-dependent – we find values

of order 0.3 ps on the majority of the Fermi surface, as well as values as large as 1.4 ps near the regions of the Fermi surface close to the X points of the BZ. Similar state-dependent RTs due to impurity scattering in copper have been predicted using the all-electron Korringa-Kohn-Rostoker Green's function method [25, 26]. These results show that our first-principles approach can access microscopic details of the e-d scattering processes.

Figure 3.4(b) shows the calculated defect-limited resistivity for vacancy defects in copper at low temperatures between 2−50 K. The calculated resistivity is independent of temperature, in agreement with experimental results [24, 27], even though the conductivity formula we use [Eq. (3.10)] depends on temperature via the Fermi-Dirac distribution. The low-temperature defect-limited resistivity for a reference 1 ppm vacancy concentration is $\sim 10^{-9}$ Ω·m. However, the equilibrium vacancy concentration of copper at 50 K is negligible (of order 10^{-119}), and the corresponding resistivity is of order 10^{-122} Ω·m. This value is negligible in comparison with the measured resistivity of $\sim 10^{-12}$ Ω·m in a highly pure copper Cu(7N) sample at low temperature, where 7N means 99.99999% purity [see Fig. 3.4(b)]. We conclude that the resistivity of real copper samples at low temperature is not limited by intrinsic vacancy defects, but rather is controlled by impurities. This is a well-known result [28–32], and our calculations are consistent with it.

Our method can predict a lower bound of the residual resistivity due to intrinsic defects in an ideally pure material. Alternatively, if the main type of defect or impurity is known from experiment, our method can estimate the defect concentration present in the sample. The so-called residual resistivity ratio (RRR) between the low temperature and room temperature resistivities is used as a figure of merit for sample quality, and a large collection of data exists [33] for RRR in metals. Since at room temperature the resistivity is usually phonon-limited, combined with e-ph calculations [10, 34] our approach allows one to compute RRR for a wide range of materials and defect types. Taken together, these capabilities expand the tool box of first-principles methods for investigating carrier dynamics in complex materials.

3.8 Summary

In conclusion, we developed a WF-based interpolation approach to efficiently compute e-d interactions and the associated matrix elements on fine BZ grids. We have shown that the interpolation method is accurate and that it can effectively compute demanding BZ integrals requiring up to 10^8-10^9 $\boldsymbol{k}$-points. The ability to efficiently

interpolate e-d matrix elements starting from moderate BZ coarse grids is a stepping stone toward perturbative calculations of defect-limited charge and spin transport and to investigate quantum transport regimes governed by e-d interactions.

As discussed in the next Chapter, our e-d interpolation method can be extended to charged defects, for which the long-range Coulomb interactions can be added in reciprocal space similar to what is done for e-ph interactions [10, 35].

References

[1] O. D. Restrepo, K. Varga, and S. T. Pantelides, Appl. Phys. Lett. **94**, 212103 (2009).

[2] V. Lordi, P. Erhart, and D. Åberg, Phys. Rev. B **81**, 235204 (2010).

[3] I.-T. Lu, J.-J. Zhou, and M. Bernardi, Phys. Rev. Mater. **3**, 033804 (2019).

[4] M. Bernardi, Eur. Phys. J. B **89**, 239 (2016).

[5] K. Kaasbjerg, J. H. J. Martiny, T. Low, and A.-P. Jauho, Phys. Rev. B **96**, 241411(R) (2017).

[6] G. Ulian, S. Tosoni, and G. Valdrè, J. Chem. Phys. **139**, 204101 (2013).

[7] L. A. Agapito and M. Bernardi, Phys. Rev. B **97**, 235146 (2018).

[8] F. Giustino, M. Cohen, and S. Louie, Phys. Rev. B **76**, 165108 (2007).

[9] M. Calandra, G. Profeta, and F. Mauri, Phys. Rev. B **82**, 165111 (2010).

[10] J. J. Zhou and M. Bernardi, Phys. Rev. B **94**, 201201(R) (2016).

[11] J.-J. Zhou, O. Hellman, and M. Bernardi, Phys. Rev. Lett. **121**, 226603 (2018).

[12] N.-E. Lee, J.-J. Zhou, L. A. Agapito, and M. Bernardi, Phys. Rev. B **97**, 115203 (2018).

[13] J. Park, J.-J. Zhou, and M. Bernardi, Phys. Rev. B **101**, 045202 (2020).

[14] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997).

[15] T. Berlijn, D. Volja, W. Ku, et al., Phys. Rev. Lett. **106**, 077005 (2011).

[16] M. J. Van Setten, M. Giantomassi, E. Bousquet, M. J. Verstraete, D. R. Hamann, X. Gonze, and G.-M. Rignanese, Comput. Phys. Commun. **226**, 39 (2018).

[17] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, J. Phys.: Condens. Matter **21**, 395502 (2009).

[18] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[19] J. R. Yates, X. Wang, D. Vanderbilt, and I. Souza, Phys. Rev. B **75**, 195121 (2007).

[20] A. A. Mostofi, J. R. Yates, G. Pizzi, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, Comput. Phys. Commun. **185**, 2309 (2014).

[21] T. Hehenkamp, W. Berger, J.-E. Kluin, C. Lüdecke, and J. Wolff, Phys. Rev. B **45**, 1998 (1992).

[22] F. Corsetti and A. A. Mostofi, Phys. Rev. B **84**, 035209 (2011).

[23] H. Bruus and K. Flensberg, *Many-body quantum theory in condensed matter physics: An introduction* (Oxford University Press, 2004).

[24] H. Nakane, T. Watanabe, C. Nagata, S. Fujiwara, and S. Yoshizawa, IEEE Trans. Instrum. Meas. **41**, 107 (1992).

[25] D. V. Fedorov, P. Zahn, M. Gradhand, and I. Mertig, Phys. Rev. B **77**, 092406 (2008).

[26] M. Gradhand, D. V. Fedorov, P. Zahn, and I. Mertig, Phys. Rev. B **81**, 020403(R) (2010).

[27] Y. C. Cho, S. Lee, M. Ajmal, W.-K. Kim, C. R. Cho, S.-Y. Jeong, J. H. Park, S. E. Park, S. Park, H.-K. Pak, et al., Cryst. Growth Des. **10**, 2780 (2010).

[28] P. Jongenburger, Phys. Rev. **90**, 710 (1953).

[29] F. J. Blatt, Phys. Rev. **99**, 1708 (1955).

[30] A. Overhauser and R. Gorman, Phys. Rev. **102**, 676 (1956).

[31] J. Friedel, Advances in Physics **50**, 539 (2001).

[32] P. L. Rossiter, *The electrical resistivity of metals and alloys* (Cambridge University Press, 1991).

[33] L. Hall, *Survey of Electrical Resistivity Measurements on 16 Pure Metals in The Temperature Range 0 to 273 K*, Vol. 365 (US Dept. of Commerce, National Bureau of Standards, 1968).

[34] J. I. Mustafa, M. Bernardi, J. B. Neaton, and S. G. Louie, Phys. Rev. B **94**, 155105 (2016).

[35] J. Sjakste, N. Vast, M. Calandra, and F. Mauri, Phys. Rev. B **92**, 054307 (2015).

Chapter 4

AB INITIO ELECTRON-DEFECT INTERACTIONS FOR IONIZED IMPURITIES IN SEMICONDUCTORS

4.1 Introduction

The charge carrier mobility in semiconductors is typically limited by phonon scattering at room temperature and ionized (or charged) impurity scattering at low temperatures or in highly disordered materials [1–5]. The phonon-limited carrier mobility can be predicted accurately using first-principles approches, currently for materials with up to 50−100 atoms in the unit cell using the state-of-the-art code PERTURBO [6]. Different from *ab initio* electron-phonon (*e*-ph) calculations, advances in methods computing the carrier mobility limited by ionized impurities have been modest. The problem is highly important because ionized impurities are nearly always present in semiconductors, but progress has been hindered by the computational cost of electron-defect (*e*-d) calculations [7–9].

Early work on the *ab initio* defect-limited carrier mobility has heavily relied on the effective mass model whereby the carrier mobility is computed with the effective mass obtained from first-principles band structure calculations [10, 11]. The effective mass model assumes that the band structure is parabolic and that no interband transition occurs. Due to its simplified approximations, the simple effective mass model fails and needs tuning parameters or higher order effective mass corrections [12, 13]. Moreover, the model cannot be extended to deal with materials with complex bandstructures and with metals, semimetals, or topological semimetals and the surface states of topological insulators [14–16] because of the linear band dispersion.

To overcome the constraints of the effective mass model, we start by analyzing calculations of *e*-d scattering rates [or their inverse, the relaxation times (RTs)] and perform Brillouin zone (BZ) summations. As discussed previously, these are the two essential steps to compute the defect-limited carrier mobility. Computing *e*-d scattering rates requires *e*-d matrix elements due to charged defects. The *e*-d matrix elements are usually computed using plane waves instead of Bloch waves as a simplifying assumption [17]. However, the plane-wave method can only treat intraband scattering and neglects interband scattering. In real materials, the

electronic states can have more complicated wave functions than plane waves, and their bandstructure may exhibit multiple bands, allowing for interband scattering. *Ab initio* methods can provide more accurate wave functions than plane waves and capture complicated band structures. In addition, performing BZ summations to obtain the e-d scattering rates is a demanding task, which requires at least about 1 million points in the BZ $\boldsymbol{k}$-grid. To sum up, current e-d interaction methods for ionized impurity scattering are limited to using approximate wave functions and band structures, and cannot rigorously compute matrix elements, RTs, and transport properties governed by ionized impurities.

So far, *ab initio* methods for the e-d matrix elements using Bloch wave functions have mainly focused on short-range defects such as neutral defects [18–21]. To deal with charged defects with long-range interactions, the current methods cannot efficiently compute the e-d matrix elements due to the challenge of properly accounting for electronic screening and carrying out the demanding BZ summations involved in RT calculations. In some cases, the curve for the *ab initio* e-d RTs as a function of carrier energy is the best fit of the unconverged e-d RTs [22]. *Ab initio* e-ph calculations can provide useful tools for e-d calculations. A relevant example is the long-range Frohlich interaction in polar materials, which is added in momentum space in the e-ph calculations workflow [23–25]. Similar methods to compute and interpolate long-range e-d interaction are yet to be developed.

In this work, we aim to develop methods to compute and interpolate the e-d matrix elements due to ionized impurity scattering. We apply this method to silicon and compare the e-d matrix elements using three different methods: the plane-wave, the direct computation, and the WF interpolation method. With the interpolated e-d matrix elements in hand, we converge the e-d RTs using importance sampling with a grid generated according to the Cauchy distribution. Lastly, we show preliminary results for electron carrier mobility as a function of doping concentration in phosphorus-doped silicon.

4.2 Electron-Defect Matrix Elements for Ionized Impurities

We model the ionized impurity as a point charge generating a Yukawa potential

$$V(r) = \frac{Ze}{\epsilon_{\mathrm{s}} r} e^{-q_0 r}, \tag{4.1}$$

where Z is the number of charges carried by the impurity, e is the electron charge, and ϵ_{s} is the dielectric constant of the material of interest. The inverse of the

screening wavelength, q_0, due to the screening from extrinsic charge carriers, is defined as

$$q_0 = \sqrt{\frac{4\pi n_c e^2}{\epsilon_\mathrm{s} k_\mathrm{B} T}}, \tag{4.2}$$

where n_c is the extrinsic carrier concentration, k_B is the Boltzmann constant, and T is the temperature. The Fourier coefficients of the screened Coulomb potential for wavevector q is

$$\tilde{V}(q) = \frac{4\pi}{\epsilon_\mathrm{s}\Omega_\mathrm{uc}(q^2 + q_0^2)}, \tag{4.3}$$

where Ω_uc is the volume of the primitive cell. The e-d matrix elements due to a charged point defect are thus defined as

$$M_{mn}(\boldsymbol{k}', \boldsymbol{k}) = \langle m\boldsymbol{k}'|V(r)|n\boldsymbol{k}\rangle = \sum_{\boldsymbol{G}} \tilde{V}(|\boldsymbol{k}' - \boldsymbol{k} - \boldsymbol{G}|) \langle u_{m\boldsymbol{k}'}|e^{-i\boldsymbol{G}\cdot\boldsymbol{r}}|u_{n\boldsymbol{k}}\rangle_\mathrm{uc}, \tag{4.4}$$

where $|n\boldsymbol{k}\rangle$ is the Bloch state with a band index n and a crystal momentum $\boldsymbol{k}$, $\boldsymbol{G}$ is the reciprocal lattice vectors of the primitive cell, and $u_{n\boldsymbol{k}}$ is the periodic part of the Bloch wave function $|n\boldsymbol{k}\rangle$. The subscript uc in the last term means that the intergation is performed in the primitive cell.

The e-d RTs $\tau_{n\boldsymbol{k}}(T)$, equal to the inverse of the scattering rates $\Gamma_{n\boldsymbol{k}}$, are computed within the lowest-order Born approximation [19, 26, 27],

$$\Gamma_{n\boldsymbol{k}}(T) = \tau^{-1}_{n\boldsymbol{k}}(T) = \frac{2\pi}{\hbar}\frac{n_\mathrm{at}C_\mathrm{d}}{N_{\boldsymbol{k}'}} \sum_{m\boldsymbol{k}'} |M_{mn}(\boldsymbol{k}', \boldsymbol{k})|^2 (1 - \cos\theta_{m\boldsymbol{k}',n\boldsymbol{k}})\delta(\varepsilon_{m\boldsymbol{k}'} - \varepsilon_{n\boldsymbol{k}}), \tag{4.5}$$

where $\hbar$ is the reduced Planck constant, n_at is the number of atoms in the primitive cell, C_d is the dimensionless impurity concentration, $N_{\boldsymbol{k}'}$ is the number of BZ $\boldsymbol{k}'$-points used in the sum, $\theta_{m\boldsymbol{k}',n\boldsymbol{k}}$ is the angle between the band velocity $\mathbf{v}_{m\boldsymbol{k}'}$ and the velocity $\mathbf{v}_{n\boldsymbol{k}}$, and $\varepsilon_{n\boldsymbol{k}}$ is the energy of the Bloch state $|n\boldsymbol{k}\rangle$. The geometric factor $(1 - cos\theta_{m\boldsymbol{k}',n\boldsymbol{k}})$ is typically used in the literature to take backscattering into account in transport calculations involving elastic e-d scattering. A more rigorous way to include backscattering is solving the linearized Boltzmann equation, for example with an iterative approximation, which will be investigated in the future work. Note that in this formalism, the RTs due to ionized impurity scattering depend on temperature, unlike the neutral impurity scattering case discussed in Chapter 2 and 3. The Dirac delta function is approximated as a Gaussian with small broadening, $\delta(x) = e^{-x^2/2\eta^2}/\sqrt{2\pi}\eta$, where η is the energy broadening.

The carrier mobility μ due to ionized impurity scattering can be computed using the linearized Boltzmann equation within the relaxation times approximation (RTA),

$$\mu_{\alpha\beta}(T) = \frac{e}{n_c}\int_{-\infty}^{+\infty} dE[-\partial f(T,E)/\partial E] \times \Sigma_{\alpha\beta}(E), \tag{4.6}$$

where $f(T,E)$ is the Fermi-Dirac distribution function and $\Sigma(T,E)$ the transport distribution function, defined as:

$$\Sigma_{\alpha\beta}(T,E) = \frac{2}{\Omega_{\text{uc}}}\sum_{n\boldsymbol{k}} \tau_{n\boldsymbol{k}}(T)\mathbf{v}^{\alpha}_{n\boldsymbol{k}}\mathbf{v}^{\beta}_{n\boldsymbol{k}}\delta\left(E - \varepsilon_{n\boldsymbol{k}}\right), \tag{4.7}$$

where α and β are Cartesian directions. The transport distribution function is computed with a tetrahedron integration method [25] using the converged e-d RTs and Wannier-interpolated band velocities $\mathbf{v}_{n\boldsymbol{k}}$ [28, 29].

4.3 Wannier Interpolation Method for Charged Defects

The WF $|j\boldsymbol{R}\rangle$ with a Wannier index j and centered at the Bravais lattice vector $\boldsymbol{R}$ can be constructed from the Bloch states,

$$|j\boldsymbol{R}\rangle = \frac{1}{N_{\boldsymbol{k}}}\sum_{n\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} U_{nj,\boldsymbol{k}}\,|n\boldsymbol{k}\rangle\,, \tag{4.8}$$

where $U_{nj,\boldsymbol{k}}$ is the unitary matrix that minimizes the spatial spread of the WF. Similarly, the Bloch state $|n\boldsymbol{k}\rangle$ can be constructed from the WFs,

$$|n\boldsymbol{k}\rangle = \sum_{j\boldsymbol{R}} e^{i\boldsymbol{k}\cdot\boldsymbol{R}} U^{\dagger}_{jn,\boldsymbol{k}}\,|j\boldsymbol{R}\rangle\,. \tag{4.9}$$

In addition to expressing the e-d matrix elements in the Bloch states, we can also express the e-d matrix elements in the WF basis,

$$M_{ij}(\boldsymbol{R}',\boldsymbol{R}) = \langle i\boldsymbol{R}'|V(r)|j\boldsymbol{R}\rangle\,. \tag{4.10}$$

Using the relations between the Bloch states and the WFs, we can show that the e-d matrix elements in the WF basis are the generalized double Fourier transformation of the e-d matrix element in the Bloch states,

$$M(\boldsymbol{R}',\boldsymbol{R}) = \left(\frac{1}{N_{\boldsymbol{k}_\text{c}}}\right)^2 \sum_{\boldsymbol{k}'_\text{c},\boldsymbol{k}_\text{c}} e^{i(\boldsymbol{k}'_\text{c}\cdot\boldsymbol{R}' - \boldsymbol{k}_\text{c}\cdot\boldsymbol{R})} U^{\dagger}_{\boldsymbol{k}'_\text{c}} M(\boldsymbol{k}'_\text{c},\boldsymbol{k}_\text{c}) U_{\boldsymbol{k}_\text{c}}, \tag{4.11}$$

where we omit the band index for simplicity. Notice that the e-d matrix elements are computed on a coarse BZ $\boldsymbol{k}_\text{c}$-grid. Through the inverse Fourier transformation,

we can interpolate any e-d matrix elements in the Bloch states for any momentum pair $(\boldsymbol{k}'_\mathrm{f}, \boldsymbol{k}_\mathrm{f})$ on a fine BZ $\boldsymbol{k}_\mathrm{f}$-grid,

$$M(\boldsymbol{k}'_\mathrm{f}, \boldsymbol{k}_\mathrm{f}) = \sum_{\boldsymbol{R}',\boldsymbol{R}} e^{-i(\boldsymbol{k}'_\mathrm{f}\cdot\boldsymbol{R}' - \boldsymbol{k}_\mathrm{f}\cdot\boldsymbol{R})} U_{\boldsymbol{k}'_\mathrm{f}} M(\boldsymbol{R}', \boldsymbol{R}) U^\dagger_{\boldsymbol{k}_\mathrm{f}}, \tag{4.12}$$

where the $U_{\boldsymbol{k}_\mathrm{f}}$ can be obtained from the Hamiltonian in the WF basis, $H(\boldsymbol{R})$ [28]. Essential to the successful application of this approach is that the matrix elements decay rapidly in real space. However, for charged defects the e-d matrix elements in the Wannier basis do not decay rapidly enough in real space due to the long-range nature of the screened Coulomb potential. As a result, the summation in Eq. (4.12) requires a large set of lattice vectors $\boldsymbol{R}$.

To overcome this challenge, we use a smooth phase approximation to remove the so-called long-range part of the screened Coulomb potential from the e-d matrix elements before performing the double-Fourier transformation in Eq. (4.11). The long-range part of the e-d matrix element on the coarse BZ $\boldsymbol{k}_\mathrm{c}$-grid is defined as

$$M^{\mathrm{LR}}_{mn}(\boldsymbol{k}'_\mathrm{c}, \boldsymbol{k}_\mathrm{c}) = \sum_{\boldsymbol{G}} \tilde{V}(|\boldsymbol{k}' - \boldsymbol{k} - \boldsymbol{G}|)(U_{\boldsymbol{k}'_\mathrm{c}} U^\dagger_{\boldsymbol{k}_\mathrm{c}})_{mn}. \tag{4.13}$$

The remaining part of the e-d matrix elements, after the long-range part is removed, is defined as the short-range part of the e-d matrix elements,

$$M^{\mathrm{SR}}_{mn}(\boldsymbol{k}'_\mathrm{c}, \boldsymbol{k}_\mathrm{c}) = M_{mn}(\boldsymbol{k}'_\mathrm{c}, \boldsymbol{k}_\mathrm{c}) - M^{\mathrm{LR}}_{mn}(\boldsymbol{k}'_\mathrm{c}, \boldsymbol{k}_\mathrm{c}), \tag{4.14}$$

which is used in the WF interpolation such that we can use a small set of lattice vectors $\boldsymbol{R}$ to interpolate the short-range part of the e-d matrix elements and add back the corresponding long-range part of the e-d matrix elements on the fine BZ $\boldsymbol{k}_\mathrm{f}$-grid.

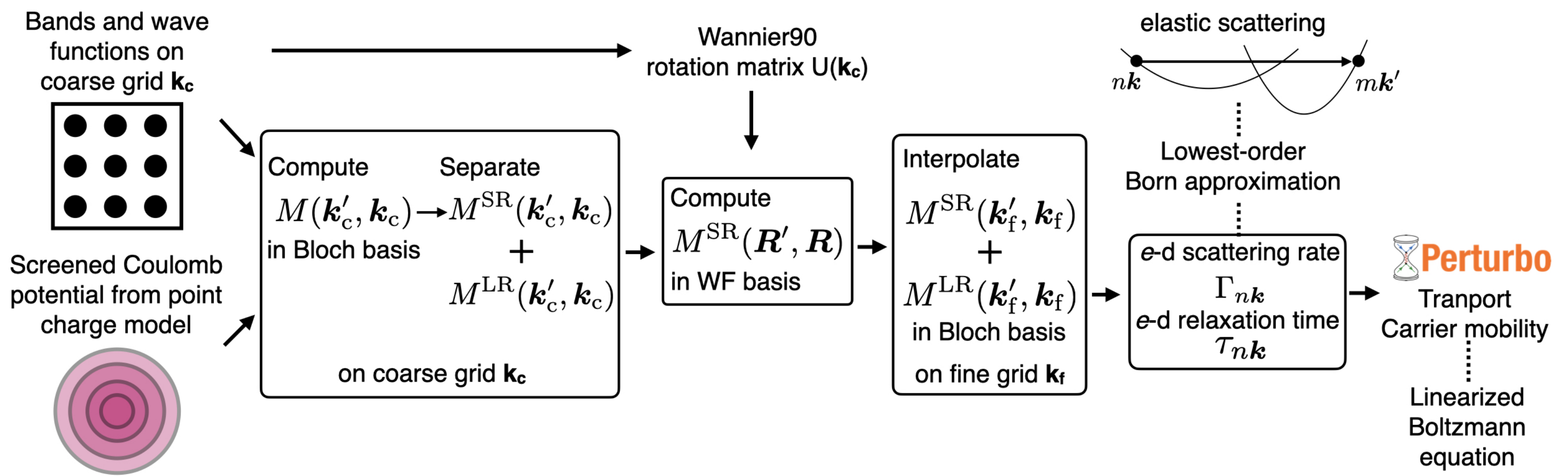

Figure 4.1: Workflow for interpolating e-d matrix elements for ionized impurity scattering. The details of the workflow are explained in the main text.

4.4 Workflow and Computation Details

Figure 4.1 shows the workflow of interpolating the e-d matrix elements due to ionized impurity scattering where the charged defect is modelled as a point charge. We obtain for Bloch states on a coarse BZ $\boldsymbol{k}_{\mathrm{c}}$-grid the wave functions and the energies, which are used to construct maximally localized WFs using Wannier90 [29]. Using the wave functions and the screened Coulomb potential, the e-d matrix elements on the coarse BZ $\boldsymbol{k}_{\mathrm{c}}$-grid are computed. Subsequently, the long-range part of the e-d matrix elements is removed, while the short-range part is transformed from the Bloch state into the WF basis using the unitary matrices $U(\boldsymbol{k}_{\mathrm{c}})$. The short-range part of the e-d matrix elements in the WF basis is used to interpolate any short-range part of the e-d matrix elements in the Bloch states on a fine $\boldsymbol{k}_{\mathrm{f}}$-grid. The long-range part of the e-d matrix elements is added back with the short-range part to form the e-d matrix elements used in the e-d scattering rate or relaxation time calculations. With the converged e-d RTs or scattering rates, the carrier mobility can be computed using PERTURBO [6].

We apply our workflow to silicon. The ground state of the primitive cell is computed using density functional theory (DFT) with the local density approximation. We use QUANTUM ESPRESSO [30], a DFT code using a plane wave basis set and pseudopotentials, and use the norm-conserving pseudopotential for silicon from the Pseudo Dojo pseudopotential repository [31]. We compute the total energy and the electron density using a lattice constant of 5.43 Å, a kinetic energy cutoff of 40 Ry, and a $12\times12\times12$ $\boldsymbol{k}$-point grid. The total energy is converged to within 10 meV/atom. We use the converged electron density to compute the wave functions and electronic energy eigenvalues on coarse $\boldsymbol{k}_{\mathrm{c}}$-grids of $4\times4\times4$ and $10\times10\times10$ $\boldsymbol{k}_{\mathrm{c}}$-points. The electronic energies and wave functions are used as input to construct maximally localized WFs in Wannier90 [29] and to obtain the corresponding Wannier unitary matrices. These quantities are used to compute and interpolate the e-d matrix elements for the screened Coulomb potential associated with charged defects. The dielectic constant for silicon is taken to be 11.3.

4.5 Electron-Defect Matrix Elements for Charged Defects

To compare the matrix elements computed using different approaches, we compute gauge-invariant e-d matrix elements to remove the gauge dependence on the Wannier

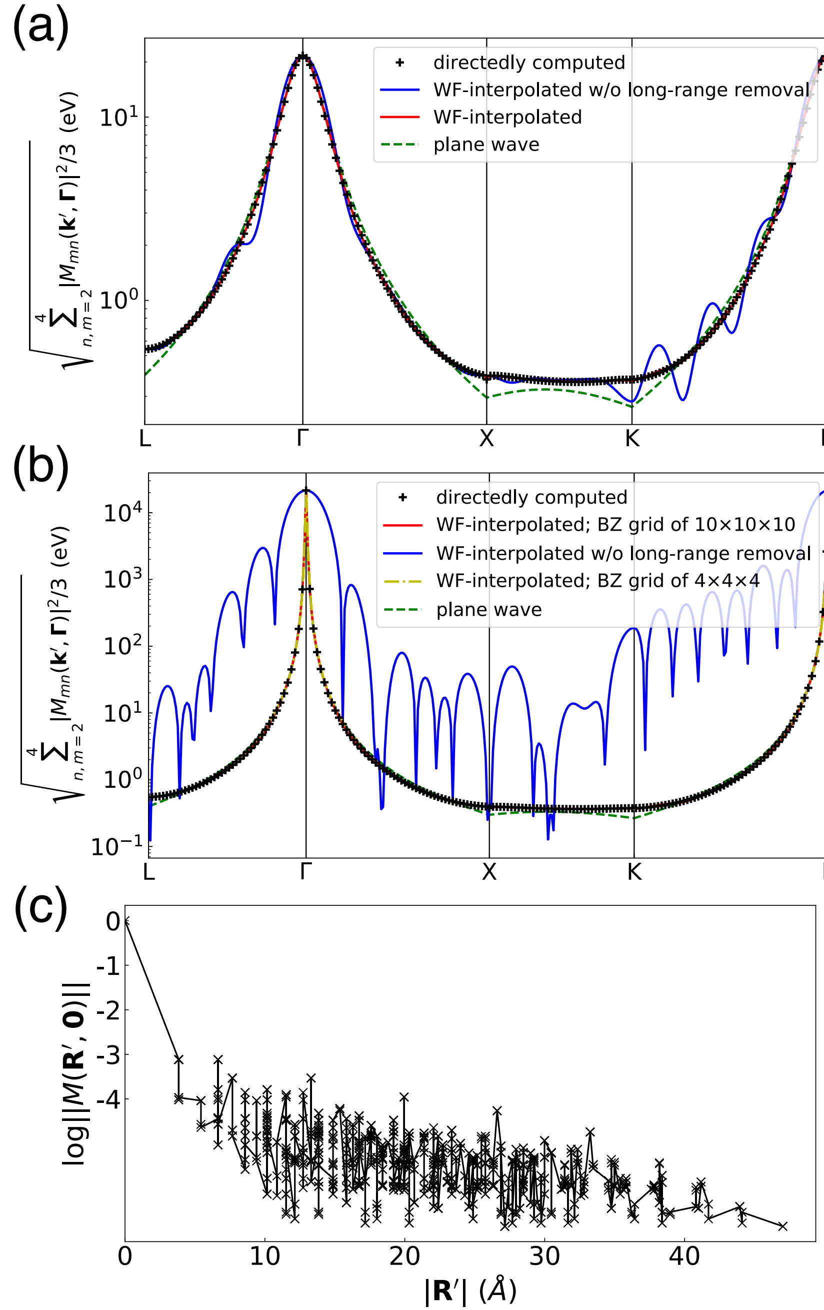

Figure 4.2: Electron-defect matrix elements for the Yukawa potential. Shown are gauge-invariant matrix elements along high symmetric lines for two carrier concentrations of 10^{19} cm^{-3} in panel (a) and 10^{16} cm^{-3} in panel (b) at 100 K; the initial $\boldsymbol{k}$-point is fixed at the Γ point. (c) Spatial behavior of the short-range e-d matrix elements in the WF basis for a carrier concentration of 10^{16} cm^{-3}.

unitary matrices:

$$|M(\boldsymbol{k}', \boldsymbol{k})| = \sqrt{\sum_{m,n=1}^{N_b} |M_{mn}(\boldsymbol{k}', \boldsymbol{k})|^2 / N_b}, \tag{4.15}$$

where N_b is the number of bands involved in the summation. The number of bands used in the summation is usually equal to the number of degenerate bands.

Figure 4.2(a) shows the invariant e-d matrix elements computed by fixing the initial k-point at Gamma and for the final momentum $\boldsymbol{k}'$ along high symmetry lines. The calculation is carried out for an extrinsic carrier concentration of 10^{19} cm^{-3} and a temperature of 100 K. The matrix elements are computed using the direct computation, WF interpolation, and plane-wave method. The two interpolated matrix element curves are computed starting from a coarse BZ $\boldsymbol{k}_\mathrm{c}$-grid of $10 \times 10 \times 10$ points; in one curve, the long-range part was removed before interpolation while in the other curve it was not. The interpolated matrix elements computed without removing the long-range part cannot reproduce the directly computed results, even though the coarse BZ grid is dense enough to capture the screened Coulomb potential. However, after removing the long-range part in the interpolation procedure, the interpolated matrix elements successfully reproduce the directly computed results. Here we also provide the matrix elements computed using plane waves,

$$M_\mathrm{PW}(\boldsymbol{k}', \boldsymbol{k}) = \tilde{V}(|\boldsymbol{k}' - \boldsymbol{k}|), \tag{4.16}$$

which is just the Fourier coefficient of the screened Coulomb potential at wavevector $|\boldsymbol{k}' - \boldsymbol{k}|$. Notice that the plane-wave matrix elements are only within the same band because the matrix elements depend only on crystal momentum, but not on band index. Interestingly, for electrons the matrix elements computed with plane waves reproduce well the full calculation with Bloch states when the initial and final momenta are close in value because the cell-periodic part of the Bloch states do not differ significantly for small changes in the crystal momentum [see Eq. (4.4)]. Nevertheless, the plane-wave results differ when the separation of the crystal momenta increases, as is seen in Fig. 4.2(a) along the X−K direction.

Figure 4.2(b) gives the invariant e-d matrix elements for a different carrier concentration of 10^{16} cm^{-3} at 100 K for the initial crystal momentum fixed at the Γ point. Similar to the previous case, the interpolated matrix elements without removing the long-range part cannot reproduce the directly computed results. As the spatial range of the screened Coulomb potential is greater at lower carrier concentrations due to less effective screening [see Eq. (4.3)], not removing the long-range part before interpolation leads to large errors, as seen in Fig. 4.2(b). After removing the long-range part, the interpolated matrix elements actually reproduce the directly computed results. Surprisingly, the interpolated matrix elements using a smaller

coarse BZ $\boldsymbol{k}_\mathrm{c}$-grid of $4\times4\times4$, with the long-range part removed before interpolation, can also reproduce the directly computed results. The plane-wave matrix elements can reproduce the matrix elements well around the Γ point similar to the previous case.

To analyze the spatial decay of the short-range e-d matrix elements in the WF basis, we define for each pair of lattice vectors $\boldsymbol{R}'$ and $\boldsymbol{R}$ the norm of the matrix $M(\boldsymbol{R}',\boldsymbol{R})$ as the absolute value of the e-d matrix elements as $||M(\boldsymbol{R}',\boldsymbol{R})|| = \max_{ij}|M_{ij}(\boldsymbol{R}',\boldsymbol{R})|$. Figure 4.2(c) shows the spatial decay of $||M(\boldsymbol{R}',\boldsymbol{R})||$ due to a point charged defect in silicon, at 100 K temperature and for a carrier concentration of 10^{16} cm^{-3}, as a function of $|\boldsymbol{R}'|$ while keeping $\boldsymbol{R} = 0$. The matrix elements decay exponentially within a few lattice constants, which indicates that the interpolation approach we have developed for charged defects works well.

4.6 Electron-Defect Relaxation Times due to Charged Defects

With the interpolation method for matrix elements in hand, we can converge the RTs with respect to the number of points in the BZ $\boldsymbol{k}_\mathrm{f}$-grid for different energy broadening values. Naively, one should use a broadening value as small as possible to simulate the Dirac delta function in Eq. (4.5); however, the smaller the broadening value, the denser the BZ grid needed to numerically converge the integral in the RT formula [see Eq. (4.5)]. To reduce the computational cost, we start with a relatively large broadening value (say, 15 meV) and converge the RTs with respect to the number of $\boldsymbol{k}_\mathrm{f}$-points. We then reduce the broadening and converge the RTs again using a denser BZ $\boldsymbol{k}_\mathrm{f}$-grid. The RTs are converged when further decreasing the broadening does not change their values.

Figure 4.3(a) shows electron and hole RTs at two arbitrary $\boldsymbol{k}$-points, respectively, using two approaches − the interpolation and plane-wave methods − to compute the e-d matrix elements at two broadening values η of 1 and 2 meV. The matrix elements that significantly contribute to the RTs are those at final momenta $\boldsymbol{k}'$ close to the initial $\boldsymbol{k}$-point because of the large contribution from the screened Coulomb potential at $q = |\boldsymbol{k}' - \boldsymbol{k}|$ close to zero. Due to the large value of the integrand at $q = 0$, using a regular BZ $\boldsymbol{k}'$-grid will converge the RTs slowly. To improve the calculation, we use a Cauchy distribution to generate random $\boldsymbol{k}_\mathrm{f}$-points around a given $\boldsymbol{k}$ point on the regular BZ $\boldsymbol{k}_\mathrm{f}$-grid. The points in the double-grid are weighted appropriately to preserve the weighting factor $\frac{1}{N_{\boldsymbol{k}'}}$ for the uniform grid in Eq. (4.5). Using the Cauchy distribution, we converge the RTs very effectively with respect

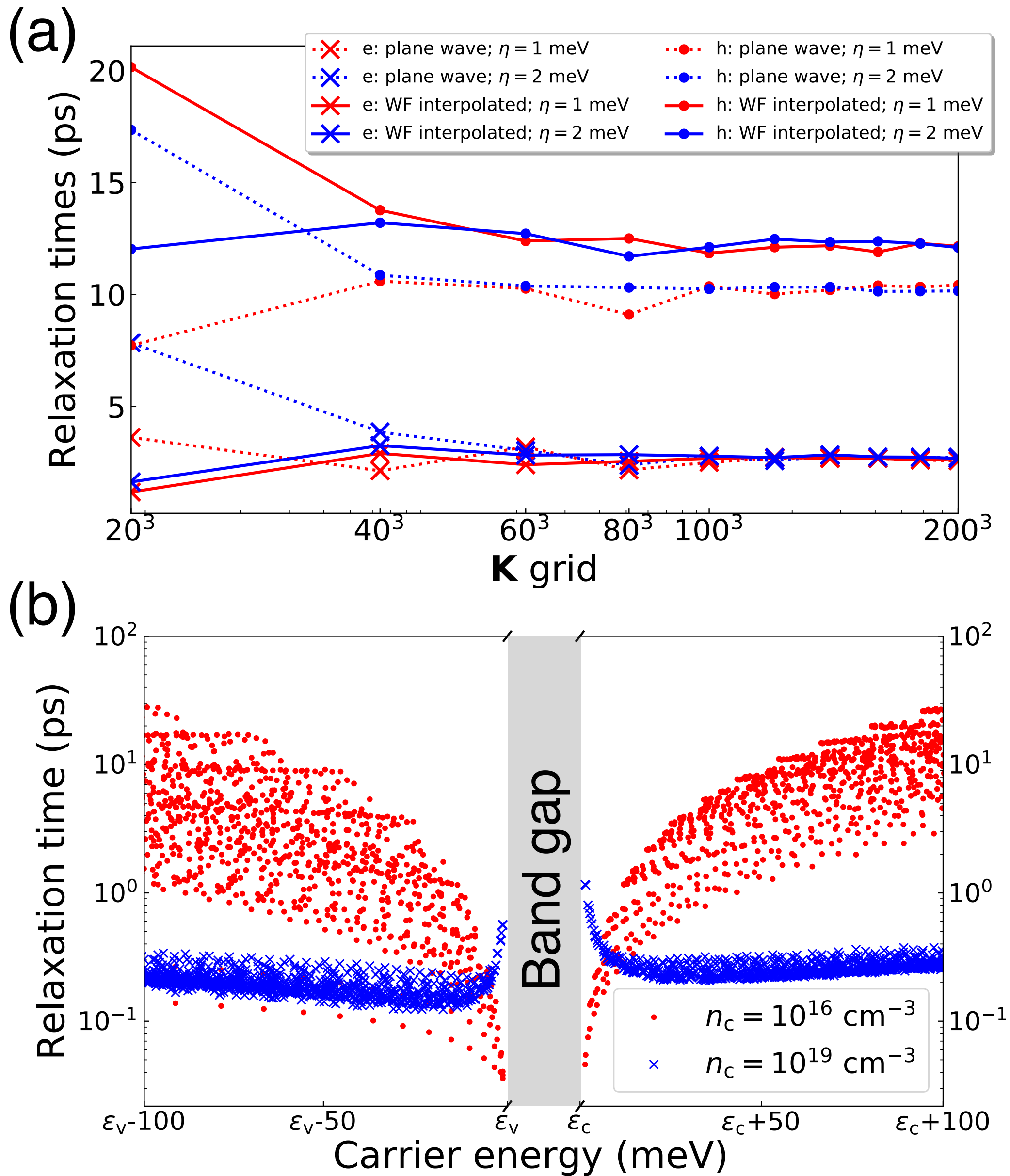

Figure 4.3: Relaxation times due to ionized impurity scattering at 100 K. (a) The convergence of the RTs as a function of the BZ $\boldsymbol{k}_\mathrm{f}$-grids. Electron and hole are labelled as e and h, respectively. (b) Carrier RTs near band edges in silicon. The RTs are computed using the interpolated e-d matrix elements.

to the number of $\boldsymbol{k}_\mathrm{f}$-points. For instance, the electron RTs using WF interpolation converge to the same value for two different broadening values when using a dense BZ grid with more than 1 million $\boldsymbol{k}_\mathrm{f}$-points. The RTs computed using the plane waves also converge to the same RT value because there is only one band (the lowest conduction band) involved for the electron while interband scatterings are

neglected. However, the hole RTs converge using 1 million $\boldsymbol{k}_{\mathrm{f}}$-points to the same value for each e-d matrix calculation method (the interpolation and plane-wave methods), but the RTs differ in the two methods because the Wannier interpolation method takes interband scattering into account.

Figure 4.3(b) shows the RTs for electrons and holes near the band edges for two different doping concentrations of 10^{16} and 10^{19} cm^{-3} at 100 K, assuming in both cases that all the dopants are fully ionized. The carrier RTs for the two concentrations vary with carrier energy and show different trends. For the lower carrier concentration, the RTs increase away from the band edges by almost three orders of magnitude within 100 meV, while for the higher concentration, the RTs decrease from the band edges to a constant within 20 meV, beyond which the RTs increase slightly. The RT trend for the higher carrier concentration is similar to that for neutral impurity, a reasonable result given that the range of the screened Coulomb potential becomes smaller when the extrinsic carrier concentration is greater.

4.7 Electron Carrier Mobility as a Function of Doping Concentration in Silicon

With the powerful machinery we have developed to treat the e-d matrix elements for ionized impurity scattering, we obtain the preliminary results for carrier mobility as a function of doping concentration. The carrier mobility μ is computed using

$$\frac{1}{\mu} = \frac{1}{\mu_{\mathrm{e-ph}}} + \frac{1}{\mu_{\mathrm{e-d}}}, \tag{4.17}$$

where $\mu_{\mathrm{e-ph}}$ is the phonon-limited carrier moblity and $\mu_{\mathrm{e-d}}$ the defect-limited carrier mobility. For this very preliminary result, we use a constant value of the phonon-limited mobility at 300 K (1450 cm^{2}V^{-1}s^{-1} [34]), although in future works we plan to include the fully temperature dependent phonon-limited mobility. We compute the defect-limited electron mobility as a function of carrier concentration at 300 K using the matrix elements computed with the plane-wave method to reduce the computational cost. Figure 4.4 shows the preliminary results for electron carrier mobility as a function of doping concentration at 300 K. We show two sets of computed carrier mobilities, one obtained by assuming fully ionized impurities, and thus a carrier concentration equal to the doping (curve labeled "fully ionized"), and the other computed by assuming that dopants are partially ionized and the carrier concentration is given by

$$n_c = n_{\mathrm{d}} e^{-E_{\mathrm{ion}}/k_{\mathrm{B}}T}, \tag{4.18}$$

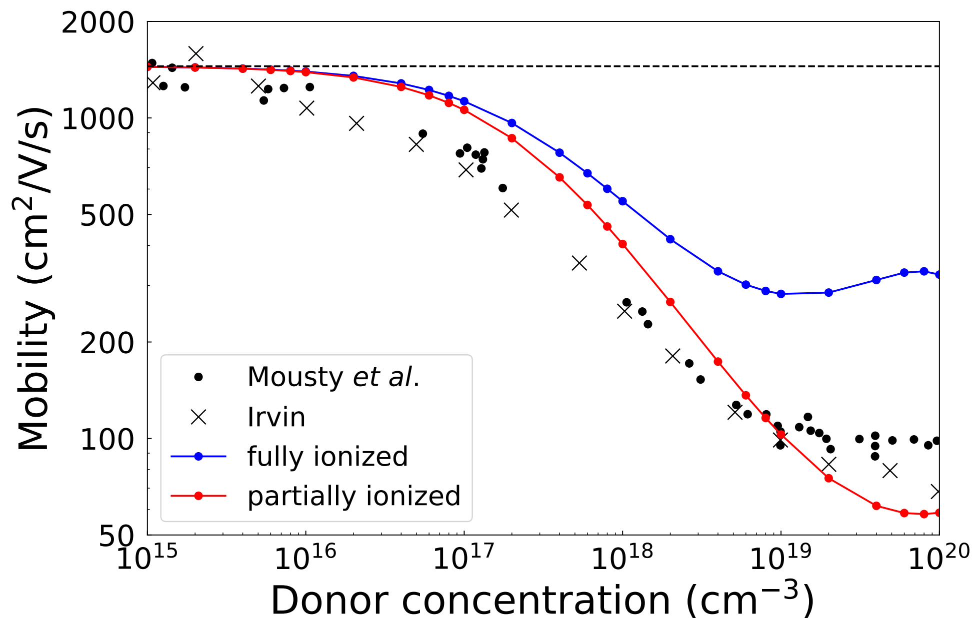

Figure 4.4: Electron mobility in Phosphorus-doped silicon as a function of doping concentration at 300 K. The calculation includes both e-ph and e-d interactions, the latter due to ionized impurity scattering. The dashed line is the phonon-limited carrier mobility at 300 K. Also shown are the experimental data from the literature [32, 33].

where n_d is the doping concentration and E_{ion} is the ionization energy of phosphorus in silicon, 45 meV [35]. For the fully ionized case, the electron carrier mobility does not agree well with the experiment, especially at higher doping concentrations. For the partially ionized case, the order of magnitude of the computed carrier mobility is within the range of the experimental measurement [32, 33], and the trend is in reasonable agreement with experiment although refinements are needed to improve the results. Future work will further examine the deviation of the computed carrier mobility from the experimental data. Possible reasons for the discrepancy include errors in the effective mass for the DFT computed band structure of silicon, the temperature dependent carrier concentrations not being taken properly into account, or the geometric factor we employed to account for backscattering, which is only a crude approximation. Future work on charged defect will also include e-ph scattering to obtain a more accurate temperature dependence of the carrier mobility.

4.8 Summary

In summary, we have developed an interpolation method for e-d matrix elements due to ionized impurity scattering. The interpolated matrix elements using WFs match well with the directly computed ones. We also systematically converge the e-d RTs with respect to the number of $\boldsymbol{k}$-points in the BZ using a Cauchy distribution. The convergence has not been shown before likely due to the prohibitive computational cost of the e-d matrix elements, which can now be computed efficiently using unit cell calculations and the interpolation method developed in this work.

The method for e-d interactions can be integrated into PERTURBO, a software developed by our group to investigate from first principles e-ph interactions, charge transport, and ultrafast electron dynamics. This effort and my contributions to the PERTURBO code are described in Chapter 5 of the thesis. Including the e-d routines into the open source Perturbo code would make the methods presented in this thesis available to the community and dramatically enhance their impact.

References

[1] S. Datta, *Electronic transport in mesoscopic systems* (Cambridge University Press, 1997).

[2] C. Jacoboni, *Theory of electron transport in semiconductors: A pathway from elementary physics to nonequilibrium green functions*, Vol. 165 (Springer Science & Business Media, 2010).

[3] V. F. Gantmakher and Y. B. Levinson, *Carrier scattering in metals and semiconductors* (Elsevier, 2012).

[4] D. A. Neamen, *Semiconductor physics and devices: Basic principles* (New York, NY: McGraw-Hill, 2012).

[5] S.-L. Li, K. Tsukagoshi, E. Orgiu, and P. Samorì, Chem. Soc. Rev. **45**, 118 (2016).

[6] J.-J. Zhou, J. Park, I.-T. Lu, I. Maliyov, X. Tong, and M. Bernardi, arXiv:2002.02045 (2020).

[7] M. H. Evans, X.-G. Zhang, J. D. Joannopoulos, and S. T. Pantelides, Phys. Rev. Lett. **95**, 106802 (2005).

[8] O. D. Restrepo, K. Varga, and S. T. Pantelides, Appl. Phys. Lett. **94**, 212103 (2009).

[9] V. Lordi, P. Erhart, and D. Åberg, Phys. Rev. B **81**, 235204 (2010).

[10] K. Krishnaswamy, B. Himmetoglu, Y. Kang, A. Janotti, and C. G. Van de Walle, Phys. Rev. B **95**, 205202 (2017).

[11] S. Poncé, E. R. Margine, and F. Giustino, Phys. Rev. B **97**, 121201(R) (2018).

[12] J. Wang, A. Rahman, A. Ghosh, G. Klimeck, and M. Lundstrom, IEEE Trans. Electron Devices **52**, 1589 (2005).

[13] K. Nehari, M. Lannoo, F. Michelini, N. Cavassilas, M. Bescond, and J. L. Autran, Appl. Phys. Lett. **93**, 092103 (2008).

[14] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).

[15] B. Yan and C. Felser, Annu. Rev. Condens. Matter Phys. **8**, 337 (2017).

[16] N. P. Armitage, E. J. Mele, and A. Vishwanath, Rev. Mod. Phys. **90**, 015001 (2018).

[17] B. K. Ridley, *Quantum processes in semiconductors* (Oxford University Press, 2013).

[18] K. Kaasbjerg, J. H. J. Martiny, T. Low, and A.-P. Jauho, Phys. Rev. B **96**, 241411(R) (2017).

[19] I.-T. Lu, J.-J. Zhou, and M. Bernardi, Phys. Rev. Mater. **3**, 033804 (2019).

[20] K. Kaasbjerg, Phys. Rev. B **101**, 045433 (2020).

[21] I.-T. Lu, J. Park, J.-J. Zhou, and M. Bernardi, Npj Comput. Mater. **6**, 17 (2020).

[22] B. Qiu, Z. Tian, A. Vallabhaneni, B. Liao, J. M. Mendoza, O. D. Restrepo, X. Ruan, and G. Chen, Europhys. Lett. **109**, 57006 (2015).

[23] C. Verdi and F. Giustino, Phys. Rev. Lett. **115**, 176401 (2015).

[24] J. Sjakste, N. Vast, M. Calandra, and F. Mauri, Phys. Rev. B **92**, 054307 (2015).

[25] J. J. Zhou and M. Bernardi, Phys. Rev. B **94**, 201201(R) (2016).

[26] G. D. Mahan, *Many-particle physics* (Springer US, 2000).

[27] H. Bruus and K. Flensberg, *Many-body quantum theory in condensed matter physics: An introduction* (Oxford University Press, 2004).

[28] J. R. Yates, X. Wang, D. Vanderbilt, and I. Souza, Phys. Rev. B **75**, 195121 (2007).

[29] A. A. Mostofi, J. R. Yates, G. Pizzi, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, Comput. Phys. Commun. **185**, 2309 (2014).

[30] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, J. Phys.: Condens. Matter **21**, 395502 (2009).

[31] M. J. Van Setten, M. Giantomassi, E. Bousquet, M. J. Verstraete, D. R. Hamann, X. Gonze, and G.-M. Rignanese, Comput. Phys. Commun. **226**, 39 (2018).

[32] F. Mousty, P. Ostoja, and L. Passari, J. Appl. Phys. **45**, 4576 (1974).

[33] J. C. Irvin, Bell Syst. Tech. J. **41**, 387 (1962).

[34] C. Jacoboni, C. Canali, G. Ottaviani, and A. Alberigi Quaranta, Solid State Electron. **20**, 77 (1977).

[35] S. O. Kasap, *Principles of electronic materials and devices* (Tata McGraw-Hill, 2006).

Chapter 5

THE PERTURBO OPEN SOURCE CODE AND ITS ELECTRON-DEFECT ROUTINES

5.1 Introduction

Understanding the dynamical processes involving electrons, lattice vibrations, and atomic defects in the solid state is key to developing the next generation of materials and devices [1, 2]. Due to the increasing complexity of functional materials, there is a critical need for computational tools that can take into account the atomic and electronic structure of materials and make quantitative predictions on their physical properties. The vision behind the PERTURBO open source code developed in the Bernardi group is to provide a unified platform and a validated code that can be applied broadly to compute the interactions, transport, and ultrafast dynamics of electrons and excited states in materials [3]. The goal is to facilitate basic scientific discoveries in materials and devices by advancing microscopic understanding of carrier dynamics, while creating a sustainable software element able to address the demands of the computational physics community.

During my PhD work, I have co-developed and tested parts of the PERTURBO code, and am currently working to include the electron-defect (e-d) routines in the code to make them available to the community. This step would dramatically enhance the impact of the work discussed in this thesis as various research groups will be able to use and expand the methods developed in this thesis. The PERTURBO code builds on established first-principles methods. It uses density functional theory (DFT) and density functional perturbation theory (DFPT) [4] as a starting point for computing electron dynamics. It reads the output of DFT and DFPT calculations, for now from the Quantum Espresso (QE) code [5, 6], and uses this data to compute electron interactions, charge transport, and ultrafast dynamics. The current distribution (version 1.0) focuses on electron-phonon (e-ph) interactions and the related phonon-limited transport properties [7], including the electrical conductivity, mobility, and the Seebeck coefficient. A future version will include the e-d methods developed in this thesis.

The transport module of PERTURBO enables accurate calculations of charge transport in a wide range of functional materials. In its most basic workflow, PERTURBO

computes the conductivity and mobility as a function of temperature and carrier concentration, either within the relaxation time approximation (RTA) or with an iterative solution of the linearized Boltzmann transport equation (BTE) [8, 9]. The ultrafast dynamics module explicitly evolves in time the electron BTE (while keeping the phonon occupations fixed), enabling investigations of the ultrafast electron dynamics starting from a given initial electron distribution [10]. The routines can carry out these calculations in metals, semiconductors, insulators, and 2D materials. An efficient implementation of long-range e-ph interactions is employed for polar bulk and 2D materials. Materials with spin-orbit coupling (SOC) are treated using fully relativistic pseudopotentials [9, 11]. Both norm-conserving and ultrasoft pseudopotentials are supported. Quantities related to e-ph interactions can be easily obtained, stored, and analyzed. The e-d portion of the code under development will follow a similar implementation style.

PERTURBO is implemented in modern Fortran with a modular code design. The code is highly efficient thanks to its hybrid MPI (Message Passing Interface) and OpenMP (Open Multi-Processing) parallelization. It can run on record-large unit cells with up to at least 50 atoms [12], and its performance scales up to thousands of CPU cores. It conveniently writes files using the HDF5 format, and is suitable for both high-performance supercomputers and smaller computer clusters.

The current version (v1.0) of PERTURBO mainly supports e-ph calculations. In this chapter, we discuss how to integrate *ab initio* electron-defect (e-d) calculations with the current PERTURBO code. We review how to compute the defect-limited carrier mobility using the BTE and then discuss the workflow for computing the e-d interaction matrix elements for both neutral and charged defects. Finally, we mention some technical details on the parallelization of the e-d codes.

5.2 Boltzmann Transport Equation for Electron-Defect Scattering

The current release of PERTURBO can compute charge transport and ultrafast dynamics in the framework of the semiclassical BTE. The BTE describes the flow of the electron occupations $f_{n\boldsymbol{k}}(\boldsymbol{r},t)$ in the phase-space variables of relevance in a periodic system, the crystal momentum $\boldsymbol{k}$, and spatial coordinate $\boldsymbol{r}$:

$$\frac{\partial f_{n\boldsymbol{k}}(\boldsymbol{r},t)}{\partial t} = -\left[\nabla_{\boldsymbol{r}} f_{n\boldsymbol{k}}(\boldsymbol{r},t)\cdot \mathbf{v}_{n\boldsymbol{k}} + \hbar^{-1}\nabla_{\boldsymbol{k}} f_{n\boldsymbol{k}}(\boldsymbol{r},t)\cdot \boldsymbol{F}\right] + \mathcal{I}\left[f_{n\boldsymbol{k}}\right], \tag{5.1}$$

where n is the band index and $\mathbf{v}_{n\boldsymbol{k}}$ are band velocities. The time evolution of the electron occupations is governed by the so-called drift term due to external fields $\boldsymbol{F}$ and the collision term $\mathcal{I}\left[f_{n\boldsymbol{k}}\right]$, which captures electron scattering processes due to

phonons or other mechanisms [13]. In PERTURBO, the fields are assumed to be slowly varying and the material homogeneous, so $f_{n\boldsymbol{k}}$ does not depend on the spatial coordinates and its spatial dependence is not computed explicitly.

The collision integral $\mathcal{I}[f_{n\boldsymbol{k}}]$ is a sum over a large number of scattering processes in momentum space, and it is very computationally expensive because it involves Brillouin zone (BZ) integrals on fine grids. Most analytical and computational treatments simplify the scattering integral with various approximations. A common one is the RTA, which assumes that the scattering integral is proportional to the deviation $\delta f_{n\boldsymbol{k}}$ of the electron occupations from the equilibrium Fermi-Dirac distribution, $\mathcal{I}[f_{n\boldsymbol{k}}] = -\delta f_{n\boldsymbol{k}}/\tau$; the relaxation time (RT) τ is either treated as a constant empirical parameter [14, 15] or as a state-dependent quantity, $\tau_{n\boldsymbol{k}}$.

PERTURBO implements the first-principles formalism of the BTE, which employs materials properties obtained with quantum mechanical approaches, using the atomic structure of the material as the only input. The electronic structure is computed using DFT. The scattering integral implemented in PERTURBO [16], which currently considers only e-ph processes, can be extended to also take into account e-d scattering. The scattering integral due to e-d processes can be written as

$$\mathcal{I}^{e\text{-d}}[f_{n\boldsymbol{k}}] = -\frac{2\pi}{\hbar}\frac{n_{\text{at}}C_{\text{d}}}{N_{\boldsymbol{k}'}}\sum_{m\boldsymbol{k}'}|M_{mn}(\boldsymbol{k}',\boldsymbol{k})|^2\,[f_{m\boldsymbol{k}'} - f_{n\boldsymbol{k}}]\,\delta(\varepsilon_{m\boldsymbol{k}'} - \varepsilon_{n\boldsymbol{k}}), \tag{5.2}$$

where n_{at} is the number of atoms in the primitive cell, C_d is the dimensionless defect concentration, $N_{\boldsymbol{k}'}$ is the number of $\boldsymbol{k}'$-points used in the summation, and $M_{mn}(\boldsymbol{k}',\boldsymbol{k})$ is the e-d matrix element quantifying the probability amplitude for an electron to scatter from an initial state $|n\boldsymbol{k}\rangle$ to a final state $|m\boldsymbol{k}'\rangle$; here and below, $\varepsilon_{n\boldsymbol{k}}$ is the energy of the Bloch state $|n\boldsymbol{k}\rangle$.

In an external electric field, the drift and collision terms in the BTE balance out at long enough times – the field drives the electron distribution out of equilibrium, while the collisions tend to restore equilibrium. At steady state, a nonequilibrium electron distribution is reached, for which $\partial f_{n\boldsymbol{k}}/\partial t = 0$. The BTE for transport at steady state becomes

$$-\frac{e\boldsymbol{E}}{\hbar}\cdot\nabla_{\boldsymbol{k}} f_{n\boldsymbol{k}}(t) = \mathcal{I}^{e\text{-d}}[f_{n\boldsymbol{k}}]. \tag{5.3}$$

When the electric field is relatively weak, the steady-state electron distribution deviates only slightly from its equilibrium value. As is usual, we expand $f_{n\boldsymbol{k}}$ around the equilibrium Fermi-Dirac distribution, $f^0_{n\boldsymbol{k}}$, and keep only terms linear in the

electric field:

$$f_{n\boldsymbol{k}} = f_{n\boldsymbol{k}}^{0} + f_{n\boldsymbol{k}}^{1} + O\left(E^2\right) = f_{n\boldsymbol{k}}^{0} + e\boldsymbol{E}\cdot\boldsymbol{F}_{n\boldsymbol{k}}\frac{\partial f_{n\boldsymbol{k}}^{0}}{\partial \varepsilon_{n\boldsymbol{k}}} + O\left(E^2\right), \tag{5.4}$$

where $\boldsymbol{F}_{n\boldsymbol{k}}$ characterizes the first-order deviation from equilibrium of the electron distribution. We substitute Eq. (5.4) into both sides of Eq. (5.3), and obtain a linearized BTE for the distribution deviation $\boldsymbol{F}_{n\boldsymbol{k}}$ keeping only terms up to first-order in the electric field:

$$\boldsymbol{F}_{n\boldsymbol{k}} = \tau_{n\boldsymbol{k}}^{e-\mathrm{d}}\mathbf{v}_{n\boldsymbol{k}} + \frac{\tau_{n\boldsymbol{k}}^{e-\mathrm{d}}}{\mathcal{N}_{\boldsymbol{k}'}}\sum_{m\boldsymbol{k}'}\boldsymbol{F}_{m\boldsymbol{k}'}W_{m\boldsymbol{k}',n\boldsymbol{k}}, \tag{5.5}$$

where $\tau_{n\boldsymbol{k}}^{e-\mathrm{d}}$ is the electron RT, computed as the inverse of the scattering rate, $\tau_{n\boldsymbol{k}}^{e-\mathrm{d}} = \left(\Gamma_{n\boldsymbol{k}}^{e-\mathrm{d}}\right)^{-1}$. The scattering rate $\Gamma_{n\boldsymbol{k}}^{e-\mathrm{d}}$ is given by

$$\Gamma_{n\boldsymbol{k}}^{e-\mathrm{d}} = \frac{1}{\mathcal{N}_{\boldsymbol{k}'}}\sum_{m\boldsymbol{k}'}W_{m\boldsymbol{k}',n\boldsymbol{k}}. \tag{5.6}$$

The scattering probability $W_{m\boldsymbol{k}',n\boldsymbol{k}}$ is

$$W_{m\boldsymbol{k}',n\boldsymbol{k}} = \frac{2\pi}{\hbar}n_{\mathrm{at}}C_{\mathrm{d}}\left|M_{mn}(\boldsymbol{k}',\boldsymbol{k})\right|^2\delta(\varepsilon_{m\boldsymbol{k}'} - \varepsilon_{n\boldsymbol{k}}). \tag{5.7}$$

Note that since $\tau_{n\boldsymbol{k}}$ is an electron quasiparticle lifetime, it can be written equivalently as the imaginary part of the e-d self-energy [17], $\left(\tau_{n\boldsymbol{k}}^{e-\mathrm{d}}\right)^{-1} = 2\mathrm{Im}\Sigma_{n\boldsymbol{k}}^{e-\mathrm{d}}/\hbar$.

In the RTA, we neglect the second term in Eq. (5.5) and obtain $\boldsymbol{F}_{n\boldsymbol{k}} = \tau_{n\boldsymbol{k}}\boldsymbol{v}_{n\boldsymbol{k}}$. In some cases, the second term in Eq. (5.5) cannot be neglected. In metals, a commonly used scheme to approximate this term is to add a factor of $(1 - \cos\theta_{\boldsymbol{k}',\boldsymbol{k}})$ to Eq. (5.7), where $\theta_{\boldsymbol{k}',\boldsymbol{k}}$ is the scattering angle between the $\boldsymbol{k}'$ and $\boldsymbol{k}$. The resulting so-called "transport relaxation time" is then used to compute the transport properties [17]. PERTURBO implements a more rigorous approach and directly solves Eq. (5.5) using an iterative method [18, 19], for which we rewrite Eq. (5.5) as

$$\boldsymbol{F}_{n\boldsymbol{k}}^{i+1} = \boldsymbol{F}_{n\boldsymbol{k}}^{0} + \frac{\tau_{n\boldsymbol{k}}^{e-\mathrm{d}}}{\mathcal{N}_{\boldsymbol{k}'}}\sum_{m\boldsymbol{k}'}\boldsymbol{F}_{m\boldsymbol{k}'}^{i}W_{m\boldsymbol{k}',n\boldsymbol{k}}. \tag{5.8}$$

In the iterative algorithm, we choose in the first step $\boldsymbol{F}_{n\boldsymbol{k}}^{0} = \tau_{n\boldsymbol{k}}^{e-\mathrm{d}}\mathbf{v}_{n\boldsymbol{k}}$, and then compute the following steps using Eq. (5.8) until the difference $\left|\boldsymbol{F}_{n\boldsymbol{k}}^{i+1} - \boldsymbol{F}_{n\boldsymbol{k}}^{i}\right|$ is within the convergence threshold.

Once $\boldsymbol{F}_{n\boldsymbol{k}}$ has been computed, either within the RTA or with the iterative solution of the BTE in Eq. (5.8), the conductivity tensor is obtained as

$$\sigma_{\alpha\beta} = \frac{1}{\Omega E_\beta} \cdot \frac{S}{\mathcal{N}_{\boldsymbol{k}}} \sum_{n\boldsymbol{k}} -e\mathbf{v}_{n\boldsymbol{k}}^{\alpha} \cdot f_{n\boldsymbol{k}}^{1} = \frac{e^2 S}{\mathcal{N}_{\boldsymbol{k}}\Omega} \sum_{n\boldsymbol{k}} \mathbf{v}_{n\boldsymbol{k}}^{\alpha} F_{n\boldsymbol{k}}^{\beta} \left(-\frac{\partial f_{n\boldsymbol{k}}^{0}}{\partial \varepsilon_{n\boldsymbol{k}}}\right), \tag{5.9}$$

where α and β are Cartesian directions, Ω is the volume of the unit cell, and S is the spin degeneracy. We also compute the carrier mobility tensor, $\mu_{\alpha\beta} = \sigma_{\alpha\beta}/(en_c)$, by dividing the conductivity tensor through the carrier concentration n_c.

In our implementation, we conveniently rewrite Eq. (5.9) as

$$\sigma_{\alpha\beta} = e^2 \int dE \left(-\partial f^0/\partial E\right) \Sigma_{\alpha\beta}(E), \tag{5.10}$$

where $\Sigma_{\alpha\beta}(E)$ is the transport distribution function (TDF) at energy E,

$$\Sigma_{\alpha\beta}(E) = \frac{S}{\mathcal{N}_{\boldsymbol{k}}\Omega} \sum_{n\boldsymbol{k}} \mathbf{v}_{n\boldsymbol{k}}^{\alpha} \boldsymbol{F}_{n\boldsymbol{k}}^{\beta} \delta\left(E - \varepsilon_{n\boldsymbol{k}}\right), \tag{5.11}$$

which is computed in PERTURBO using the tetrahedron integration method [20]. The integrand in Eq. (5.10) can be used to characterize the contributions to transport as a function of electron energy [8, 9].

The e-d relaxation times computed in the previous chapters are based on the RTA rather than the iterative approach. Using the interpolation methods for neutral and charged defects developed in this work, one can compute the defect-limited mobility or the mobility in the presence of both e-ph and e-d interactions. Both the RTA and iterative BTE can be implemented; the latter is expected to more accurately take into account backscattering than the approximate approach of using the cosine of the scattering angle given in Eq. (4.5). The current PERTURBO has a similar workflow for e-ph interactions to compute the e-ph relaxation times $\tau_{n\boldsymbol{k}}^{e-\mathrm{ph}}$ and phonon-limited carrier mobility $\mu^{e-\mathrm{ph}}$ [16]. With the machinery built in this work, we can compute the e-d relaxation times $\tau_{n\boldsymbol{k}}^{e-\mathrm{d}}$ and defect-limited carrier mobility $\mu^{e-\mathrm{d}}$. To compute charge transport that includes both e-ph and e-d interactions, we can use the Matthiessen's rule for the relaxation times $\tau_{n\boldsymbol{k}}$:

$$\frac{1}{\tau_{n\boldsymbol{k}}} = \frac{1}{\tau_{n\boldsymbol{k}}^{e-\mathrm{ph}}} + \frac{1}{\tau_{n\boldsymbol{k}}^{e-\mathrm{d}}} \tag{5.12}$$

and the Mattiessen's rule for the carrier mobility μ:

$$\frac{1}{\mu} = \frac{1}{\mu^{e-\mathrm{ph}}} + \frac{1}{\mu^{e-\mathrm{d}}}. \tag{5.13}$$

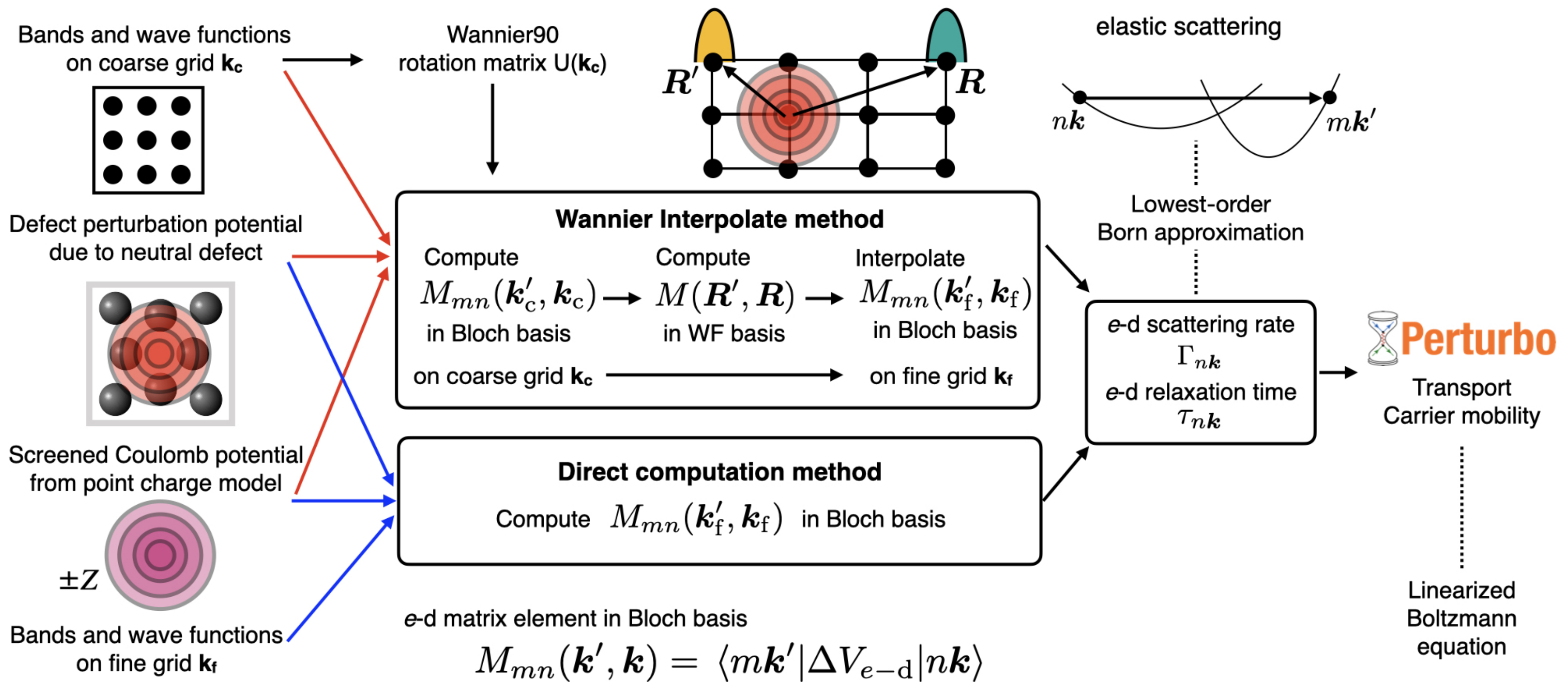

Figure 5.1: Workflow for electron-defect calculations in PERTURBO.

5.3 Software to Compute Electron-Defect Matrix Elements for Neutral and Charged Defects

Figure 5.1 shows the workflow of *ab initio* e-d calculations. The main inputs to compute the e-d matrix elements, RTs, and carrier mobility are the wave functions and band structure and the defect perturbation potential due to neutral or charged defects.

Our routines include two approaches to compute e-d matrix elements: the direction computation and the interpolation method. The input wave functions and band structure depend on which method is selected. For the direct computation method, a list of $\boldsymbol{k}$-points needed in the transport calculations is specified by the user. For a given BZ grid and energy window of relevance for the transport calculation (usually, within 100 meV of the band edges in semiconductors), the required initial-state $\boldsymbol{k}$-points are those with Bloch states whose energies are within the window; the final-state $\boldsymbol{k}$-points are determined by the energy broadening for the Gaussian function approximating the Dirac delta function for energy conservation in our implementation. For improved efficiency, we use the Wannier-interpolated band structure to select the relevant $\boldsymbol{k}$-points, and then run non-self-consistent DFT calculations at the selected $\boldsymbol{k}$-points to obtain the wave functions. In contrast, for the interpolation method, the wave functions and band structure are computed on a coarse uniform BZ $\boldsymbol{k}_{\text{c}}$-grid and used to construct Wannier functions and obtain the associated unitary matrices. The wave functions on the fine BZ $\boldsymbol{k}_{\text{c}}$-grid points are not needed in the interpolation approach, which affords significant saving of computational resources.

In our current e-d calculations, the defect perturbation potential for neutral defects is obtained as the difference between the KS Hamiltonian of a pristine supercell and that of a supercell containing one defect. For charged defects, for now we examined a Yukawa potential as we verified that the short-range part gives a much smaller contribution than the long-range, which is governed by electrostatic effects. To better capture the atomic relaxation around a charged defect, one can add the defect perturbation potential of the neutral defect to the charged defect potential, perform DFT calculations for a charged defect instead of using the Yukawa potential, or use the potential generated from a narrow point charge distribution screened by a material-specific dielectric function computed from first principles. These approaches will be examined in future works.

Once the wave functions, band structure, and defect perturbation potential have been

computed and read as input, the e-d matrix elements are calculated using either the direct computation or the interpolation method. Compared to the direct computation method, which is straightforward but computationally costly, the interpolation method involves the Fourier-Wannier transformation. The e-d matrix elements in the Bloch states are first computed on the coarse BZ $\boldsymbol{k}_c$-grid and then transformed from their Bloch form, $M(\boldsymbol{k}', \boldsymbol{k})$, to the Wannier function basis, $M(\boldsymbol{R}', \boldsymbol{R})$, where the band index is omitted for simplicity. Using the Wannier basis, the matrix elements can be interpolated for any momentum pair on a fine BZ $\boldsymbol{k}_f$-grid. Note that for charged defects, one has to remove the long-range part of the e-d matrix elements on the coarse BZ $\boldsymbol{k}_c$-grid before the transformation and add it back subsequently in momentum space. The details of how to compute e-d matrix elements can be found in Chapter 2. The methods of how to interpolate the e-d matrix elements using Wannier functions for neutral defects and charged defects are described in Chapters 3 and 4, respectively.

The resulting e-d matrix elements are used to compute the e-d RTs due to elastic scattering within the lowest-Born approximation in Eq. (5.6). With the RTs in hand, we can further compute the defect-limited carrier mobility using the RTA or the iterative BTE approach in Eq. (5.5).

5.4 Parallelization of Electron-Defect Scattering Calculations

Parallel computing is needed to carry out our e-d interaction calculations within a few hours instead of months or years. Parallel computing executes the program instructions in parallel on multiple computer CPUs, rather than serially executing the program instructions one after another. However, writing codes using parallel programming is not a simple task because it typically involves entirely redesigning the code and data structures of an existing serial code. A parallel program code tells each computer processor which segments of the code it should execute simultaneously with the other processors working on different segments. This demanding task rests on the skills of code developers, who need to choose which parallel programming languages and which code and data structures to use. In our codes for *ab initio* e-d interactions, we combine two parallel programming paradigms and libraries, MPI and OpenMP, which are complementary to each other. MPI is a widely used distributed memory scheme to parallelize the workload by distributing it among different compute nodes, while OpenMP is a shared memory approach to leverage on-node parallel operations.

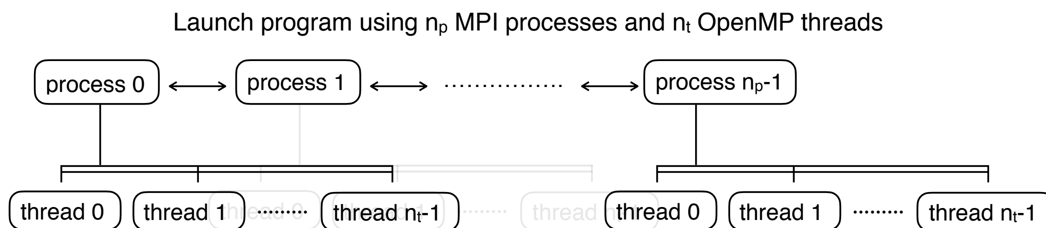

Figure 5.2: The simple concept of hybrid parallel programming.

MPI stands for Message Passing Interface on distributed memory systems. When a program is launched in parallel, a copy of the same executable program runs on each MPI process. Each MPI process has its own memory, which cannot be accessed by other MPI processes; therefore, any data to be shared among MPI processes must be explicitly specified and transmitted from one process to another. The communication among MPI processes happens in a defined group called a communicator within which each MPI process has a unique rank. Each MPI process can belong to various communicators, but gets assigned a unique rank in each communicator. As a result, a code with complicated several-tier parallelization may need a wrapper to the MPI libraries, as is done in the QE code [5, 6], whose MPI wrapper was used in our routines. OpenMP is an application program interface that uses shared-memory multithreading, using two or more CPUs within a single compute node. Each process, which has its own memory space, has at least one thread of execution (usually called a master thread). The master thread can fork a specified number of slave threads, dividing a task among them and executing the task in parallel. All the threads created by a process share the same address space of that process, although each thread can also have its own local variables.

Instead of relying heavily on one parallel programming paradigm, we use a hybrid parallel programming technique combining MPI and OpenMP in our codes for *e*-d calculations. Such a hybrid MPI plus OpenMP approach is widely regarded as a highly desirable new standard for efficient scientific computing. It allows one to launch a few MPI processes that have their own copy of the data with memory size of order 1-10 GB, and uses several OpenMP threads per MPI process to perform each computational task. This approach can save memory by not replicating data to too many MPI processes while also taking advantage of the shared memory of the OpenMP threads.

Our hybrid programming approach for *e*-d calculations is akin to Figure 5.2. When

we launch the executable program, the first parallelization layer is MPI and the second is OpenMP. Their combination allows us to run the program using less memory while achieving higher speed. Because the data in each MPI process is not shared with the other processes, message-passing subroutines are needed to send, receive, broadcast, and collect data among MPI processes. When designing codes with MPI parallelization, the developer has to specify when and how to share data among processes, which complicates the coding process. To simplify the task, QE developers have designed several subroutines and modules using the interface block structure in Fortran. The MPI subroutines in QE need to be used with care as they include several parallelization layers. Up to version 6.5 of QE, there are several MPI parallelization levels, and each level is assigned to a communicator. For example, one level involves all the MPI processes (`comm_world`), another one includes only the MPI processes associated with a pool of $\boldsymbol{k}$-points in the BZ grid (e.g., `intra_pool_comm` and `inter_pool_comm`), and so on. These details are documented on the QE website [21]. When using the QE subroutines directly, as we do in parts of our e-d codes, it is important to make sure that the parallelization implemented in the code is compatible with the one used in QE. For example, we ran into issues when loading the wave functions and related data from the QE output directory `prefix.save`, and overcoming these technical problems required time-consuming debugging.

In the e-d codes, we only use simple OpenMP functionalities to improve efficiency, such as parallelizing for loops in Fortran, for which we place a simple statement `!$omp parallel do` to launch OpenMP threads and perform parallel calculations.

In the following, we discuss our implementation of MPI and OpenMP parallelization in the e-d codes using the neutral defect case as an example. The main tasks involved in the e-d calculations are: 1) Reading the wave functions and related physical quantities from the QE directory `prefix.save`; 2) Obtaining the Fourier coefficients of the defect perturbation potential; 3) Computing the local and nonlocal parts of the e-d matrix elements; 4) Performing the Fourier-Wannier transformation of the e-d matrix elements; 5) Computing the e-d RTs using the interpolated e-d matrix elements.

In the first step, we launch the e-d codes to read the wave functions from the QE output directory `prefix.save` using MPI processes with the default values for all QE parallelization levels. All the MPI processes work on the plane-wave parallelization level of the QE subroutines. Consider the wave function for a Bloch

```
allocate(wfc_uc(ucdata%npol*ucdata%npwx,ucdata%nbnd,ucdata%nkstot)); wfc_uc = 0.0_DP

dirname = trim(prefix)//postfix

if (wfc_is_collected) then
   do ik = 1, nks
      ik_g = global_kpoint_index(nkstot, ik)
      call read_collected_wfc(dirname,ik,evc)
      do ig = 1, ngk(ik_g)
         igst = 0
         do imp = 1, mpime
            igst = igst+ngk_each(imp,ik)
         end do
         ig_g = igst + ig
         do ipol = 1, npol
            wfc_uc(ig_g+(ipol-1)*ucdata%npwx,1:ucdata%nbnd,ik_g) = evc(ig+(ipol-1)*npwx,1:nbnd)
         end do
      end do
   end do
end if

deallocate(ngk_each)
call mp_sum(wfc_uc,world_comm)
```

Figure 5.3: A portion of the electron-defect routine to read wave functions.

state $|n\boldsymbol{k}\rangle$:

$$|n\boldsymbol{k}\rangle = \sum_{\boldsymbol{G}} C_{n\boldsymbol{k}}(\boldsymbol{G}) e^{i\boldsymbol{G}\cdot\boldsymbol{r}}, \tag{5.14}$$

where $C_{n\boldsymbol{k}}(\boldsymbol{G})$ are the Fourier coefficients at the reciprocal lattice vector $\boldsymbol{G}$. In the MPI routines, the $\boldsymbol{G}$ vectors and the corresponding coefficients are divided and distributed among the MPI processes. Figure 5.3 shows a portion of the e-d routine used to collect all the coefficients from the MPI processes. First, we create a new variable called `wfc_uc` to store the Fourier coefficients of the wave functions for all Bloch states. The variable `ucdata` is a derived data type we define to store all the quantities related to the primitive cell: for example, `ucdata%npol` is the number of spin components of the wave functions (npol =1 for scalars, npol=2 for spinors), `ucdata%nbnd` is the number of bands, and `ucdata%nkstot` is the total number of $\boldsymbol{k}$-points.

In the first do-loop (starting with `ik`), for each $\boldsymbol{k}$-point we use a QE subroutine called `read_collected_wfc` to read the corresponding wave functions from the QE output directory into the variable `evc`. Because the $\boldsymbol{G}$ vectors are distributed over the MPI processes, each $\boldsymbol{k}$-point has only a portion of `ngk(ik_g)` $\boldsymbol{G}$ vectors (where `ngk` is the number of plane waves and `ik_g` is the global index of the $\boldsymbol{k}$-

point), and `igst` is a variable to specify where the data should be put inside the `wfc_uc`. After the do-loop, we use the MPI subroutine from QE, `mp_sum`, to sum the wavefunctions from all processes and pass the result to the master process, which then broadcasts all the wave functions.

In the second step, we read the local potential from the QE output directory and use QE subroutines with several MPI parallelization layers. The defect potential is stored in a QE defined array `vrs`, which has a number equal to `dfftp%nnr`, where `dfftp` is a derived data type defined in QE to handle the Fast Fourier Transform (FFT) grid. The elements of the FFT grid (`dfftp%nnr`) are divided and distributed among the MPI processes. The FFT grid in a serial code has a `nr1*nr2*nr3` elements, where `nr1`, `nr2`, and `nr3` are the number of points along the three crystallographic directions. However, when we run the calculations in parallel, the `nr3` points are divided and distributed among the MPI processes, and thus the variable `dfftp%nnr` is no longer equal to `nr1*nr2*nr3`, and the size of `vrs` is not the same as the one used in the serial calculations. To obtain the entire local potential, we need to define a real variable with a size of `nr1*nr2*nr3` and use the QE subroutine `gather_grid` to collect all the data from the other MPI processes. Once the master process obtains all the elements for the FFT grid, it broadcasts the whole local defect perturbation potential to all MPI processes.

We use a hybrid MPI plus OpenMP scheme also for step 2 to obtain the Fourier coefficients of the local defect perturbation potential. Suppose we want to obtain the Fourier coefficients of the perturbation potential for the `nqc` wavevectors. We first use our own subroutine `mp_split_processes`, which splits the `nqc` points into the MPI processes. We then apply OpenMP threads to each portion of the `nqc` points assigned to each MPI process. Each OpenMP thread runs the Fourier transformation for each point for a given MPI process, and the Fourier coefficients are then stored in the variable `vqc`. After the calculation, we use the QE parallel subroutine `mp_sum` to collect and broadcast the Fourier coefficients among the MPI processes.

In step 3, we compute the local and nonlocal e-d matrix elements using the property of the complex conjugate of the matrix elements,

$$M_{nm}(\boldsymbol{k},\boldsymbol{k}') = M^{*}_{mn}(\boldsymbol{k}',\boldsymbol{k}), \tag{5.15}$$

so we need to compute only the matrix elements for each pair of Bloch states, which corresponds to a scattering channel. Before the matrix element calculations, the

codes determine the number of scattering channels and distribute them among the MPI processes, using OpenMP threads to parallelize the calculations for the matrix element for each scattering channel.

In step 4, we use hybrid parallel programming to perform a double Fourier transformation of the e-d matrix elements in the Bloch states to the Wannier basis. For each pair of Bravais lattice points, we use the subroutine `mp_split_processes` to split the number of pairs among MPI processes, and use OpenMP threads to perform the double Fourier transformation.

Finally, in step 5, we use hybrid parallel programming to calculate the e-d scattering rates using the interpolated e-d matrix elements. Given a list of $\boldsymbol{k}$-points, we split them among the MPI processes. Each MPI process launches a set of OpenMP threads to distribute the load associated with the $\boldsymbol{k}$-points assigned to the MPI process. Each OpenMP thread interpolates the e-d matrix elements, from which the scattering rates are computed. Once all the scattering rates are computed, we use the QE subroutine `mp_sum` to collect them from all the MPI processes.

More details about the e-d codes will be made available when the routines are released to the public. The e-d codes will be released in the future version of PERTURBO.

5.5 Summary

We reviewed calculations of the defect-limited carrier mobility using two approaches, the RTA and the iterative approach of the BTE. We then discussed calculations of the associated e-d matrix elements for neutral and charged defects using the direct computation and the interpolation method, and combining the e-d routines with PERTURBO. Lastly, we discussed hybrid MPI+OpenMP parallelization, which is indispensable in the e-d codes to speed up the calculations from a technical viewpoint.

References

[1] F. Rossi and T. Kuhn, Rev. Mod. Phys. **74**, 895 (2002).

[2] R. Ulbricht, E. Hendry, J. Shan, T. F. Heinz, and M. Bonn, Rev. Mod. Phys. **83**, 543 (2011).

[3] D. A. Egger, A. Bera, D. Cahen, G. Hodes, T. Kirchartz, L. Kronik, R. Lovrincic, A. M. Rappe, D. R. Reichman, and O. Yaffe, Adv. Mater. **30**, 1800691 (2018).

[4] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi, Rev. Mod. Phys. **73**, 515 (2001).

[5] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, J. Phys.: Condens. Matter **21**, 395502 (2009).

[6] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. Buongiorno Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. Dal Corso, S. De Gironcoli, P. Delugas, R. A. Distasio, A. Ferretti, A. Floris, G. Fratesi, G. Fugallo, R. Gebauer, U. Gerstmann, F. Giustino, T. Gorni, J. Jia, M. Kawamura, H. Y. Ko, A. Kokalj, E. Kücükbenli, M. Lazzeri, M. Marsili, N. Marzari, F. Mauri, N. L. Nguyen, H. V. Nguyen, A. Otero-De-La-Roza, L. Paulatto, S. Poncé, D. Rocca, R. Sabatini, B. Santra, M. Schlipf, A. P. Seitsonen, A. Smogunov, I. Timrov, T. Thonhauser, P. Umari, N. Vast, X. Wu, and S. Baroni, J. Phys.: Condens. Matter **29**, 465901 (2017).

[7] J. M. Ziman, *Electrons and phonons: The theory of transport phenomena in solids* (Oxford University Press, 2001).

[8] J. J. Zhou and M. Bernardi, Phys. Rev. B **94**, 201201(R) (2016).

[9] J.-J. Zhou, O. Hellman, and M. Bernardi, Phys. Rev. Lett. **121**, 226603 (2018).

[10] V. A. Jhalani, J.-J. Zhou, and M. Bernardi, Nano Lett. **17**, 5012 (2017).

[11] J. Park, J.-J. Zhou, and M. Bernardi, Phys. Rev. B **101**, 045202 (2020).

[12] N.-E. Lee, J.-J. Zhou, L. A. Agapito, and M. Bernardi, Phys. Rev. B **97**, 115203 (2018).

[13] M. Bernardi, Eur. Phys. J. B **89**, 239 (2016).

[14] G. K. Madsen and D. J. Singh, Comput. Phys. Commun. **175**, 67 (2006).

[15] G. Pizzi, D. Volja, B. Kozinsky, M. Fornari, and N. Marzari, Comput. Phys. Commun. **185**, 422 (2014).

[16] J.-J. Zhou, J. Park, I.-T. Lu, I. Maliyov, X. Tong, and M. Bernardi, arXiv:2002.02045 (2020).

[17] G. D. Mahan, *Many-particle physics* (Springer US, 2000).

[18] W. Li, Phys. Rev. B **92**, 75405 (2015).

[19] M. Fiorentini and N. Bonini, Phys. Rev. B **94**, 085204 (2016).

[20] P. E. Blöchl, O. Jepsen, and O. K. Andersen, Phys. Rev. B **49**, 16223 (1994).

[21] QUANTUM ESPRESSO official website, https://www.quantum-espresso.org/.

Chapter 6

FUTURE EXTENSION OF *AB INITIO* ELECTRON-DEFECT INTERACTIONS

This thesis developed three powerful tools for investigating the electron-defect (e-d) interactions, and extended the current tool box of first-principles methods.

The first approach developed in Chapter 2 enables the new efficient calculations of e-d matrix elements. Our formalism uses only the wave functions of the primitive cell and can efficiently compute matrix elements. The approach is applied to neutral defects, in particular vacancy and interstitial defects in silicon, for which we demonstrate how to converge the e-d relaxation times using over 1 million points in the Brillouin zone (BZ), a daunting task that was prohibitively expensive using pre-existing methods that compute the e-d matrix elements using supercell wave functions. This first tool is a stepping stone of further development for *ab initio* e-d interactions in the plane-wave pseudopotential DFT formalism. Though we study atomic defects, our formalism is general and can be applied to extended defects such as dislocations and grain boundaries, with important implications for materials science and novel electronic materials and devices [1–4].

The second tool developed in Chapter 3 is an interpolation method for e-d matrix elements using Wannier functions to further speed up e-d calculations. The interpolation method takes advantage of the localized nature of Wannier functions and the short-range nature of the perturbation potential for neutral (non-charged) defects. As a result, the e-d matrix elements in the Wannier function basis can be represented using only a small set of Wannier functions, from which any e-d matrix elements in the Bloch states can be obtained by interpolation through a double generalized Fourier transform. Using this newly developed interpolation method, we show how to converge the defect-limited carrier mobility as a function of BZ grid size for a neutral vacancy in silicon, a demanding computational task requiring BZ grids of size $200 \times 200 \times 200$ or greater. The interpolation method is faster by two to four orders of magnitude than the direct computation with our already efficient approach. Therefore, the interpolation tool is a milestone in *ab initio* e-d calculations and is particularly suitable for short-range defects such as neutral defects or defects in metals.

The third method developed in Chapter 4 is a Wannier interpolation method for e-d matrix elements for charged defects that correctly takes into account the long-range nature of the perturbation potential. We apply the interpolation method to a phosphorus ionized impurity in silicon, modeling it as a source of Yukawa potential. We show that our approach can reproduce the directly computed e-d matrix elements. To this end, the long-range part of the potential is first subtracted on the coarse grid, and then added back in momentum space on the fine BZ grid after interpolation. We compute and converge the e-d relaxation times using the interpolated e-d matrix elements together with random BZ grids distributed according to a Cauchy distribution, which effectively samples scattering events with small momentum transfer $q = |\boldsymbol{k}' - \boldsymbol{k}|$. Future work will focus on including electronic screening in the random-phase approximation and free carrier screening on the same footing. This third tool and the related techniques developed in Chapter 4 are the first step toward the development of *ab initio* methods to treat e-d scattering by the charged defects in semiconductors.

Despite the great progress we have made so far, *ab initio* e-d interaction in the plane-wave DFT framework with the pseudopotential method is still in its infancy; there are many research directions and corresponding numerical methods yet to be explored. In the following, we outline future extensions of our work on *ab initio* e-d interactions.

In this thesis, we studied electronic states without including their spin explicitly, which is adequate for non-spin polarized materials in the absence of spin-orbit coupling. Our tools can be generalized to treat electronic states as spinors and compute the corresponding e-d perturbation potential, which becomes a 2×2 matrix in spin space in the presence of spin-orbit coupling or noncollinear spin. To include the spin degree of freedom, we obtain the spin up and down components of the electronic wave functions using DFT calculations either in the local spin density approximation for collinear spin states or using noncollinear spin calculations. In the former case, the Kohn-Sham (KS) Hamiltonians for spin up and down are solved separately to obtain a separate KS e-d perturbation potential for each spin up and spin down state. The methods we have developed here can be extended straightforwardly to compute scattering rates and relaxation times for each spin direction. The noncollinear spin calculations pose additional challenges and require fully relativistic pseudopotentials to treat spin-orbit coupling, together with noncollinear magnetic states when necessary. With the machinery for spinor-defect interaction, one can explore e-

d interactions and spin-flip events in materials with strong spin-orbit coupling or magnetic impurities, and investigate e-d processes in quantum materials such as topological insulators or Weyl semimetals [5–9]. An interpolation method for spin-defect matrix elements can be derived following a recently developed approach for spinor-phonon interaction [10]. Preliminary work on this topic has been carried out by an undergraduate intern, Shreshth Malik, under my supervision and has led to the development of routines to treat spin-orbit coupling in e-d interactions that will be further tested and refined in future work.

For ionized impurity scattering, we have focused on a simple point charge modelled as a Yukawa potential in 3D materials. The charged defect potential can also be computed using DFT, but one needs to be cautious about the finite-size effects and interactions from the image simulation cells. Due to the long-range Coulomb potential, the defect perturbation potential converges slowly with system size and so do the associated e-d matrix elements. In addition, plane-wave DFT makes the simulation cell charge neutral by adding a background charge (technically, by removing the $\boldsymbol{G} = \mathbf{0}$ component of the Hartree potential), which creates noteworthy technical challenges. One approach to overcome these issues is to separate the long-range Coulomb potential from the DFT-computed defect perturbation potential using a simple model such as a Yukawa potential. The associated e-d matrix elements for the remaining short-range perturbation potential are computed using either the direct computation method or the Wannier interpolation method as described in Chapters 2 and 3, respectively. For the long-range Coulomb potential modelled by the simple model, the short-range and the long-range part of the e-d matrix elements are treated using the computation methods developed in Chapter 4. The dielectric screening is crucial to modeling the screened Coulomb potential; we plan to compute it from first-principles in future work using the random-phase approximation already widely employed in the GW method and available, for example, in the Yambo [11, 12] and BerkeleyGW code [13]. In addition to charged defects in 3D materials, obtaining correct charged defect potential in 2D materials is also another challenge [14]. However, our methods can directly be generalized to charged defects in 2D and 3D.

Finally, the e-d interaction employed in this work is based on the lowest-order Born approximation. For stronger defect perturbation potentials, higher-order term corrections are important and need to be included. My work has preliminarily explored the T-matrix formalism to include higher-order e-d interactions. The T-matrix approximation can apply to defects of arbitrary strength, while the Born

approximation is only valid for weak defects [15]. As the higher-order corrections require a large number of e-d matrix elements, the methods developed in this thesis form the basis for future efficient *ab initio* calculations of higher-order e-d interactions. In turn, the higher order corrections will enable the study of a much broader range of defect-related physical processes, including resonant scattering, bound states, electron localization and antilocalization, and quantum corrections to transport, among others. Recently, the T-matrix approach using atomic orbitals has been developed by another group to compute carrier mobility in 2D materials and to study valleytronics in 2D materials [15, 16]. However, plane-wave e-d calculations, especially in bulk systems where a 3D BZ grid is needed, are still out of reach for first-principles methods. The results shown in this thesis will make them possible in the near future.

In conclusion, we believe the methods developed in this work form a solid foundation for first-principles calculations of e-d interactions in bulk and nanostructured materials. They can be applied broadly to address materials design challenges in electronics, energy, and quantum technologies.

References

[1] A. Rycerz, J. Tworzydło, and C. W. Beenakker, Nat. Phys. **3**, 172 (2007).

[2] D. Gunlycke and C. T. White, Phys. Rev. Lett. **106**, 136806 (2011).

[3] S. I. Kim, K. H. Lee, H. A. Mun, H. S. Kim, S. W. Hwang, J. W. Roh, D. J. Yang, W. H. Shin, X. S. Li, Y. H. Lee, G. J. Snyder, and S. W. Kim, Science **348**, 109 (2015).

[4] M. A. U. Absor, I. Santoso, N. Yamaguchi, and F. Ishii, Phys. Rev. B **101**, 155410 (2020).

[5] P. Hosur, S. A. Parameswaran, and A. Vishwanath, Phys. Rev. Lett. **108**, 046602 (2012).

[6] Z. Huang, D. P. Arovas, and A. V. Balatsky, New J. Phys. **15**, 123019 (2013).

[7] G. Yin, D. Wickramaratne, Y. Zhao, and R. K. Lake, Appl. Phys. Lett. **105**, 33118 (2014).

[8] P. E. C. Ashby and J. P. Carbotte, Phys. Rev. B **89**, 245121 (2014).

[9] M. Shiranzaei, F. Parhizgar, J. Fransson, and H. Cheraghchi, Phys. Rev. B **95**, 235429 (2017).

[10] J. Park, J.-J. Zhou, and M. Bernardi, Phys. Rev. B **101**, 045202 (2020).

[11] A. Marini, C. Hogan, M. Grüning, and D. Varsano, Comput. Phys. Commun. **180**, 1392 (2009).

[12] D. Sangalli, A. Ferretti, H. Miranda, C. Attaccalite, I. Marri, E. Cannuccia, P. Melo, M. Marsili, F. Paleari, A. Marrazzo, et al., J. Phys.: Condens. Matter **31**, 325902 (2019).

[13] J. Deslippe, G. Samsonidze, D. A. Strubbe, M. Jain, M. L. Cohen, and S. G. Louie, Comput. Phys. Commun. **183**, 1269 (2012).

[14] J. Xiao, K. Yang, D. Guo, T. Shen, H.-X. Deng, S.-S. Li, J.-W. Luo, and S.-H. Wei, Phys. Rev. B **101**, 165306 (2020).

[15] K. Kaasbjerg, Phys. Rev. B **101**, 045433 (2020).

[16] K. Kaasbjerg, J. H. J. Martiny, T. Low, and A.-P. Jauho, Phys. Rev. B **96**, 241411(R) (2017).

Appendix A

ELASTIC AND INCOHERENT ELECTRON-DEFECT SCATTERING RATE

The scattering rate $\Gamma_{n\boldsymbol{k}}$ can be written using Fermi's golden rule within the lowest-order perturbation theory:

$$\Gamma_{n\boldsymbol{k}} = \frac{2\pi}{\hbar} \sum_{n'\boldsymbol{k}'} \left| \langle \psi_{n'\boldsymbol{k}'} | \Delta\hat{H} | \psi_{n\boldsymbol{k}} \rangle \right|^2 \delta\left(\varepsilon_{n'\boldsymbol{k}'} - \varepsilon_{n\boldsymbol{k}}\right), \tag{A.1}$$

where the perturbation $\Delta\hat{H}$ is the difference between the Kohn-Sham Hamiltonian $\hat{H}^{(\mathrm{d})}$ of a crystal containing N_{d} identical defects and the Hamiltonian $\hat{H}^{(\mathrm{p})}$ of the same crystal with no defects, namely, $\Delta\hat{H} = \hat{H}^{(\mathrm{d})} - \hat{H}^{(\mathrm{p})}$. The crystal is made up by $N_{\boldsymbol{k}}$ primitive cells, and we apply Born-von Karman (BvK) periodic boundary conditions; the crystal volume is $\Omega_{\mathrm{B}vK} = N_{\boldsymbol{k}}\Omega_{\mathrm{u}c}$, where $\Omega_{\mathrm{u}c}$ is the volume of the primitive cell. Above, $|\psi_{n\boldsymbol{k}}\rangle$ are unperturbed Bloch wave functions with energy $\varepsilon_{n\boldsymbol{k}}$, which in coordinate space read:

$$\langle \boldsymbol{r} | \psi_{n\boldsymbol{k}} \rangle = \frac{1}{\sqrt{N_{\boldsymbol{k}}}} \langle \boldsymbol{r} | n\boldsymbol{k} \rangle = \frac{1}{\sqrt{N_{\boldsymbol{k}}}} u_{n\boldsymbol{k}}(\boldsymbol{r})\, e^{i\boldsymbol{k}\cdot\boldsymbol{r}}, \tag{A.2}$$

where $|n\boldsymbol{k}\rangle$ is the Bloch wavefunction without the prefactor, and $u_{n\boldsymbol{k}}(\boldsymbol{r})$ is the periodic part of the Bloch wavefunction, normalized in the primitive cell as

$$\int_{\Omega_{\mathrm{u}c}} d^3r\, u^*_{n\boldsymbol{k}}(\boldsymbol{r})\, u_{n\boldsymbol{k}}(\boldsymbol{r}) = 1. \tag{A.3}$$

Since the kinetic energy is the same in the pristine and defect-containing systems, the difference of their Hamiltonians equals the sum of the perturbations due to all defects:

$$\Delta\hat{H} = \hat{H}^{(\mathrm{d})} - \hat{H}^{(\mathrm{p})} = \sum_{i=1}^{N_{\mathrm{d}}} \Delta V_{\mathrm{e}-d}(\boldsymbol{r} - \boldsymbol{r}_i), \tag{A.4}$$

where $\Delta V_{\mathrm{e}-d}(\boldsymbol{r} - \boldsymbol{r}_i)$ denotes the perturbation potential due to a defect located at $\boldsymbol{r}_i$, and we consider non-interacting defects of the same kind. Assuming that the scattering events are independent, we can write the scattering rate for elastic and incoherent scattering processes due to all defects as

$$\Gamma_{n\boldsymbol{k}} = \frac{2\pi}{\hbar} \frac{n_{\mathrm{at}} C_{\mathrm{d}}}{N_{\boldsymbol{k}'}} \sum_{n'\boldsymbol{k}'} \left| M_{n'\boldsymbol{k}',n\boldsymbol{k}} \right|^2 \delta\left(\varepsilon_{n'\boldsymbol{k}'} - \varepsilon_{n\boldsymbol{k}}\right), \tag{A.5}$$

where n_{at} is the number of atoms in the primitive cell, $C_{\rm d}$, which is formally equal to $N_{\rm d}/(N_{\boldsymbol{k}} \times n_{at})$, and is in practice an assumed value of the defect concentration, and $M_{n'\boldsymbol{k}',n\boldsymbol{k}}$ is defined as the e-d matrix element for the perturbation due to a single defect:

$$M_{n'\boldsymbol{k}',n\boldsymbol{k}} = \langle n'\boldsymbol{k}'|\Delta V_{\rm e-\mathit{d}}|n\boldsymbol{k}\rangle \, . \tag{A.6}$$

Within our approximations, the scattering rate is proportional to the defect concentration, and can be computed at any desired value of the defect concentration, provided that the scattering events remain uncorrelated and the defects non-interacting throughout the concentration range of interest.

Appendix B

DERIVATION OF LOCAL PART OF ELECTRON-DEFECT MATRIX ELEMENTS

The local matrix elements can be written as:

$$\begin{aligned} M^{\mathrm{L}}_{n'\boldsymbol{k}',n\boldsymbol{k}} &= \langle n'\boldsymbol{k}'|\Delta V_{\mathrm{L}}(\boldsymbol{r})|n\boldsymbol{k}\rangle \\ &= \int_{\Omega_{\mathrm{BvK}}} d^3r\, u^*_{n'\boldsymbol{k}'}(\boldsymbol{r})\, e^{-i\boldsymbol{k}'\cdot\boldsymbol{r}}\Delta V_{\mathrm{L}}(\boldsymbol{r})\, u_{n\boldsymbol{k}}(\boldsymbol{r})\, e^{i\boldsymbol{k}\cdot\boldsymbol{r}}. \end{aligned} \tag{B.1}$$

We define the forward and inverse Fourier transforms of the local perturbation potential, respectively, as

$$\Delta V_{\mathrm{L}}(\boldsymbol{r}) = \frac{1}{N_{\boldsymbol{k}}}\sum_{\boldsymbol{q}} \Delta\widetilde{V}_{\mathrm{L}}(\boldsymbol{q})\, e^{i\boldsymbol{q}\cdot\boldsymbol{r}} \tag{B.2}$$

and

$$\begin{aligned} \Delta\widetilde{V}_{\mathrm{L}}(\boldsymbol{q}) &= \frac{1}{\Omega_{\mathrm{uc}}}\int_{\Omega_{\mathrm{B}vK}} d^3r\, \Delta V_{\mathrm{L}}(\boldsymbol{r})\, e^{-i\boldsymbol{q}\cdot\boldsymbol{r}} \\ &\approx \frac{1}{\Omega_{\mathrm{uc}}}\int_{\Omega_{sup}} d^3r\, \Delta V_{\mathrm{L}}(\boldsymbol{r})\, e^{-i\boldsymbol{q}\cdot\boldsymbol{r}}, \end{aligned} \tag{B.3}$$

where in the last line we replace the crystal volume $\Omega_{\mathrm{B}vK}$ with the supercell volume Ω_{sup}, using the fact that the local perturbation potential vanishes at the supercell boundary, which is typically the case for supercells larger than a few primitive cells due to the localized nature of the perturbation potential. Inserting Eq. (B.2) into Eq. (B.1) and using the translational invariance of $u_{n\boldsymbol{k}}(\boldsymbol{r})$, we have

$$M^{\mathrm{L}}_{n'\boldsymbol{k}',n\boldsymbol{k}} = \sum_{\boldsymbol{G}} \Delta\widetilde{V}_{\mathrm{L}}(\boldsymbol{k}'-\boldsymbol{k}+\boldsymbol{G})\, \langle u_{n'\boldsymbol{k}'}|e^{i\boldsymbol{G}\cdot\boldsymbol{r}}|u_{n\boldsymbol{k}}\rangle_{\mathrm{uc}}, \tag{B.4}$$

with the plane-wave matrix elements defined as

$$\langle u_{n'\boldsymbol{k}'}|e^{i\boldsymbol{G}\cdot\boldsymbol{r}}|u_{n\boldsymbol{k}}\rangle_{\mathrm{uc}} = \int_{\Omega_{\mathrm{uc}}} d^3r\, u^*_{n'\boldsymbol{k}'}(\boldsymbol{r})\, e^{i\boldsymbol{G}\cdot\boldsymbol{r}} u_{n\boldsymbol{k}}(\boldsymbol{r}), \tag{B.5}$$

where $\boldsymbol{G}$ are reciprocal lattice vectors of the primitive cell. This formula is valid for any basis set. Here we use a plane wave basis for $u_{n\boldsymbol{k}}(\boldsymbol{r})$, and write

$$u_{n\boldsymbol{k}}(\boldsymbol{r}) = \frac{1}{\sqrt{\Omega_{\mathrm{uc}}}}\sum_{\boldsymbol{G}} C_{n\boldsymbol{k}}(\boldsymbol{G})\, e^{i\boldsymbol{G}\cdot\boldsymbol{r}}, \tag{B.6}$$

where $C_{n\boldsymbol{k}}\,(\boldsymbol{G})$ is the Fourier coefficient of $u_{n\boldsymbol{k}}\,(\boldsymbol{r})$ at the reciprocal lattice vector $\boldsymbol{G}$. The local matrix element formula in a plane wave basis becomes:

$$\begin{aligned} M^{\mathrm{L}}_{n'\boldsymbol{k}',n\boldsymbol{k}} &= \sum_{\boldsymbol{G}} \Delta\widetilde{V}_{\mathrm{L}}(\boldsymbol{k}' - \boldsymbol{k} + \boldsymbol{G}) \\ &\times \left[\sum_{\mathbf{G}''} \sum_{\mathbf{G}'} C^{*}_{n'\boldsymbol{k}'}(\mathbf{G}'')\, C_{n\boldsymbol{k}}(\mathbf{G}')\, \delta_{\mathbf{G}'',\mathbf{G}'+\boldsymbol{G}} \right]. \end{aligned} \tag{B.7}$$

This is the formula implemented in our code and used in this work.

Appendix C

DERIVATION OF NONLOCAL PART OF ELECTRON-DEFECT MATRIX ELEMENTS

To obtain an expression for the nonlocal matrix elements, we focus on the scalar product

$$\begin{aligned}\langle \beta_j^{(s)} \,|\, n\boldsymbol{k} \rangle &= \int_{\Omega_{\mathrm{B}vK}} d^3 r \beta_j^* (\boldsymbol{r} - \tau_s)\, e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{n\boldsymbol{k}} (\boldsymbol{r}) \\ &= \int_{\Omega_{\mathrm{B}vK}} d^3 r \left[\frac{\beta_j (\boldsymbol{r} - \tau_s)}{\sqrt{\Omega_{\mathrm{uc}}}} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \right]^* \sqrt{\Omega_{\mathrm{uc}}} u_{n\boldsymbol{k}} (\boldsymbol{r})\,.\end{aligned} \tag{C.1}$$

We first fix the atomic position at the origin (by setting $\tau_s = 0$), and then generalize the result to arbitrary atomic positions. We define generalized forward and inverse Fourier transforms of the Kleinman-Bylander (KB) projectors, respectively, as:

$$\frac{\beta_j (\boldsymbol{r})}{\sqrt{\Omega_{\mathrm{uc}}}} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} = \frac{1}{\Omega_{\mathrm{B}vK}} \sum_{\boldsymbol{q}} B_{j\boldsymbol{k}} (\boldsymbol{q})\, e^{i\boldsymbol{q}\cdot\boldsymbol{r}} \tag{C.2}$$

and

$$\begin{aligned}B_{j\boldsymbol{k}} (\boldsymbol{q}) &= \int_{\Omega_{\mathrm{B}vK}} d^3 r \left[\frac{\beta_j (\boldsymbol{r})}{\sqrt{\Omega_{\mathrm{uc}}}} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \right] e^{-i\boldsymbol{q}\cdot\boldsymbol{r}} \\ &= \int_{\Omega_{sup}} d^3 r \left[\frac{\beta_j (\boldsymbol{r})}{\sqrt{\Omega_{\mathrm{uc}}}} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \right] e^{-i\boldsymbol{q}\cdot\boldsymbol{r}},\end{aligned} \tag{C.3}$$

where we replace the crystal volume $\Omega_{\mathrm{B}vK}$ with the supercell volume Ω_{sup} because the KB projector is localized around the core atomic region. If a general atomic position τ_s is chosen, the Fourier coefficient $B_{j\boldsymbol{k}}^{(s)} (\boldsymbol{q})$ becomes, using the properties of the Fourier transforms,

$$B_{j\boldsymbol{k}}^{(s)} (\boldsymbol{q}) = e^{-i(\boldsymbol{k}+\boldsymbol{q})\cdot\tau_{\mathrm{s}}} B_{j\boldsymbol{k}} (\boldsymbol{q})\,. \tag{C.4}$$

The scalar product in Eq. (C.1), after inserting Eq. (C.2) into Eq. (C.1), becomes

$$\langle \beta_j^{(s)} \,|\, n\boldsymbol{k} \rangle = \frac{1}{\sqrt{\Omega_{\mathrm{uc}}}} \sum_{\boldsymbol{G}} B_{j\boldsymbol{k}}^{(s)*} (\boldsymbol{G})\, \langle e^{i\boldsymbol{G}\cdot\boldsymbol{r}} \,|\, u_{n\boldsymbol{k}} \rangle_{\mathrm{uc}}\,, \tag{C.5}$$

where we used the translational invariance of $u_{n\boldsymbol{k}} (\boldsymbol{r})$. This formula is valid for any basis set. Here we use a plane wave basis, so the matrix elements of $\hat{V}_{\mathrm{NL}}$, for each

of the pristine and defect-containing supercells (labelled by α = d, p), read:

$$\begin{aligned} \langle n'\boldsymbol{k}'|\hat{V}^{(\alpha)}_{\mathrm{NL}}|n\boldsymbol{k}\rangle = \sum_{s_\alpha=1}\sum_{ij} D^{(s_\alpha)}_{ij}\times \\ \left[\sum_{\mathbf{G}'} B^{(s_\alpha)}_{i\boldsymbol{k}'}(\mathbf{G}')\, C^{*}_{n'\boldsymbol{k}'}(\mathbf{G}')\right]\left[\sum_{\boldsymbol{G}} B^{(s_\alpha)*}_{j\boldsymbol{k}}(\boldsymbol{G})\, C_{n\boldsymbol{k}}(\boldsymbol{G})\right], \end{aligned} \tag{C.6}$$

where we use the same notation as in Eq. (2.11). The nonlocal matrix elements are computed as the difference in Eq. (2.10). This is the formula implemented in our code and used in this work.

Appendix D

RELATIONSHIP BETWEEN DFT AND WS SUPERCELLS

The DFT supercell used to obtain the perturbation potential is not equivalent to the Wigner-Seitz (WS) supercell used to construct the WFs [Fig. D.1(a)]. The size of the WS supercell is equal to the size of the BZ coarse grid, and does not depend on the size of the DFT supercell. Let the lattice vector $\boldsymbol{R}_{\mathrm{d}}$ be the location of the unit cell containing the defect within the DFT supercell. The lattice vector $\boldsymbol{R}_{\mathrm{d}}$ can lie outside the WS supercell if the latter is not large enough to enclose the perturbation potential [see Fig. D.1(a)]. This scenario should be avoided as it leads to large interpolation errors. To include the perturbation potential in the WS supercell, we move the lattice point $\boldsymbol{R}_{\mathrm{d}}$ to the origin of the WS supercell; using a well-known property of the Fourier transform, we multiply the e-d matrix elements computed using the DFT supercell by the phase factor $e^{i(\boldsymbol{k}'-\boldsymbol{k})\cdot\boldsymbol{R}_{\mathrm{d}}}$; the resulting e-d matrix elements correctly take into account the translation of the defect site, and can be used in the e-d interpolation procedure.

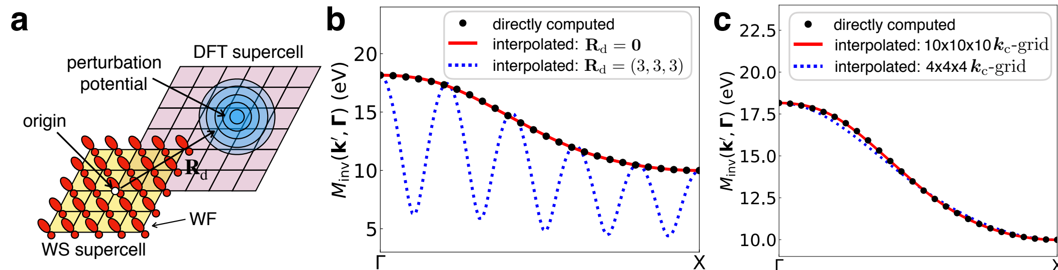

Figure D.1: Supercells and matrix elements. (a) Schematic of the DFT and WS supercells. (b) Effect of the defect position $\boldsymbol{R}_{\mathrm{d}}$ on the invariant e-d coupling $M_{\mathrm{inv}}(\boldsymbol{k}', \boldsymbol{k})$. The interpolated e-d matrix elements are computed starting from a coarse $10 \times 10 \times 10$ BZ grid. The crystal momentum of the initial state is set to the $\boldsymbol{\Gamma}$ point, and the crystal momentum $\boldsymbol{k}'$ of the final state is varied along the high-symmetry BZ line shown in figure.

To verify the accuracy of this approach, we define and compute an invariant e-d coupling, $M_{\mathrm{inv}}(\boldsymbol{k}', \boldsymbol{k})$, as

$$M_{\mathrm{inv}}(\boldsymbol{k}', \boldsymbol{k}) = \sqrt{\sum_{mn} |M_{mn}(\boldsymbol{k}', \boldsymbol{k})|^2}, \tag{D.1}$$

which is invariant under the operation of the fine-grid unitary matrix $U_{\boldsymbol{k}_\mathrm{f}}$ used to interpolate e-d matrix elements. The invariant e-d coupling computed from the interpolated matrix elements depends on the defect-site lattice vector $\boldsymbol{R}_\mathrm{d}$ through the e-d matrix elements in the Wanniner representation [see Eq. (5)]. Figure D.1(b) shows how the choice of the defect-site lattice vector $\boldsymbol{R}_\mathrm{d}$ affects the invariant e-d coupling $M_\mathrm{inv}(\boldsymbol{k}',\boldsymbol{k})$. To this end, we interpolate the e-d matrix elements starting from a coarse $10 \times 10 \times 10$ BZ grid. When the defect is placed at the lattice vector $\boldsymbol{R}_\mathrm{d} = (3,3,3)$ (in Bravais lattice vector units), the interpolated e-d matrix elements deviate significantly from the directly computed ones because the perturbation potential lies outside the WS supercell. We can improve the interpolated results by moving the lattice point $\boldsymbol{R}_\mathrm{d}$ to the origin of the WS supercell, and multiplying the matrix elements by a phase factor, as discussed above. Using this approach, the interpolated e-d matrix elements are in excellent agreement with the directly computed ones, as is shown in the $\boldsymbol{R}_\mathrm{d} = \boldsymbol{0}$ curve in Figure D.1(b). All the results in this work have been obtained by moving the defect-site lattice vector $\boldsymbol{R}_\mathrm{d}$ to the origin of the WS supercell after computing the e-d matrix elements on the coarse BZ grid. Figure D.1(c) shows how the WS supercell size affects the invariant e-d matrix elements $M_\mathrm{inv}(\boldsymbol{k}',\boldsymbol{k})$ when $\boldsymbol{R}_\mathrm{d}$ is kept fixed at the origin. The interpolated e-d matrix elements using two different coarse BZ grids of $4 \times 4 \times 4$ and $10 \times 10 \times 10$ $\boldsymbol{k}_\mathrm{c}$-points agree well with the directly computed ones. The larger WS supercell gives more accurate interpolated results. Therefore, the quality of the WF-based interpolated e-d matrix elements depends on the relative position of the defect perturbation potential in the WS supercell and the WS supercell size.