Development and application of
first-principles real-space transport calculations

Shigeru Iwase

Doctoral Program in Physics

Submitted to the Graduate School of

Pure and Applied Sciences

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Science

at the

University of Tsukuba

Abstract

Spurred by the strong demand for scaling dimension of the transistor to nanometer region where the quantum mechanical effects play a crucial role, the development of fast and reliable theoretical tools to investigate and understand such systems becomes an urgent issue. At the nanometer scale, the quantum transport calculation based on the density functional theory is promising. However, such calculation is still computationally expensive even with the state-of-the-art supercomputer and therefore more efficient algorithms are needed to perform the realistic device simulations. In this thesis, new efficient formulation, implementation, and algorithm for first-principles transport calculation are presented. In particular, the computational aspect of calculating the transport properties of Landauer-Büttiker two-probe systems within the non-equilibrium Green's function method is discussed. Herein, I employ the real-space grids to construct the Kohn-Sham Hamiltonian matrix but the extension of the developed method to the tight-binding approximation or any localized basis set is straightforward.

The most time-consuming parts of quantum transport calculation are evaluating the self-energy matrices of electrode and solving the Kohn-Sham equation under the open boundary condition. Self-energy matrices describe the boundary matching condition between electrodes and central region, and a solution of the Kohn-Sham equation is obtained by matrix inversion associated with the Green's function. For the calculation of the self-energy matrices, the formulation and implementation with the use of the partitioning and singular value decomposition techniques are derived. In addition, the contour integral eigensolver based on Sakurai-Sugiura method is developed to obtain the self-energy matrices and leads to a reduction of the computational burden by up to orders of magnitude. The efficient formula for the retarded and lesser part of the Green's functions is presented so as to compute the charge density and transmission efficiently. Using the mathematically strict relationship between the retarded and unperturbed Green's function, the physically important quantities are obtained from the first and last block matrix columns of the unperturbed Green's function which is solved efficiently by the newly developed iterative solver, modified shifted conjugate-orthogonal conjugate gradient method. To illustrate the capability of the proposed methods, several benchmarks to measure serial and parallel performance tests are conducted. As applications, transport properties of the SiC/SiO_2 interface based on power electronic device and silicene based on the nanoscale electronic devices are investigated.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Looking back on the history of the silicon integrated circuit, there is no doubt that "Moore's law" is the most important keyword. Moore's law, named from Intel co-founder Gordon Moore, states that the density of complementary metal-oxide-semiconductor (CMOS) transistor doubles every 18 to 24 months. For decades, the integrated circuit manufacturers have endeavored to remain on track of this empirical rule by pursuing the miniaturization of the transistor. As a result, the drastic progress owed to high speed, small, and low power consumption transistor has brought great deals of benefits for our life, for example, laptops, smartphone, tables have become an indispensable part of our life. However, in the mid-2000s, it became much harder to continue downsizing the transistor due to several physical limitations, and therefore conventional scaling is no longer an effective way to improve the device performance. Nowadays, to pursue further transistor downsizing, innovation of the transistor structure and novel material has undergone with rapid advancement. As successful examples in this direction, high quality native insulator SiO_2 has been replaced by the so-called "high-k" insulators, which have higher dielectric constants rather than that for SiO_2, and conventional two-dimensional MOS field effect transistors (MOSFETs) has evolved into three-dimensional FinFET. Recently, extensive efforts have been devoted to develop the future devices such as all surrounding gate nanowire FET, carbon nanotube FET, and tunnel FET, but none of them is decisive. The quest for novel device structure and material can also be seen in the research field of the power electronic device. Although silicon is currently also mainstream semiconductor in power electronic device, wide band gap semiconductors such as silicon carbide and gallium nitride emerged for high-power and high-temperature electronics. In this situation, researchers can no longer rely solely on their intuitions and past knowledge accumulated on silicon technology to find the candidates of the future devices.

Driven by the fast development of new material, new structure, and new principle electronics, atomistic-scale simulations based on the quantum mechanics become increasingly important for understanding and predicting transport properties. In the past, technology computer aided design (TCAD) tool based on the drift-diffusion and Boltzsmann transport equation was frequently used to predict the device properties with practically sufficient accuracy. However, as decreasing the dimension of the transistor, the effects of quantum mechanical effects like quantum confinement and tunneling on

1

device properties become too large to ignore. The empirical tight-binding approximation is one of the simplest ways to include the quantum mechanical effects and it may give the reasonable results in some cases. Although the tight-binding approximation shows a good balance between accuracy and efficiency, there is severe problem that it is difficult to obtain the transferable parameters from the experimental measurements. Then, the tight-binding approximation often fails in the system containing impurity, surface and interface, where the influence of the individual atoms is essential. At the atomistic scale, the first-principles approach whose Hamiltonian matrix is constructed self-consistently without any empirical parameter should be the ultimate solution.

The great majority of first-principles electron transport calculations have been based on the density functional theory (DFT) within the local density or generalized gradient approximations. The first attempts to the understand electron transport were based on the Landauer formula, implemented by the Lippmann-Schwinger method. The current-voltage curve of the nanostructure between semi-infinite electrodes was evaluated by solving the time-independent Kohn-Sham equation under the open boundary condition. At this stage, there were many limitations and restrictions such that crystalline electrodes are approximated by jellium electrodes, contributions from electron-phonon and electron-electron interactions are difficult to include, and self-consistent calculations are numerically unstable. To overcome these difficulties, theoretical and computational improvements have been studied extensively. Currently, the most successful approach at the level of DFT calculation is the non-equilibrium Green's function (NEGF) method often combined with the localized basis set. However, in the DFT calculations of electron transport, a reasonable description of actual experiments involving over thousands atoms is still computationally demanding, and therefore, the development of an efficient computational method remains a vital task. Especially in the framework of the NEGF method, the evaluation of the self-energy matrices of electrodes and retarded Green's function is the hotspot of the whole computation for large systems. A major contribution of this thesis is that the quantum transport calculation of nanoscale conductor containing thousands of atoms makes it feasible.

In this work, I present new computational methods for quantum transport simulations based on the real-space finite-difference scheme. In order to obtain the transmission probability and other physical quantities, I have developed the efficient formulas and algorithms for the computationally most expensive parts of the transport calculations. In addition, fully utilizing the sparsity of the Hamiltonian matrix, the proposed methods are suitable for massively parallel computing. Although all derivations in this thesis are assumed to use real-space grids, extensions to the tight-binding approximation and localized basis set calculations are straightforward, and I believe that the efficient algorithms presented here makes first-principles investigations of transport properties in large-scale atomic structures feasible. My contributions to this work includes

- Derivation of the efficient formula and implementation for the self-energy matrices based on real-space finite-difference scheme combined with partitioning and singular value decomposition techniques.

- Implementation of the Green's function method based on the sparsity of the self-energy matrices

and Dyson equation. Implementation of the shifted conjugate-orthogonal conjugate-gradient method for the unperturbed Green's function to perform transmission calculations efficiently.

- Development and implementation of a contour integral method for calculation of generalized Bloch states, complex band structure, and self-energy matrices.

- The study of transport properties of the SiC/SiO2 interface based on the power electronic device.

- The study of transport properties of the silicene interface based on the nanoscale electronic device.

1.1   Units and notation

Unless otherwise noted, Hartree atomic units ($m_e = \hbar = e = 1$) are used throughout this thesis. Following unit conversions are often used.

$$
\begin{aligned}
1 \text{ unit of Length} &= 0.5292 \text{ (Å)} \\
1 \text{ unit of Energy} &= 27.2114 \text{ (eV)} \\
1 \text{ unit of Time} &= 0.02418 \text{ (fs)} \\
1 \text{ unit of Electron mass} &= 9.1096 \times 10^{-31} \text{ (kg)} \\
1 \text{ unit of Electron charge} &= 1.6022 \times 10^{-19} \text{ (C)} \\
1 \text{ unit of angular momentum} &= 1.0546 \times 10^{-34} \text{ (Js)}
\end{aligned}
$$

The following mathematical notations are used in this thesis. Scalars are written as italic characters, a, b, x, y, z, etc. Integer variables are predominantly given in i, j, k, N, etc. Vectors are denoted by bold face characters, $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$, etc, and matrices are predominantly with hat, $\hat{A}, \hat{B}, \hat{X}, \hat{Y}, \hat{Z}$, etc. To indicate the specific element of vector, I use the notation $a_i = [\boldsymbol{a}]_i$ and similarly $A_{i,j} = [\hat{A}]_{i,j}$ for matrices. Finally, the particular block matrix element of vectors and matrices are denoted by $\boldsymbol{a}_i$ and $\hat{A}_{i,j}$, respectively.

1.2   Outline

The thesis is composed of 7 chapters, 4 appendices, and the bibliography. In Chap. 2, I briefly review the framework of the density functional theory and derive the Kohn-Sham equation for the norm-conserving pseudopotential. In Chap. 3, I discuss the state-of-the-art theoretical modeling of the quantum transport based on the Landauer picture in weakly interacting phase-coherent nanoscale conductor. In Chap. 4, I describe the formulation and implementation of self-energy matrices evaluation based on the real-space finite-difference scheme. In Chap. 5, I present the formalism for a more efficient implementation of the retarded and non-equilibrium Green's functions to compute the charge density and transmission. In addition, an efficient numerical solver to obtain the reduced solution of

the shifted linear systems arising from the unperturbed Green's function calculation is implemented. In Chap. 6, I describe a contour integral method for the fast evaluation of the generalized Bloch states, complex band structure, and self-energy matrices within the formalism of the Green's function or wavefunction-matching method. In Chap. 7, I give a short conclusion and outlook.

Chapter 2

Density Functional Theory

Density functional theory (DFT) is the first-principles theoretical approach to treat the many-body electron system in the ground state. From the pioneering work by Hohenberg and Kohn [1] in 1964, numerous efforts have been devoted to enhance the capability of the DFT calculations, for examples, proposals of the Kohn-Sham equation [2], *ab initio* norm-conserving pseudopotential [3], various exchange correlation functionals [4–6], Car-Parrinello molecular dynamics method [7], and so on. Owing to these theoretical breakthroughs and remarkable advance on the super computers, current DFT-based simulations enable us to investigate the electronic properties of a wide range of materials accurately as well as the structural changes and phonon spectral theoretically. Nowadays, beyond the ground state properties, electronic excitations including electron-transport and optical properties are tractable by combined with theories such as Landauer-Büttiker formalism [8] and time-dependent DFT [9]. I would like to begin this thesis from the minimum review of the DFT.

2.1 Born-Oppenheimer approximation

The purpose of condensed matter physics is to understand the behavior of electrons in matter including gas, liquid, and solids that consist of a lot of electrons and atomic nuclei. The non-relativistic Hamiltonian for the system of electrons and atomic nuclei can be written as below:

$$\mathscr{H} = \sum_i (-\frac{1}{2}\nabla_i^2) + \frac{1}{2}\sum_{i\neq j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,n}\frac{Z_n}{|\mathbf{r}_i - \mathbf{R}_n|} + \sum_n(-\frac{1}{2M_n}\nabla_n^2) + \frac{1}{2}\sum_{n\neq m}\frac{Z_nZ_m}{|\mathbf{R}_n - \mathbf{R}_m|}, \quad (2.1)$$

where $\mathbf{r}_i$ is the position of the i-th electron and $\mathbf{R}_n, M_n$ and Z_n are the position, mass and charge of the n-th nucleus, respectively.

In addition, Schrödinger equation for the Hamiltonian in Eq. (2.1) is

$$\mathscr{H}\Phi = \mathscr{E}\Phi. \quad (2.2)$$

It is assumed that the eigenfunction of the above equation can be expressed by the product of the wavefunction of electrons and that of nuclei

$$\Phi(\mathbf{r_1}, ...; \mathbf{R_1}, ...) = \Psi(\mathbf{r_1}, ...; \mathbf{R_1}, ...)\Lambda(\mathbf{R_1}, ...). \quad (2.3)$$

Here Ψ is the eigenfunction of the electronic Hamiltonian (2.1)

$$H\Psi(\mathbf{r_1}, ...; \mathbf{R_1}, ...) = E(\mathbf{R_1}, ...)\Psi(\mathbf{r_1}, ...; \mathbf{R_1}, ...), \tag{2.4}$$

where

$$H = \sum_i (-\frac{1}{2}\nabla_i^2) + \frac{1}{2}\sum_{i\neq j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,n}\frac{Z_n}{|\mathbf{r}_i - \mathbf{R}_n|}. \tag{2.5}$$

Note that the atomic position $\mathbf{R}_n$ contributes to the electron system as just a parameter, and E depends on the atomic positions. However, one might think that the vibration of the atomic nuclei around the equilibrium position at the finite temperature will affect the electrons. To see this effect, substituting Eq. (2.3) into Eq. (2.2) and using Eq. (2.4), one might obtain

$$\mathscr{H}\Phi = \Psi\Big[\sum_n(-\frac{1}{2M_n}\nabla_n^2) + E(\mathbf{R_1}, ...) + \frac{1}{2}\sum_{n\neq m}\frac{Z_n Z_m}{|\mathbf{R}_n - \mathbf{R}_m|}\Big]\Lambda - \sum_n\frac{1}{2M_n}\Big(2\nabla_n\Psi\nabla_n\Lambda + \nabla_n^2\Psi\Lambda\Big). \tag{2.6}$$

The second term in Eq. (2.6) is the electron-phonon interaction. If the electron-phonon interaction can be ignored, the freedoms of degree for electrons and nuclei systems are separated completely, and therefore, this approximation is called adiabatic approximation or Born-Oppenheimer approximation. Actually, by calculating the expected values of the first and second terms of the electron-phonon interaction in Eq. (2.6), its effect is much smaller than the kinetic energy of electrons by the order of $1/M_n = 10^{-3} \sim 10^{-5}$. This approximation seems to be quite good.

Thanks to the Born-Oppenheimer approximation, electron-nuclei interactions are treated as the external potentials, and the target problem becomes solving Eq. (2.4) instead of Eq. (2.2). However, it is still very difficult to solve Eq. (2.4) due to the electron-electron interactions.

2.2   Hohenberg-Kohn theorem

In this section I explain the theoretical framework of DFT to obtain the ground state of Schrödinger equation for Hamiltonian of N-electron system. As a preparation for deriving the Hohenberg-Kohn theorem [1], I introduce the external potential V_{ext} and electron particle number operator $\hat{n}(\mathbf{r})$ defined as

$$V_{ext} = \sum_i^N v_{ext}(\mathbf{r}_i), \tag{2.7}$$

$$\hat{n}(\mathbf{r}) = \sum_{i=1}^N \delta(\mathbf{r} - \mathbf{r}_i), \tag{2.8}$$

where $v_{ext}(\mathbf{r}_i)$ is an external potential acting on the i-th electron:

$$v_{ext}(\mathbf{r}_i) = \sum_n\frac{Z_n}{|\mathbf{r}_i - \mathbf{R}_n|}. \tag{2.9}$$

The electron density is obtained as a expected value of $\hat{n}(\boldsymbol{r})$ for the N-electron wavefunction,

$$
\begin{aligned}
n(\mathbf{r}) &= \langle\Psi|\,\hat{n}\,|\Psi\rangle \\
&= \int \Psi(\mathbf{r}_1,...,\mathbf{r}_N)\hat{n}(\mathbf{r})\Psi(\mathbf{r}_1,...,\mathbf{r}_N)d\mathbf{r}_1...d\mathbf{r}_N.
\end{aligned}
\tag{2.10}
$$

In addition, potential energy for V_{ext} is expressed using the electron density,

$$
\begin{aligned}
\langle\Psi|\,V_{ext}\,|\Psi\rangle &= \int \Psi(\mathbf{r}_1,...,\mathbf{r}_N)\sum_i^N v_{ext}(\mathbf{r_i})\Psi(\mathbf{r}_1,...,\mathbf{r}_N)d\mathbf{r}_1...d\mathbf{r}_N \\
&= \int \Psi(\mathbf{r}_1,...,\mathbf{r}_N)\sum_i^N \left(\int v_{ext}(\mathbf{r})\delta(\mathbf{r}-\mathbf{r_i})d\mathbf{r}\right)\Psi(\mathbf{r}_1,...,\mathbf{r}_N)d\mathbf{r}_1...d\mathbf{r}_N \\
&= \int v_{ext}(\mathbf{r})n(\mathbf{r})d\mathbf{r}.
\end{aligned}
\tag{2.11}
$$

Following the quantum mechanics, if V_{ext} is given, N-electron wavefunction and electron density are determined uniquely except for degeneration. The first Hohenberg-Kohn theorem is its reverse.

Theorem 1 *For a given electron density, external potential and ground state wavefunction are determined uniquely except for degeneration.*

Proof. Assume that there are two different external potentials $v_{ext}(\mathbf{r})$ and $v'_{ext}(\mathbf{r})$ that give the same electron density $n(\mathbf{r})$. If H and H' denote the corresponding Hamiltonians for $v_{ext}(\mathbf{r})$ and $v'_{ext}(\mathbf{r})$, ground state wavefunctions Ψ_{GS} and Ψ'_{GS} are different unless $v_{ext}(\mathbf{r}) - v'_{ext}(\mathbf{r}) = const$. From the variational principle, ground state energy E_{GS} must satisfy

$$
\begin{aligned}
E_{GS} = \langle\Psi_{GS}|\,H\,|\Psi_{GS}\rangle &< \langle\Psi'_{GS}|\,H\,|\Psi'_{GS}\rangle \\
&= \langle\Psi'_{GS}|\,(H'-V'+V)\,|\Psi'_{GS}\rangle \\
&= E'_{GS} + \int (v_{ext}(\mathbf{r})-v'_{ext}(\mathbf{r}))n(\mathbf{r})d\mathbf{r},
\end{aligned}
\tag{2.12}
$$

and also for E'_{GS},

$$
\begin{aligned}
E'_{GS} = \langle\Psi'_{GS}|\,H'\,|\Psi'_{GS}\rangle &< \langle\Psi_{GS}|\,H'\,|\Psi_{GS}\rangle \\
&= \langle\Psi_{GS}|\,(H-V+V')\,|\Psi_{GS}\rangle \\
&= E_{GS} + \int (v_{ext}(\mathbf{r})'-v_{ext}(\mathbf{r}))n(\mathbf{r})d\mathbf{r}.
\end{aligned}
\tag{2.13}
$$

Adding Eq.(2.12) and Eq.(2.13) yields

$$
E_{GS} + E'_{GS} < E_{GS} + E'_{GS}.
\tag{2.14}
$$

The above equation is contradictory. Thus, $v_{ext}(\mathbf{r})$ and $v'_{ext}(\mathbf{r})$ that give the same $n(\mathbf{r})$ do not exist (End of Proof).

The first Hohenberg-Kohn theorem asserts the one-to-one mapping between $n(\mathbf{r})$ and $v_{ext}(\mathbf{r})$, that is, all physical properties of ground state are described by the universal functional of $n(\mathbf{r})$. For example, ground state energy can be described as a functional of $n(\mathbf{r})$

$$E_{GS}[n] = F[n] + \int v_{ext}(\mathbf{r})n(\mathbf{r})d\mathbf{r}, \tag{2.15}$$

where

$$F[n] = \langle \Psi_{GS}| \left(T + V_{ee}\right) |\Psi_{GS}\rangle . \tag{2.16}$$

Here, T and V_{ee} are the kinetic energy of electrons and electron-electron interactions, respectively. The first Hohenberg-Kohn theorem only shows the existence and uniqueness of $n(\mathbf{r})$ but says nothing about how to obtain the true electron density. The second Hohenberg-Kohn theorem concerns this issue.

Theorem 2 *Electron density which minimizes the ground state energy is the exact electron density.*

Proof. Following the first Hohenberg-Kohn theorem, for a trial electron density $n'(\mathbf{r})$, there exists the corresponding external potential $v'_{ext}(\mathbf{r})$ and ground state wavefunction Ψ'_{GS}. From the variational principles, ground state energy for $n'(\mathbf{r})$ is obtained as below:

$$
\begin{aligned}
E_{GS}[n'] &= \langle \Psi'_{GS}| H |\Psi'_{GS}\rangle \\
&= \int v_{ext}(\mathbf{r})n'(\mathbf{r})d\mathbf{r} + F[n'] \\
&\geq E_{GS}[n] = \int v_{ext}(\mathbf{r})n(\mathbf{r})d\mathbf{r} + F[n].
\end{aligned}
\tag{2.17}
$$

(End of Proof)

Second Hohenberg-Kohn theorem asserts that the ground state energy E_{GS} is obtained once we know the exact electron density $n(\mathbf{r})$. The second theorem is too simple but important because it reduces the problem of finding the wavefunction $\Psi(\mathbf{r}_1, ..., \mathbf{r}_N)$ with $3N$ variables into finding the electron density $n(\mathbf{r})$ with 3 variables. Although the complexity of the problem is dramatically reduced, the specific form of the functional $F[n]$ in Eq. (2.16) is still unclear.

Note that the first and second Hohenberg-Kohn theorems assume the existence of the external potentials that reproduce the ground state electron density (V-representability). In addition, the second theorem further assumes that the ground state electron density is obtained from the anti-symmetrized wavefunction (N-representability). Strictly speaking, these are not correct. However, using the constrained search formulation proposed by Levy [10, 11], it is enough to consider the N-representability only. Fortunately, N-representability is not problematic in practical calculations because the electron density constructed by the solutions of the Kohn-Sham equation to be appeared in the next section satisfies the N-representability.

2.3   Kohn-Sham equation

In the following year of the paper by Hohenberg and Kohn, Kohn and Sham derived the single-particle equation known as Kohn-Sham equation and opened the way of the electronic structure calculation based on DFT. The central idea of the Kohn-Sham method is introducing the fictitious system, which is a set of non-interacting electrons whose electron density is identical to that of the true many-body system. The electrons of the fictitious system moves in the local effective potential $v_{eff}(\mathbf{r})$, instead of ignoring the electron-electron interaction. So the Hamiltonian operator is described by just a sum of the one-electron operator, and we simply solve the single-particle (Kohn-Sham) equation;

$$\left[-\frac{1}{2}\nabla^2 + v_{eff}(\mathbf{r}) \right]\psi_i(\mathbf{r}) = \varepsilon_i\psi_i(\mathbf{r}) \tag{2.18}$$

for N non-interacting electrons. Here, $\psi_i(\mathbf{r})$ and ε_i are known as the Kohn-Sham orbital and energy, respectively. When the full wavefunction is expressed by the Kohn-Sham orbitals as the form of a Slater determinant, the true electron density will be given by

$$n(\mathbf{r}) = \sum_i^N |\psi_i(\mathbf{r})|^2. \tag{2.19}$$

In order to obtain $v_{eff}(\mathbf{r})$, $F[n]$ might be decomposed into three terms:

$$F[n] = T_s[n] + \frac{1}{2}\int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}d\mathbf{r}d\mathbf{r}' + E_{xc}[n], \tag{2.20}$$

where $T_s[n]$ is the kinetic energy of non-interaction electrons, and the second term is the classical Coulomb interaction contributed from the charge density $n(\mathbf{r})$, and $E_{xc}[n]$ is the exchange-correlation energy which includes the kinetic energy and electron-electron interaction of interacting electrons except for the first and second terms in Eq. (2.20). Using the Kohn-Sham orbitals $\psi_i(\mathbf{r})$, $T_s[n]$ can be described as

$$\begin{aligned} T_s[n] &= \sum_i^N \int \psi_i^*(\mathbf{r})(-\frac{1}{2}\nabla^2)\psi_i(\mathbf{r})d\mathbf{r} \\ &= \sum_i^N \varepsilon_i - \int v_{eff}(\mathbf{r})n(\mathbf{r})d\mathbf{r}. \end{aligned} \tag{2.21}$$

The ground state energy $E_{GS}[n]$ in the true many-body system is

$$E_{GS}[n] = \sum_i^N \varepsilon_i - \int v_{eff}(\mathbf{r})n(\mathbf{r})d\mathbf{r} + \int v_{ext}(\mathbf{r})n(\mathbf{r})d\mathbf{r} + \frac{1}{2}\int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}d\mathbf{r}d\mathbf{r}' + E_{xc}[n]. \tag{2.22}$$

By the way, the guideline of determining the unknown $v_{eff}(\mathbf{r})$ is the variational principle for $E_{GS}[n]$. Under the constraint that total number of electron is constant, we can derive the variational equation

$$\frac{\delta}{\delta n}\left[E_{GS}[n] - \mu\left(\int n(\mathbf{r})d\mathbf{r} - N\right)\right] = -v_{eff}(\mathbf{r}) + v_{ext}(\mathbf{r}) + \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}d\mathbf{r}' + v_{xc}(\mathbf{r}) - \mu = 0, \tag{2.23}$$

where μ is the Lagrange multiplier and

$$v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[n]}{\delta n}. \tag{2.24}$$

The $v_{xc}(\mathbf{r})$ is called exchange-correlation potential and its analytical expression will be given in next section. As can be seen from Eq. (2.18), because μ uniformly shift ε_i, we put $\mu = 0$. The forth term in Eq. (2.23), the Hartree potential, is

$$v_H(\mathbf{r}) = \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}'. \tag{2.25}$$

Because the direct integration of the $v_H(\mathbf{r})$ is very expensive, it is usually evaluated by solving the Poisson equation

$$\nabla^2 v_H(\mathbf{r}) = -4\pi n(\mathbf{r}), \tag{2.26}$$

under the appropriate boundary condition. Typical numerical techniques to solve the Poisson equation are the conjugate-gradient method, multi-grid method, and fast Fourier transform method. Thus, the $v_{eff}(\mathbf{r})$ can be rewritten as below:

$$v_{eff}(\mathbf{r}) = v_{ext}(\mathbf{r}) + v_H(\mathbf{r}) + v_{xc}(\mathbf{r}). \tag{2.27}$$

As a result, our target problem, that is, calculating the ground state of the N-electron system is reduced to solve Eq. (2.18), Eq. (2.19), and Eq. (2.27) self-consistently. In the actual calculation, the self-consistent solution is obtained by following iterative steps.

Step 1. Guess the trial electron density $\tilde{n}(\mathbf{r})$ and Kohn-Sham orbital $\tilde{\psi}_i(\mathbf{r})$

Step 2. Construct the effective local potential $v_{eff}(\mathbf{r})$

Step 3. Solve the Kohn-Sham equation (2.18) to obtain the new $n(\mathbf{r})$ and $\psi_i(\mathbf{r})$.

Step 4. if $|n(\mathbf{r}) - n'(\mathbf{r})|$ is sufficiently small, the procedure stops and evaluate the physical quantities such as E_{GS}. Otherwise mix the new and old densities and go to Step 2.

Fig. 2.1 illustrates the self-consistent procedure in a DFT calculation.

2.4   Exchange correlation functionals

The Kohn-Sham equation derived in the previous section is a highly efficient approach for quantum many-body problem, but the exact exchange-correlation energy $E_{xc}[n]$ and its functional derivative $v_{xc}(\mathbf{r})$ are still unclear, and unfortunately it might be quite difficult to give its exact expression. Kohn and Sham have already pointed out in their paper that many of solids can be regarded as a homogeneous electron gas limit. In this limit, $E_{xc}[n]$ can be well approximated by the local or semi-local functional of the electron density. Although various analytical forms for $E_{xc}[n]$ have been proposed so far, it is beyond the scope of this thesis to derive them. Here, I only mention typical ones, the local density approximation and generalized gradient approximation, that are used in the current calculations.
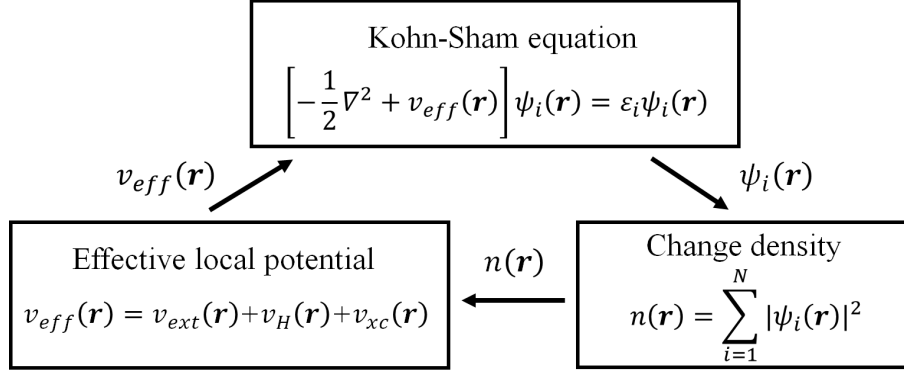
Figure 2.1: Schematics of self-consistent calculation in Kohn-Sham scheme. The cycle is continued until solution of the Kohn-Sham equation $\psi_i(\mathbf{r})$, charge density $n(\mathbf{r})$ and $v_{eff}(\mathbf{r})$ become self-consistent.

2.4.1 Local density approximation

Local density approximation (LDA) is the simplest approximation which assumes $E_{xc}[n]$ is given by

$$E_{xc}[n] = \int n(\mathbf{r})\epsilon_{xc}^{LDA}(n(\mathbf{r}))d\mathbf{r}, \tag{2.28}$$

where ϵ_{xc}^{LDA} is the exchange-correlation energy per electron for the homogeneous electron gas. Furthermore, ϵ_{xc}^{LDA} can be divided into exchange and correlation terms

$$\epsilon_{xc}^{LDA}(n(\mathbf{r})) = \epsilon_x^{LDA}(n(\mathbf{r})) + \epsilon_c^{LDA}(n(\mathbf{r})). \tag{2.29}$$

ϵ_x^{LDA} is so called Slater exchange energy and given by

$$\epsilon_x^{LDA}(n) = -\frac{3}{4}\left(\frac{6}{\pi}n\right)^{1/3}. \tag{2.30}$$

For the correlation energy ϵ_c^{LDA}, several analytical forms have been proposed. The commonly used analytical representations of ϵ_c^{LDA} were formulated by Perdew and Zunger [12], Vosko, Wilks, and Nusair [13], and Perdew and Wang [14]. Because all these expressions were obtained by fitting the quantum Monte Carlo calculation by Ceperley and Alder [15] for the homogeneous electron gas, their results are qualitatively similar. For example, the analytical expression formulated by Perdew and Wang is given by

$$
\begin{aligned}
\epsilon_c^{LDA}(r_s) = {} & -2 \times 0.031091(1 + 0.21370r_s) \\
& \times \ln\left(1 + \frac{1}{2 \times 0.031091(7.5957r_s^{1/2} + 3.5876r_s + 1.6382r_s^{3/2} + 0.49294r_s^2)}\right),
\end{aligned} \tag{2.31}
$$

where r_s is the density parameter given by

$$r_s = \left(\frac{3}{4\pi n}\right)^{1/3}. \tag{2.32}$$

The exchange-correlation potential $v_{xc}(\mathbf{r})$ for LDA can be obtained by the functional derivative of $E_{xc}[n]$ in Eq. (2.28), that is,

$$
\begin{aligned}
v_{xc}(\mathbf{r}) &= \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})} \\
&= \int \left[\frac{\delta n(\mathbf{r}')}{\delta n(\mathbf{r})} \epsilon_{xc}^{LDA}(n(\mathbf{r}')) + n(\mathbf{r}') \frac{\delta \epsilon_{xc}^{LDA}(n(\mathbf{r}'))}{\delta n(\mathbf{r})} \right] d\mathbf{r}' \\
&= \int \left[\epsilon_{xc}^{LDA}(n(\mathbf{r}')) + n(\mathbf{r}') \frac{\delta \epsilon_{xc}^{LDA}(n(\mathbf{r}'))}{\delta n(\mathbf{r}')} \right] \frac{\delta n(\mathbf{r}')}{\delta n(\mathbf{r})} d\mathbf{r}' \\
&= \epsilon_{xc}^{LDA}(n(\mathbf{r})) + n(\mathbf{r}) \frac{\partial \epsilon_{xc}^{LDA}(n(\mathbf{r}))}{\partial n(\mathbf{r})}.
\end{aligned}
\tag{2.33}
$$

It is known that the LDA is generally accurate enough to describe the structural properties including atomic structure, lattice constant and bulk modulus for a wide range of materials. The reason why the LDA is good approximation even for inhomogeneous gas can be explained by two reasons: (i) exchange-correlation energy only depends on the angle-averaged exchange hole and the LDA result is similar to the exact one, (ii) the LDA functional satisfies the charge sum rule [16].

2.4.2　Generalized gradient approximations

While the LDA works well beyond our expectation, it turned out that the LDA results are unsatisfactory in some cases such as the overbinding of the molecules, overestimation of cohesive energies, poor accuracy of activation energy barriers in chemical reactions, wrong orders of the structural energy differences, descriptions for $3d$-metals and its oxides, and so on.

Generalized gradient approximations (GGAs) are exchange correlation energies that involve the first-order gradient of the electron density to improve the description of the systems where the density changes rapidly such as in molecules. The GGA exchange-correlation energy is expressed by

$$
\begin{aligned}
E_{xc}[n] &= \int n(\mathbf{r}) \epsilon_{xc}^{GGA}(n^{\uparrow}, n^{\downarrow}, |\nabla n^{\uparrow}|, |\nabla n^{\downarrow}|) d\mathbf{r} \\
&= \int n(\mathbf{r}) \epsilon_{xc}^{LDA}(n^{\uparrow}, n^{\downarrow}) F_{xc}(n^{\uparrow}, n^{\downarrow}, |\nabla n^{\uparrow}|, |\nabla n^{\downarrow}|) d\mathbf{r},
\end{aligned}
\tag{2.34}
$$

where $n^{\uparrow}$ and $n^{\downarrow}$ are up- and down-spin electron densities. Although various successful forms of F_{xc} have been proposed so far, I only introduce the GGA proposed by Perdew, Burke, and Ernzerhof (PBE) [6], which might be the simplest and most widely used GGA functional in the field of the solid state physics. From the scaling rule for the exchange energy, it is enough to consider the spin-degenerated PBE exchange energy

$$
F_x^{PBE}(s) = 1 + \kappa - \frac{\kappa}{1 + \mu s^2 / \kappa},
\tag{2.35}
$$

where $\kappa = 0.804, \mu = 0.21951$ and

$$
s = \frac{|\nabla n|}{2 k_F n}.
\tag{2.36}
$$

Here, $k_F = (3\pi^2 n)^{1/3}$. On the other hand, for the correlation energy, the spin dependencies must be treated explicitly,

$$F_c^{PBE}(n^\uparrow, n^\downarrow, |\nabla n^\uparrow|, |\nabla n^\downarrow|) = 1 + \frac{H^{PBE}(r_s, t, \zeta)}{\epsilon_c^{LDA}(r_s, \zeta)}, \tag{2.37}$$

where

$$t = \frac{|\nabla n|}{2\xi k_s n}, \tag{2.38}$$

$$\zeta = \frac{n^\uparrow - n^\downarrow}{n}. \tag{2.39}$$

Here, $k_s = (4k_F/\pi)^{1/2}$ and $\xi = [(1+\zeta)^{2/3} + (1-\zeta)^{2/3}]/2$. The PBE correlation term $H(r_s, t, \zeta)$ in Eq. (2.37) is given by

$$H^{PBE}(r_s, t, \zeta) = \gamma\xi^3 \ln\left(1 + \frac{\beta}{\gamma}t^2\left[\frac{1 + At^2}{1 + At^2 + A^2t^4}\right]\right), \tag{2.40}$$

where $\beta = 0.066725, \gamma = 0.031091$, and

$$A = \frac{\beta}{\gamma}[e^{-\epsilon_c^{LDA}(r_s, \zeta)/(\gamma\xi^3)} - 1]^{-1}. \tag{2.41}$$

The GGA exchange-correlation potential $v_{xc}(\mathbf{r})$ can be obtained by the functional derivative for E_{xc} with respective to the n and ∇n

$$\delta E_{xc}[n] = \sum_\sigma \int \left[\epsilon_{xc}^{GGA}\delta n^\sigma + n\frac{\partial\epsilon_{xc}^{GGA}}{\partial n^\sigma}\delta n^\sigma + n\frac{\partial\epsilon_{xc}^{GGA}}{\partial|\nabla n^\sigma|}\delta|\nabla n^\sigma|\right]d\mathbf{r}, \tag{2.42}$$

where $\sigma = \uparrow$ or $\downarrow$. Using $\delta\nabla n = \nabla\delta n$,

$$\delta|\nabla n| = \delta|\nabla n|\frac{|\nabla n|}{|\nabla n|} = \delta\nabla n \cdot \frac{\nabla n}{|\nabla n|} = \frac{\nabla n}{|\nabla n|} \cdot \nabla\delta n. \tag{2.43}$$

Substituting Eq. (2.43) into Eq. (2.42) results in

$$\delta E_{xc}[n] = \sum_\sigma \int \left[\epsilon_{xc}^{GGA} + n\frac{\partial\epsilon_{xc}^{GGA}}{\partial n^\sigma} + n\frac{\partial\epsilon_{xc}^{GGA}}{\partial|\nabla n^\sigma|}\frac{\nabla n^\sigma}{|\nabla n^\sigma|} \cdot \nabla\right]\delta n^\sigma d\mathbf{r}. \tag{2.44}$$

Thus, the tractable form of $v_{xc}(\mathbf{r})$ can be obtained as below

$$v_{xc}^\sigma(\mathbf{r}) = \frac{\delta E_{xc}[n]}{\delta n^\sigma(\mathbf{r})} = \left[\epsilon_{xc}^{GGA} + n\frac{\partial\epsilon_{xc}^{GGA}}{\partial n^\sigma} + n\frac{\partial\epsilon_{xc}^{GGA}}{\partial|\nabla n^\sigma|}\frac{\nabla n^\sigma}{|\nabla n^\sigma|} \cdot \nabla\right]_{\mathbf{r}, \sigma}. \tag{2.45}$$

The last term in Eq. (2.45) acts as the differential operator, resulting in $v_{xc}^\sigma(\mathbf{r})$ is the non-local potential. For further details on how to handle the last term, we refer the reader to Ref. [17].

2.5   Pseudopotential method

I have reviewed the minimal theoretical foundations of the DFT in the sections above. Next, I shall consider solving the Kohn-Sham equation (2.18) numerically. In practice, the Kohn-Sham orbital is expanded by the real-space grid, plane-wave basis set, or pseudo-atomic basis, which transforms the Kohn-Sham (second-order differential) equation into the matrix eigenvalue problem. Because the computational cost of solving the eigenvalue problem depends on its dimension, i.e., the number of basis set, it is useful to discuss how many numbers of the basis set are required to describe the Kohn-Sham orbital. Here, I choose the aluminum (Al) in bulk and plane wave basis set as a simple example. The Al in bulk forms the face-centered cubic structure with lattice constant $a = 4.05$Å. Based on the hydrogen model, the radius of the lowest $1s$ orbital of the Al is roughly $Z_{\mathrm{Al}} (= 13)$ times smaller than that of the hydrogen, that is, $b = 0.529/13 = 0.04$Å. To describe the $1s$ orbital, plane waves with the wave vector $k_{max} \approx 2\pi/b$ will be required, and the reciprocal lattice vector of Al, which can be considered as a minimum wave vector, is $k_{min} \approx 2\pi/a$. Thus, the required number of plane-wave basis is estimated by $(k_{max}/k_{min})^3 = (a/b)^3 \approx 10^6$. This number is very large. On the other hand, the eigenvalue of the $1s$ orbital of the Al is Z_{Al}^2 times deeper than that of the hydrogen, that is, -2300 eV. Considering that the metallic properties of the Al bulk can be well approximated by the nearly free electron model, the deep and specially localized $1s$ orbital of the Al in bulk only weakly affects on the physical property. Thus, it is reasonable and computationally efficient to ignore the inner-shell electrons and deal with the outer-shell valence electrons explicitly. This is the basic concept of the pseudopotential method.

The pseudopotential method has a long history. Several elementary ideas associated with pseudopotentials were recognized in mid-1930s [18,19], however the practical calculations for predicting the band structure in solids had not been achieved until the empirical pseudopotential method (EPM) was developed. The EPM is based on the Philips-Kleinman cancellation theorem [20], which enables us to transform the hard all-electron potential into a smooth pseudopotential. Because of this property, plane-wave basis set can be used efficiently with the pseudopotentials. While the EPM has been used to understand the optical and dielectric properties of semiconductors [21], this method only works for the systems on which parameters are fitted. This problem is known as the transferability problem and finding the transferable pseudopotentials is still elusive within the empirical pseudopotential approach. Moreover, there is no reason to believe the EPM for predicting the structural properties of materials because the empirical parameters are fitted to the only optical excitation. To overcome these difficulties, the DFT-based pseudopotentials have been studied extensively until now and result in the impressive advances for understanding the electronic structure of materials. In this section, I will introduce the norm-conserving pseudopotential method which is one of the most popular method used in the current DFT calculations.

2.5.1   Norm-conserving pseudopotentials

The pseudopotential method incorporates the inner-shell electrons into the external potentials and handles valence electrons only, that is, external potentials for nuclei in Eq. (2.18) are replaced by the pseudopotetials for nulcei plus inner-shell electrons. Pseudopotentials are constructed to make the pseudo wavefunctions as smooth as possible while keeping the accuracy and transferability. Here *transferable* pseudopotentials are those constructed in the isolated atom, but describe the behaviors of valence electrons in different environments including molecules, clusters, and solids. However, pseudopotentials are not determined uniquely due to its fuzzy definition. The most frequently used pseudopotentials are probably the norm-conserving pseudopotenials (NCPPs) developed by the Hamann, Schlüter and Chiang (HSC) [3], and its improvements [22–25]. According to the HSC paper, for an atomic reference configuration, NCPPs must satisfy the following requirements:

(1) $\varepsilon_l^{ps} = \varepsilon_l^{ae}$,

(2) $\phi_l^{ps}(r) = \phi_l^{ae}(r)$ for $r > r_c$,

(3) $\phi_l^{ps}(r_c) = \phi_l^{ae}(r_c)$ and $\left.\frac{d\phi_l^{ps}(r)}{dr}\right|_{r=r_c} = \left.\frac{d\phi_l^{ae}(r)}{dr}\right|_{r=r_c}$,

(4) $\int_0^{r_c} |\phi_l^{ps}(r)|^2 r^2 dr = \int_0^{r_c} |\phi_l^{ae}(r)|^2 r^2 dr = Q_c$,

where ε_l^{ps} and ε_l^{ae} are pseudo and all-electron valence eigenvalues for angular momentum l, and $\phi_l^{ps}(r)$ and $\phi_l^{ae}(r)$ are the radial part of pseudo- and all-electron wavefunctions, respectively. The core radii r_c is a parameter to be chosen appropriately. Immediately from requirements (1) and (2), pseudopotentials agree with the all-electron potentials outside r_c because the potentials can be determined uniquely for a given wavefunction and eigenvalue. The requirement (3), the wavefunction and its derivative are continuous at r_c, results in the smooth potential. The requirement (4) is the essence of the NCPP. First, through the Gauss theorem, the norm-conservation condition guarantees that the Hartree potentials outside r_c are identical for pseudo and all-electron densities. Furthermore, the norm-conservation condition guarantees that the first energy derivative of the logarithmic derivative of pseudo and all-electron wavefunctions agree at r_c,

$$\frac{1}{2}\left[(r\phi_l^{ps}(r;\varepsilon))^2 \frac{d}{d\varepsilon}\frac{d}{dr}\ln\phi_l^{ps}(r;\varepsilon)\right]_{r=r_c} = \frac{1}{2}\left[(r\phi_l^{ae}(r;\varepsilon))^2 \frac{d}{d\varepsilon}\frac{d}{dr}\ln\phi_l^{ae}(r;\varepsilon)\right]_{r=r_c} = Q_c. \qquad (2.46)$$

In general, the scattering properties for the atomic potential are determined by the logarithmic derivative of the wavefunction, and the logarithmic derivative are monotonically increasing function. Thus, if the logarithmic derivative and its first energy derivative for pseudo wavefunction coincide with those for all-electron wavefunction at the reference energy, it is expected that pseudopotentials give the satisfactory results in different environments around the reference energy. Figure 2.2 shows the pseudo and all-electron valence wavefunctions and ionic pseudopotentials for aluminum generated with Troullier-Martins scheme [24] using OPIUM code [26].
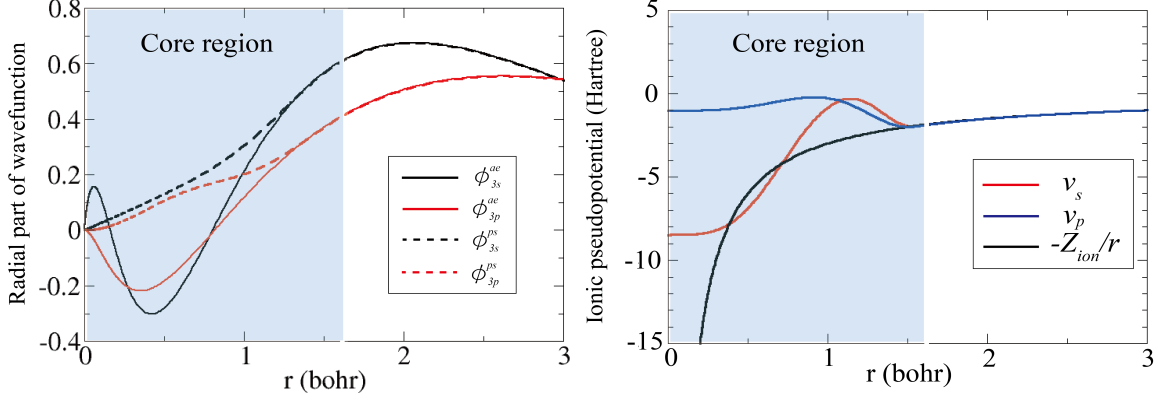
Figure 2.2: The pseudo and all-electron valence wavefunctions and ionic pseudopotentials for aluminum treated in LDA. Pseudofunctions are generated with Troullier-Martins scheme using OPIUM code. I set $r_c = 1.6$ bohr.

2.5.2 Kleinman-Bylander separable form

After removal of the valence contribution to the pseudopotential, the ionic pseudopotential $v_l(r)$ is usually separated into the local (l-independent) and non-local parts,

$$v_l(r) = v_{loc}(r) + \delta v_l(r). \tag{2.47}$$

Based on the discussion in the previous section, $v_{loc}(r) = -\frac{Z_{ion}}{r}$ and $\delta v_l(r) = 0$ for $r > r_c$, where Z_{ion} is the charge of nuclei minus inner-core electron. Because $\delta v_l(r)$ depends on the angular momentum l, the pseudopotentials act on the wavefunction as a semi-local operator:

$$\hat{v}^{ps} = v_{loc}(r) + \sum_{l,m} |Y_{lm}\rangle \, \delta v_l(r) \, \langle Y_{lm}| , \tag{2.48}$$

where Y_{lm} is the spherical harmonics. However, semi-local form is inefficient from the computational point of view. The calculation of $\hat{v}^{ps}\psi_j(\mathbf{r})$ is

$$\hat{v}^{ps}\psi_j(\mathbf{r}) = v_{loc}(r)\psi_j(\mathbf{r}) + \sum_{l,m} Y_{l,m}(\theta,\phi) \int Y_{lm}(\theta',\phi')\psi_j(r',\theta',\phi')d(\cos\theta')d\phi'. \tag{2.49}$$

In the DFT calculation, we frequently compute

$$\langle \psi_i | \, \hat{v}^{ps} \, | \psi_j \rangle = \int \psi_i(r,\theta,\phi) \Big[\hat{v}^{ps}\psi_j(\mathbf{r}) \Big]_{r,\theta,\phi} d\mathbf{r}. \tag{2.50}$$

The above operation requires that $O(N_{basis}^2)$ complexity for both computational cost and memory consumptions, where N_{basis} is the number of (real-space) basis surrounding each atom.

Kleinman and Bylander (KB) [27] proposed the efficacious form for $\hat{v}^{ps}$:

$$
\begin{aligned}
\hat{v}^{ps} &\approx v_{loc}(r) + \hat{v}_{KB} \\
&= v_{loc}(r) + \sum_{l,m} \frac{|\delta v_l \phi_{lm}^{ps}\rangle \langle \phi_{lm}^{ps} \delta v_l|}{\langle \phi_{lm}^{ps}| \delta v_l |\phi_{lm}^{ps}\rangle} \\
&= v_{loc}(r) + \sum_{l,m} |\beta_{lm}\rangle D_l \langle \beta_{lm}|,
\end{aligned}
\tag{2.51}
$$

where $\phi_{lm}^{ps} = \phi_l^{ps} Y_{lm}$, $|\beta_{lm}\rangle = |\delta v_l \phi_{lm}^{ps}\rangle$, and $D_l = \langle \phi_{lm}^{ps}| \delta v_l |\phi_{lm}^{ps}\rangle$. The Kleinman-Bylander form is much more efficient than the conventional semi-local form because

$$
\langle \psi_i| \hat{v}_{KB} |\psi_j\rangle = \sum_{l,m} D_l \langle \psi_i|\beta_{lm}\rangle \langle \beta_{lm}|\psi_j\rangle,
\tag{2.52}
$$

where

$$
\langle \beta_{lm}|\psi_j\rangle = \langle \psi_j|\beta_{lm}\rangle^* = \int \beta^*(\mathbf{r})\psi_j(\mathbf{r})d\mathbf{r}.
\tag{2.53}
$$

Thus, the operation in Eq. (2.52) can be executed in only $O(N_{basis})$ calculation and memory usage. Note that the Kleinman-Bylander form is approximation except for the atomic reference configuration, leading the problem of the occurrence of the unphysical eigenstates (Ghost states), but can be avoided with special care [28, 29].

Consequently, the Kohn-Sham equation for the pseudopotential is defined as below:

$$
H_{KS}^{ps}\psi_i^{ps}(\mathbf{r}) = \varepsilon_i \psi_i^{ps}(\mathbf{r}),
\tag{2.54}
$$

where

$$
H_{KS}^{ps} = -\frac{1}{2}\nabla^2 + v_{loc}(\mathbf{r}) + \hat{v}_{KB} + v_H(\mathbf{r}) + v_{xc}(\mathbf{r}).
\tag{2.55}
$$

The formulations and calculations in the current study are based on the Kohn-Sham equation (2.54) and Kohn-Sham Hamiltonian (2.55).

2.6   Comment on other issues

To perform the DFT calculation steadily, efficiently, and accurately, it is necessary to discuss about the selection of the basis sets, mixing the charge density, iterative eigensolvers, sampling the k-points in the Brillouin zone, algorithms for molecular dynamics. I also skip the recent advancements on the development for the exchange-correlation functionals, ultrasoft pseudopotentials [30] and projector-augmented wave method [31], and the theories beyond the DFT. For further details about the DFT, I refer the reader to Parr and Yang [32], and Martin [17].

Chapter 3

First-principles quantum transport approach

This chapter presents the theoretical modeling of quantum transport through nanoscale conductors where conduction electrons preserve the feature as a quantum mechanical wave, which gives rise to the following transport phenomena not observed in the macroscopic conductors: (i) quantum interference (ii) quantum tunneling (iii) quantum confinement (iv) inelastic scattering by electron-electron and electron-phonon interaction. The (i)-(iii) phenomena can be treated within the scattering theory for transport (Landauer-Büttiker formalism [33–35]) combined with the single-particle equations based on the DFT, Hartree-Fock theory, empirical pseudopotential method, tight-binding approximation, and effective mass approximation. For the fourth phenomena such as Coulomb blockade, electron-electron and electron-phonon interactions can be handled as self-energy matrices [36] based on the non-equilibrium Keldysh formalism which is often called the non-equilibrium Green's function (NEGF) method [37,38]. In this thesis, I focus on the system where inelastic scattering effects on the transport can be neglected, that is, the system is assumed to be at low temperature and low source-drain bias regime for simplicity.

Reminding that DFT is originally formulated for handling the ground-sate many-body problems, there is no guarantee that Kohn-Sham equation (2.54) is applicable to the electron transport beyond the linear response regime. In fact, the conventional LDA/GGA calculations usually underestimate the band-gap and therefore overestimate the conductance at the Fermi level. To overcome the shortcoming of the DFT for electron transport, the self-energy correction scheme using the Hedin's GW approximation has been proposed, and it has been verified that GW calculations give the qualitatively accurate conductance compared with the experimental values [39,40]. The only drawback of the GW method is that the computational cost is prohibitively expensive for all but very small systems. Thus, while the the use of DFT-based approach to the electron transport involves many restrictions, it has still been the workhorse approach for obtaining the insight of the transport properties of wide rage of materials due to its computational efficiency.

By neglecting the inelastic effects, results of Landauer-Büttiker formalism and NEGF method are

formally same since both approaches are conceptually identical. On the other hand, their implementations are considerably different each other, and therefore appropriate choices of numerical algorithms and basis sets are indispensable to perform large-scale transport calculations. To begin with this chapter, I will present the both approaches from NEGF method to wavefunction matching (WFM) method which is one of the theoretical method based on Landauer-Büttiker formalism. Later on that, I will show the relationship between NEGF and WFM methods, and their advantage and disadvantage.

3.1   Phase-coherent transport

Before going to the theoretical modeling, I shall consider why unique phenomena have been observed in nanoscale conductors at low temperature. It is well known that electrons in solids are scattered by potentials of impurities and defects frequently. These scatterings only change the direction of the motion (momentum), but the energy is preserved during the scattering. Another type of scattering, inelastic scattering such as electron-phonon interaction changes both momentum and energy of electrons. Roughly speaking, elastic and inelastic scatterings are distinguished by means of the mean free path L_{mfp}. If the size of conductor is larger than L_{mfp}, the inelastic scattering process is dominant and as a result the phase coherence of electrons is completely lost by collisions between electrons. In this regime, the motion of electrons is described by the classical Boltzmann transport equation, and the conductivity is proportional to its width and inversely proportional to its length (Ohm's law). On the other hand, if the length of the conductor is smaller than L_{mfp}, electrons go through the conductor while keeping the information of its phase, leading the experimental observations of the quantum mechanical effects such as Aharonov-Bohm effect, quantization of the conductance, and quantum Hall effect. This type of conduction is called phase-coherent transport. In metals, the relaxation time is short as 10^{-13} (10^{-11}) s at room temperature (1K), and therefore, the motion of electrons tends to be diffusive. Thus, to see the quantum interference in metal, the size of samples must be smaller than 1 μm. In semiconductor, L_{mfp} is much longer than that of metal, and quantum confinement effect is also expected due to its large fermi wavelength. Most of the theoretical works on the phase-coherent (ballistic) transport scheme is based on the Landauer-Büttiker formalism as introduced in the next section.

3.1.1   Two-probe system

Most of the theoretical works for phase-coherent transport have been based on the Landauer picture as generalized by Buttiker for multi-probe system. This approach has been successful in explaining many experimental observation qualitatively. I briefly explain the concept of the Landaur picture using the simplest two-probe system as depicted in Fig. 3.1. Consider the conductor sandwiched by two ideal leads that are connected with reservoirs whose chemical potentials are μ_L and μ_R for left and right reservoirs, respectively. Suppose that the motion of electrons is one-dimensional and its lowest energy
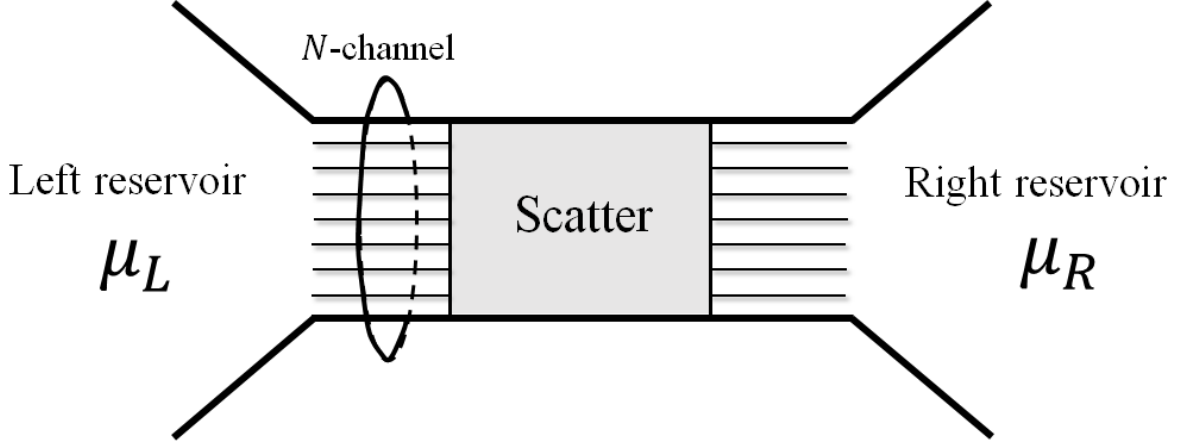
Figure 3.1: Illustration of one-dimensional narrow conductor to derive the Landauer formula. Conductor is connected to electron reservoirs with different chemical potentials μ_L and μ_R.

level is given by $\varepsilon = \hbar^2 k_z^2 / 2m^*$. The density of states of one-dimensional system is

$$
\begin{aligned}
D_{1D}(\varepsilon) &= \frac{dn}{d\varepsilon} \\
&= \frac{1}{\pi}\sqrt{\frac{m}{2\hbar^2\varepsilon}} \\
&= \frac{1}{\pi\hbar v_z},
\end{aligned}
\tag{3.1}
$$

where group velocity $v_z = \hbar k_z / m^*$. Then, the current through two-prob system around zero temperature is

$$
\begin{aligned}
I &= (-e)\int_{\infty}^{\infty} v_z(\varepsilon)D_{1D}(\varepsilon)T(\varepsilon)[f(\varepsilon-\mu_L) - f(\varepsilon-\mu_R)]d\varepsilon \\
&\approx -\frac{e}{\pi\hbar}\int_{\mu_R}^{\mu_L} T(\varepsilon)d\varepsilon \\
&\approx -\frac{e}{\pi\hbar}T(\varepsilon_F)(\mu_L - \mu_R),
\end{aligned}
\tag{3.2}
$$

where $f(\varepsilon - \mu)$ is the Fermi distribution function given by

$$
f(\varepsilon - \mu) = \frac{1}{1 + \exp(\frac{\varepsilon-\mu}{k_B T_e})}.
\tag{3.3}
$$

Here k_B and T_e are Boltzman constant and electron temperature, respectively. It is assumed that the current is carried by the single channel around the Fermi level ε_F at zero temperature. The transmission probability $T(\varepsilon)$ can be computed from the Kohn-Sham equation of two-probe system. Because the potential difference between left and right electrons is given by $-eV = \mu_L - \mu_R$, conductance between two electrodes is

$$
G = \frac{I}{V} = \frac{e^2}{\pi\hbar}T(\varepsilon_F).
\tag{3.4}
$$

This is the Landauer formula for the single channel. In general, sub-bands with discretely quantized energy levels perpendicular to the transport direction contribute to the transport. For case that total number of channels is N_{ch}, the Landauer formula is generalized as follows:

$$G = \frac{e^2}{\pi\hbar} \sum_{m,n}^{N_{ch}} |\tau_{mn}|^2, \tag{3.5}$$

where τ_{mn} is the probability amplitude in which the n-th channel of the left electrode transfers to the m-th channel of the right electrode. In this way, a procedure to investigate the transport properties in a small conductor is reduced to calculating the transmission probability from the Kohn-Sham equation of two-probe system and determining the conductivity by Landauer formula.

3.1.2   Hamiltonian in two-probe systems

The standard approach of electronic structure calculations for isolated and periodic systems is transforming the Kohn-Sham equation into the matrix eigenvalue problem. This transformation is achieved by expanding the wavefunction as a linear combination of the basis set or directly mapping on real-space grids. In this case, the size of Hamiltonian matrix becomes finite and therefore one can easily solve the Kohn-Sham equation and investigate the electronic structure in large-scale systems. In contrast, the two-probe system for transport calculation is neither isolated nor periodic system due to the external bias between the semi-infinite electrodes, leading the size of Hamiltonian matrix in two-probe system becomes infinite. It is numerically inaccessible to deal with the infinite but not periodic system. Thus, I first show how to reduce the infinite size of the Hamiltonian matrix in two-probe system to the numerically tractable form which we can deal with.

Figure 3.2 depicts the typical two-probe system where the carbon nanotube encapsulating the fullerene (peapod) embedded by the semi-infinite metallic carbon nanotubes. The peapod and semi-infinite carbon nanotubes can be regarded as a scatter and electrodes, respectively. The lower panel shows the plane-averaged electrostatic potential along the transport direction at zero bias. It can be seen that the influence of the central region is screened within several angstrom and electrostatic potential smoothly connects to the bulk potential. Thus, the whole system is divided into three part: left (L) electrode, central region (C), and right (R) electrode, and it is assumed that the effective local potential and scattering states connect to the bulk potential and Bloch states inside the electrodes. Note that the Hamiltonian matrix of electrode region is obtained by the self-consistent calculation for the bulk system. The system can be further divided into the principal layers perpendicular to the transport direction by introducing the basis set such as real-space grids, localized atomic orbitals, or Laue representation. Here the principal layers are defined so as to interact only the nearest neighboring layers. Based on the above system setup, what we need to consider is the finite-dimensional Hamiltonian matrix in the "transition" region (T) which is the central region plus two adjacent principal layers for left and right electrodes as illustrated in Fig. 3.2.

Due to the locality of the real-space representation of the Hamiltonian matrix, the infinite Hamil-
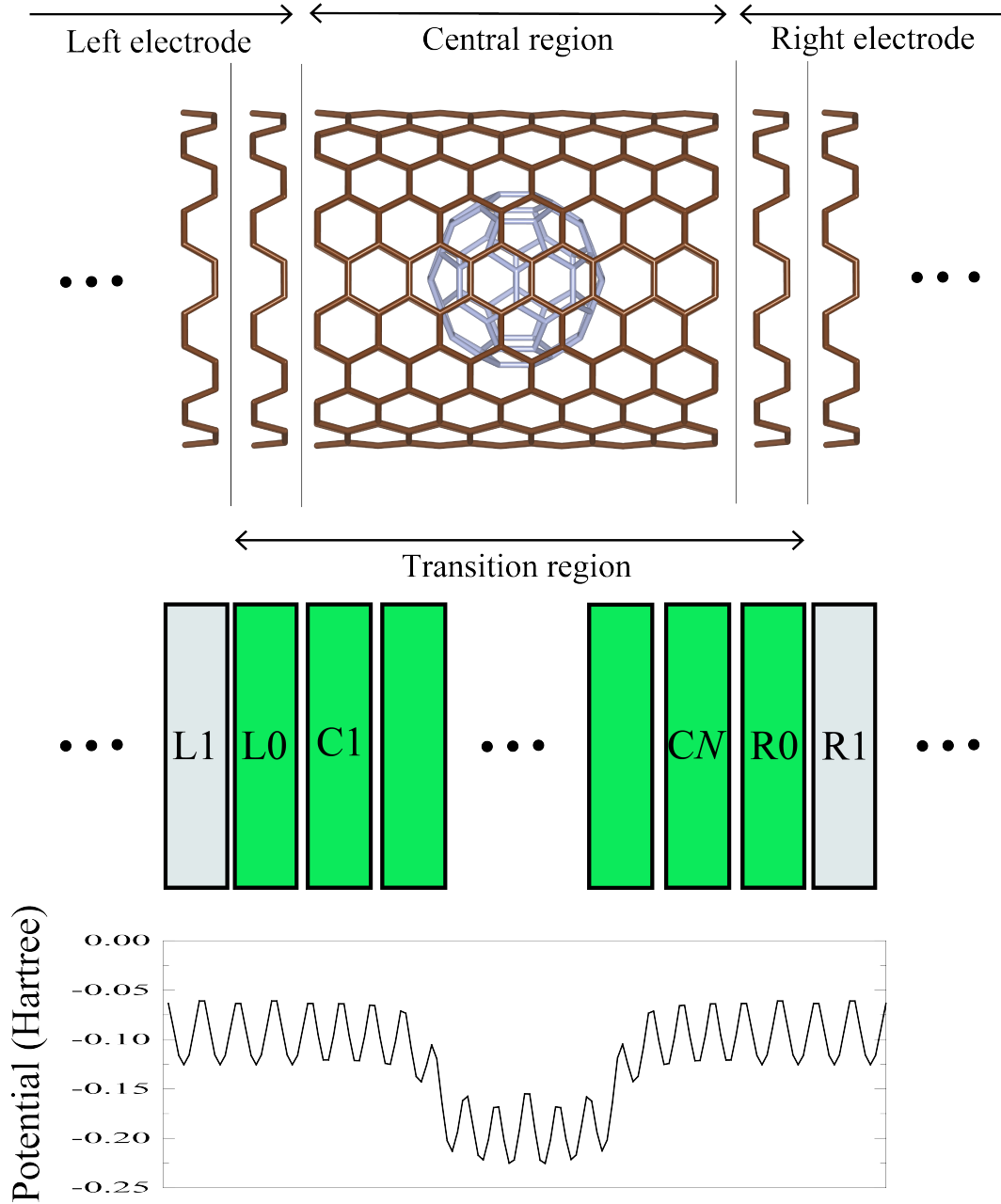
Figure 3.2: Schematic representation of a quasi-one-dimensional conductor sandwiched by two semi-infinite electrodes. As an example, atomic structure of $(10, 10)$carbon nanotube (CNT) with single C60 is illustrated in the upper panel. Because of the localized feature of the real-space grid approach, the system can be divided into principal layers denoted by $\cdots$,L1,L0,C1,$\cdots$,CN,R0,R1,$\cdots$. Transition region incorporate L0 and R0 layers as wavefunction matching planes. The lower panel shows the plane-averaged electrostatic potential along the transport direction at zero bias. It can be seen that the influence of the central region is screened within several angstrom and electrostatic potential smoothly connects to the bulk potential due to the screening effect of metal.

tonian is partitioned into the following matrix form:

$$
\hat{H} = \begin{bmatrix} \hat{H}_L & \hat{H}_{LT} & 0 \\ \hat{H}_{TL} & \hat{H}_T & \hat{H}_{TR} \\ 0 & \hat{H}_{RT} & \hat{H}_R \end{bmatrix},
\tag{3.6}
$$

where $\hat{H}_L$, $\hat{H}_T$ and $\hat{H}_R$ are the Hamiltonian matrices for the left electrode, transition region and right electrode, respectively. $\hat{H}_{TL}$ and $\hat{H}_{LT}$ ($\hat{H}_{TR}$ and $\hat{H}_{RT}$) are the coupling matrices between the central region and the left (right) electrode. Note that $\hat{H}_{TL} = \hat{H}_{LT}^{\dagger}$ and $\hat{H}_{TR} = \hat{H}_{RT}^{\dagger}$. The transition region should be taken large enough in order to eliminate the coupling between the left and right electrodes. The representation of Eq. (3.6) is also valid for periodic in the direction parallel to the electron transport. In this case, translation of the transverse direction can be characterized in terms of the Bloch vector $\boldsymbol{k}_{\parallel}$ in the two-dimensional Brillouin zone, which is still a good quantum number in the transport calculation.

3.2 Green's function method: GF

At present, the non-equilibrium Green's function (NEGF) method combined with Kohn-Sham Hamiltonian in DFT is the de-facto standard approach to calculate the electronic structure and transmission probability in two-probe systems. The advantages of NEGF method is summarized as the following reasons: (i) the non-equilibrium Green's function is compactly described by the localized basis set, (ii) the contribution of bound states to the charge density can be included by the contour integral of the retarded Green's function, (iii) the inelastic effects on the transport properties such as electron-electron and electron-phonon interactions can be included as self-energy matrices. Owing to these advantages, the NEGF method has been implemented in several DFT codes with the localized basis set [41–44], and there are many successful applications for transport properties.

3.2.1 Basic formalism

The Kohn-Sham equation (2.54) for the two-probe system can be written as a matrix equation,

$$
\begin{bmatrix} \hat{H}_L & \hat{H}_{LT} & 0 \\ \hat{H}_{TL} & \hat{H}_T & \hat{H}_{TR} \\ 0 & \hat{H}_{RT} & \hat{H}_R \end{bmatrix} \begin{bmatrix} \boldsymbol{\psi}_L \\ \boldsymbol{\psi}_T \\ \boldsymbol{\psi}_R \end{bmatrix} = \varepsilon \begin{bmatrix} \boldsymbol{\psi}_L \\ \boldsymbol{\psi}_T \\ \boldsymbol{\psi}_R \end{bmatrix}.
\tag{3.7}
$$

From the first and third equations, the wavefunction in electrodes and transition region can be coupled with as below:

$$
\boldsymbol{\psi}_L = \hat{g}_L(\varepsilon)\hat{H}_{LT}\boldsymbol{\psi}_T, \ \ \boldsymbol{\psi}_R = \hat{g}_R(\varepsilon)\hat{H}_{RT}\boldsymbol{\psi}_T,
\tag{3.8}
$$

where $\hat{g}_L(\varepsilon)$ and $\hat{g}_R(\varepsilon)$ are surface Green's functions for left and right electrodes,

$$\hat{g}_L(\varepsilon) \;=\; [\varepsilon\hat{I} - \hat{H}_L]^{-1}, \tag{3.9}$$

$$\hat{g}_R(\varepsilon) \;=\; [\varepsilon\hat{I} - \hat{H}_R]^{-1}. \tag{3.10}$$

Using Eq. (3.8), the wavefunction in the transition region becomes

$$
\begin{aligned}
[\varepsilon\hat{I} - \hat{H}_T]\boldsymbol{\psi}_T &\;=\; \hat{H}_{TL}\boldsymbol{\psi}_L + \hat{H}_{TR}\boldsymbol{\psi}_R \\
&\;=\; [\hat{\Sigma}_L(\varepsilon) + \hat{\Sigma}_R(\varepsilon)]\boldsymbol{\psi}_T,
\end{aligned} \tag{3.11}
$$

where $\hat{\Sigma}_{L,R}(\varepsilon)$ are self-energy matrices for left and right electrodes defined as

$$\hat{\Sigma}_L(\varepsilon) = \hat{H}_{TL}\hat{g}_L(\varepsilon)\hat{H}_{LT}, \tag{3.12}$$

$$\hat{\Sigma}_R(\varepsilon) = \hat{H}_{TR}\hat{g}_R(\varepsilon)\hat{H}_{RT}. \tag{3.13}$$

Equation (3.11) can be rewritten as

$$[\varepsilon\hat{I} - H_T - \hat{\Sigma}_L(\varepsilon) - \hat{\Sigma}_R(\varepsilon)]\boldsymbol{\psi}_T = 0. \tag{3.14}$$

The retarded Green's function of the transition region [1] can be defined as a resolvent of the coefficient matrix of Eq. (3.14), that is,

$$\hat{G}_T(\varepsilon) = [(\varepsilon + i\eta)\hat{I} - \hat{H}_T - \hat{\Sigma}_L(\varepsilon) - \hat{\Sigma}_R(\varepsilon)]^{-1}, \tag{3.15}$$

where η is the positive infinitesimal. It is worth noting that the size of the matrices in Eq. (3.15) is finite in contrast to the infinite size of the whole Hamiltonian of Eq. (3.6), which is numerically tractable. Thus, the matrix inversion procedure can be executed once the self-energy matrices for electrodes are obtained.

In the NEGF method. the charge density of the transition region is calculated using the relationship between density matrix and Green's function,

$$
\begin{aligned}
n_T(\mathbf{r}) &\;=\; \mathrm{Tr}\hat{D} \\
&\;=\; \frac{1}{2\pi i}\int_{-\infty}^{\infty} G_T^{<}(\varepsilon; \mathbf{r}, \mathbf{r})d\varepsilon,
\end{aligned} \tag{3.16}
$$

where $\hat{D}$ is the density matrix and $\hat{G}_T^{<}(\varepsilon)$ is the lesser part of the non-equilibrium Green's function. Assuming that the edge of the transition region, i.e., L0 and R0 layers are close to the equilibrium,

---

[1] We can also derive Eq. (3.15) from the identity

$$
A = \begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{bmatrix}, \quad A_{22}^{-1} = \frac{a_{11}a_{33}}{a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{33}a_{12}a_{21}}.
$$

one might obtain

$$\begin{aligned}
\hat{G}_T^<(\varepsilon) &= \hat{G}_T(\varepsilon)\hat{\Sigma}_T^<(\varepsilon)\hat{G}_T^\dagger(\varepsilon) \\
&\simeq \hat{G}_T(\varepsilon)[if(\varepsilon - \mu_L)\hat{\Gamma}_L(\varepsilon) + if(\varepsilon - \mu_R)\hat{\Gamma}_R(\varepsilon)]\hat{G}_T^\dagger(\varepsilon).
\end{aligned} \tag{3.17}$$

Here, I introduced the matrices $\hat{\Gamma}_{L,R}(\varepsilon)$ that are called coupling constants and defined by the imaginary part of the self-energy matrices, that is,

$$\hat{\Gamma}_L(\varepsilon) = i[\hat{\Sigma}_L(\varepsilon) - \hat{\Sigma}_L^\dagger(\varepsilon)], \ \hat{\Gamma}_R(\varepsilon) = i[\hat{\Sigma}_R(\varepsilon) - \hat{\Sigma}_R^\dagger(\varepsilon)]. \tag{3.18}$$

Equation (3.17) can be divided into equilibrium and non-equilibrium parts as below:

$$\hat{G}_T^<(\varepsilon) = i\hat{G}(\varepsilon)_T[\hat{\Gamma}_L(\varepsilon) + \hat{\Gamma}_R(\varepsilon)]\hat{G}_T^\dagger(\varepsilon)f(\varepsilon - \mu_R) + i\hat{G}_T(\varepsilon)\hat{\Gamma}_L(\varepsilon)\hat{G}_T^\dagger(\varepsilon)[f(\varepsilon - \mu_L) - f(\varepsilon - \mu_R)]. \tag{3.19}$$

If the Hamiltonian matrix is independent from $\boldsymbol{k}_\parallel$, the equilibrium part can be simplified into a well-known formula,

$$i\hat{G}_T(\varepsilon)[\hat{\Gamma}_L(\varepsilon) + \hat{\Gamma}_R(\varepsilon)]\hat{G}_T^\dagger(\varepsilon)f(\varepsilon - \mu_R) = -2if(\varepsilon - \mu_R)\text{Im}\hat{G}_T(\varepsilon). \tag{3.20}$$

Substituting Eqs. (3.19) and (3.20) into Eq. (3.16) leads to

$$n_T(\mathbf{r}) = n_T^{\text{eq}}(\mathbf{r}) + n_T^{\text{neq}}(\mathbf{r}), \tag{3.21}$$

where

$$\begin{aligned}
n_T^{\text{eq}}(\mathbf{r}) &= -\frac{1}{\pi}\int_{-\infty}^{\infty} d\varepsilon\,\text{Im}G_T(\varepsilon; \mathbf{r}, \mathbf{r})f(\varepsilon - \mu_R), \tag{3.22} \\
n_T^{\text{neq}}(\mathbf{r}) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\varepsilon\,\text{Tr}[\hat{G}_T(\varepsilon)\hat{\Gamma}_L(\varepsilon)\hat{G}_T^\dagger(\varepsilon)][f(\varepsilon - \mu_L) - f(\varepsilon - \mu_R)]. \tag{3.23}
\end{aligned}$$

Here, I assume that $\mu_L \geq \mu_R$. The first term, the equilibrium charge density, is evaluated from the retarded Green's function which is an analytic function for $\text{Im}(z) > 0$ on the complex plane. In practice, the integration of the retarded Green's function along the real axis is replaced by the contour integration on the upper half complex plane, which makes it efficient to evaluate the equilibrium charge density because the retarded Green's function on complex plane is very smooth. On the other hand, the second term, non-equilibrium correction term for the charge density, should be evaluated at real axis because the integrant is not analytic apart from the real axis and thus one cannot apply the counter integral approach. As same as the standard DFT calculation for isolated or periodic system, one can construct the effective local potential $v_{eff}(\mathbf{r})$ in the transition region from the charge density $n_T(\mathbf{r})$.

The transport properties of two-probe system are evaluated by the Landauer formula, for example, the current is given by following formula,

$$I = \frac{e}{\pi\hbar}\int_{-\infty}^{\infty} T(\varepsilon)[f(\varepsilon - \mu_L) - f(\varepsilon - \mu_R)]d\varepsilon, \tag{3.24}$$

where $T(\varepsilon)$ is computed by the Fisher-Lee formula [8]:

$$T(\varepsilon) = \text{Tr}[\hat{G}_T(\varepsilon)\hat{\Gamma}_L(\varepsilon)(\hat{G}_T(\varepsilon))^\dagger\hat{\Gamma}_R(\varepsilon)], \qquad (3.25)$$

which means that the transmission probability is determined by the matrices $\hat{\Sigma}_L$, $\hat{\Sigma}_R$, and $\hat{G}_T$.

As mentioned in the sections above, it is possible to include the effect of the electron-electron and electron-phonon interaction into the transport properties, which is also one of the merits of the NEGF method. For further details on the NEGF method, I refer the reader Ref. [8] and Ref [45].

3.2.2   Self-energy matrices

To determine the self-energy matrices, it is needed to compute the surface Green's functions defined by Eqs. (3.9) and (3.10). However, as mentioned in the sections above, the Hamiltonian matrices of electrodes are semi-infinite, resulting in that the matrix inversion procedure is infeasible. Thus, I introduce several numerical techniques to calculate the surface Green's functions. Considering that the Hamiltonian matrices of electrodes are periodic and the couplings exist only between the nearest neighbor unit cells, which makes it possible to write the Hamiltonian matrices of electrodes as the following block tri-diagonal form:

$$\hat{H}_L = \begin{bmatrix} \ddots & \ddots & & 0 \\ \ddots & \hat{A} & \hat{B} & \\ & \hat{B}^\dagger & \hat{A} & \hat{B} \\ 0 & & \hat{B}^\dagger & \hat{A} \end{bmatrix}, \quad \hat{H}_R = \begin{bmatrix} \hat{A} & \hat{B} & & 0 \\ \hat{B}^\dagger & \hat{A} & \hat{B} & \\ & \hat{B}^\dagger & \hat{A} & \ddots \\ 0 & & \ddots & \ddots \end{bmatrix}, \qquad (3.26)$$

where $\hat{A}$ and $\hat{B}$ correspond to the on-site and hopping Hamiltonian matrices in the unit cell, respectively. By construction, these matrices are same for each unit cell. I here assume the left electrode is identical with the right electrode to keep the notation as simple as possible, but the extension to the hetero electrodes is straightforward. Because the transition region is defined so as to include the outermost unit cells L0 and R0, the self-energy matrices is simplified into a more compact form, for example, $\hat{\Sigma}_L(\varepsilon)$ can be written as below,

$$\begin{aligned} \hat{\Sigma}_L(\varepsilon) &= \hat{H}_{TL}\hat{g}_L(\varepsilon)\hat{H}_{LT} \\ &= \begin{bmatrix} \cdots & 0 & \hat{B}^\dagger \\ & 0 & 0 \\ & & \\ \cdot^{\cdot^\cdot} & & \vdots \end{bmatrix} \begin{bmatrix} \ddots & & & \vdots \\ & \hat{g}_{22}^L & \hat{g}_{21}^L & \hat{g}_{20}^L \\ & \hat{g}_{12}^L & \hat{g}_{11}^L & \hat{g}_{10}^L \\ \cdots & \hat{g}_{02}^L & \hat{g}_{01}^L & \hat{g}_{00}^L \end{bmatrix} \begin{bmatrix} \vdots & & \cdot^{\cdot^\cdot} \\ & & \\ 0 & 0 & \\ \hat{B} & 0 & \cdots \end{bmatrix} \\ &= \hat{B}^\dagger \hat{g}_{00}^L \hat{B}. \qquad (3.27) \end{aligned}$$

Similar expression for $\hat{\Sigma}_R(\varepsilon)$ can be obtained by

$$
\begin{aligned}
\hat{\Sigma}_R(\varepsilon) &= \hat{H}_{TR}\hat{g}_R(\varepsilon)\hat{H}_{RT} \\
&= \begin{bmatrix} \vdots & & \iddots \\ & & \\ 0 & 0 & \\ \hat{B} & 0 & \cdots \end{bmatrix} \begin{bmatrix} \hat{g}^R_{00} & \hat{g}^R_{01} & \hat{g}^R_{02} & \cdots \\ \hat{g}^R_{10} & \hat{g}^R_{11} & \hat{g}^R_{12} & \\ \hat{g}^R_{20} & \hat{g}^R_{21} & \hat{g}^R_{22} & \\ \vdots & & & \ddots \end{bmatrix} \begin{bmatrix} \cdots & 0 & \hat{B}^\dagger \\ & 0 & 0 \\ & & \\ \iddots & & \vdots \end{bmatrix} \\
&= \hat{B}\hat{g}^R_{00}\hat{B}^\dagger,
\end{aligned}
\tag{3.28}
$$

where $\hat{g}^{L/R}_{00}$ is the surface Green's function of the outermost unit cell, i.e., L0 or R0 layer in Fig. 3.2.

Recursive Green's function method

The most straightforward way for calculating the outermost surface Green's function $\hat{g}^{L/R}_{00}$ is the recursive Green's function (RGF) method proposed by Haydock *et al.* [46,47]. In the following, I will drop the subscript L/R for the sake of the simplicity and focus on the calculation of the surface Green's function for right electrode because the left surface Green's function can be obtained by the analogous manner. Using the identity for the matrix inversion

$$
\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{A-BD^{-1}C} & -A^{-1}B\frac{1}{D-CA^{-1}B} \\ -D^{-1}C\frac{1}{A-BD^{-1}C} & \frac{1}{D-CA^{-1}B} \end{bmatrix},
\tag{3.29}
$$

and putting $A = \varepsilon\hat{I} - \hat{A}, B = -\hat{H}_{RC}, C = -\hat{H}_{CR}, D = \varepsilon\hat{I} - \hat{H}_R$, one can derive the continued fraction expression for $\hat{g}_{00}$:

$$
\hat{g}^{(n+1)}_{00}(\varepsilon) = [\varepsilon\hat{I} - \hat{A} - \hat{B}\hat{g}^{(n)}_{00}(\varepsilon)\hat{B}^\dagger]^{-1},
\tag{3.30}
$$

or equivalently,

$$
\hat{\Sigma}^{(n+1)}_L(\varepsilon) = \hat{B}[\varepsilon\hat{I} - \hat{A} - \hat{\Sigma}^{(n)}_L(\varepsilon)]^{-1}\hat{B}^\dagger,
\tag{3.31}
$$

where index n is the number of iteration. Note that one needs to add infinitesimal η to the energy, that is, $\varepsilon \to \varepsilon + i\eta$ in order to avoid the singularities of the surface Green's function. It is evident that the procedure of the $(n+1)$-th iteration means that adding the single periodic layer to the isolated n slices, and the exact surface Green's function is obtained when $n \to \infty$. It is well known that the RGF method requires many iterations until the self-consistency is achieved because the only n layers are taken into account after n iteration. While the RGF method is inefficient for the surface Green's function calculation, it is frequently used to evaluate the retarded and lesser Green's functions in $O(N)$ operation [48].

Quick decimation method

To overcome the slow convergency of the RGF method, Lopez Sancho *et al.* [49, 50] proposed the alternative iterative method which takes into 2^n layer after n iteration. I call this method as quick decimation method in this thesis. Owing to its numerical efficiency and stability, the quick iterative method is one of the most popular methods to evaluate the surface Green's function.

Since the surface Green's function for right electrode satisfies

$$[\varepsilon\hat{I} - \hat{H}_R]\hat{g}_R = \hat{I}, \tag{3.32}$$

matrix elements obey that

$$[\varepsilon\hat{I} - \hat{A}]\hat{g}_{00} - \hat{B}\hat{g}_{10} = \hat{I}, \tag{3.33}$$

$$-\hat{B}^\dagger\hat{g}_{n-1,0} + [\varepsilon\hat{I} - \hat{A}]\hat{g}_{n0} - \hat{B}\hat{g}_{n+1,0} = 0, \tag{3.34}$$

for $n \geq 1$. Equation (3.34) can be written as

$$\hat{g}_{n0} = \hat{t}_0\hat{g}_{n-1,0} + \tilde{t}_0\hat{g}_{n+1,0}, \tag{3.35}$$

where

$$\hat{t}_0 = [\varepsilon\hat{I} - \hat{A}]^{-1}\hat{B}^\dagger, \tag{3.36}$$

$$\tilde{t}_0 = [\varepsilon\hat{I} - \hat{A}]^{-1}\hat{B}. \tag{3.37}$$

Using the Eqs. (3.34)-(3.37) iteratively, Eq. (3.35) can be generalized as

$$\hat{g}_{2^n,0} = \hat{t}_n\hat{g}_{00} + \tilde{t}_n\hat{g}_{2^{n+1},0}, \tag{3.38}$$

$$\hat{t}_n = [\hat{I} - \hat{t}_{n-1}\tilde{t}_{n-1} - \tilde{t}_{n-1}\hat{t}_{n-1}]^{-1}\hat{t}_{n-1}^2, \tag{3.39}$$

$$\tilde{t}_n = [\hat{I} - \hat{t}_{n-1}\tilde{t}_{n-1} - \tilde{t}_{n-1}\hat{t}_{n-1}]^{-1}\tilde{t}_{n-1}^2. \tag{3.40}$$

Thus, one might obtain

$$\begin{aligned}
\hat{g}_{10} &= \hat{t}_0\hat{g}_{00} + \tilde{t}_0\hat{g}_{20} \\
&= \hat{t}_0\hat{g}_{00} + \tilde{t}_0(\hat{t}_1\hat{g}_{00} + \tilde{t}_1\hat{g}_{40}) \\
&= \cdots \\
&= (\hat{t}_0 + \tilde{t}_0\hat{t}_1 + \tilde{t}_0\tilde{t}_1\hat{t}_2 + \cdots + \tilde{t}_0\tilde{t}_1...\tilde{t}_n\hat{t}_n)\hat{g}_{00} + \tilde{t}_0\tilde{t}_1...\tilde{t}_n\hat{g}_{2^{n+1},0}. \tag{3.41}
\end{aligned}$$

If $|\tilde{t}_0\tilde{t}_1...\tilde{t}_n|$ is negligibly small,

$$\hat{g}_{10} = \hat{T}_n\hat{g}_{00}, \tag{3.42}$$

where $\hat{T}_n = \hat{t}_0 + \tilde{t}_0\hat{t}_1 + \tilde{t}_0\tilde{t}_1\hat{t}_2 + \cdots + \tilde{t}_0\tilde{t}_1...\tilde{t}_n\hat{t}_n$. Substituting Eq. (3.42) into Eq. (3.33), the surface Green function is obtained

$$\hat{g}_{00} = [\varepsilon\hat{I} - \hat{A} - \hat{B}^\dagger\hat{T}_n]^{-1}. \tag{3.43}$$

Note that it is necessary to add infinitesimal η to the energy, that is, $\varepsilon \to \varepsilon + i\eta$ in order to avoid the singularities of the surface Green's function. As mentioned above, the quick decimation method calculates the surface Green's function of the finite 2^n layers after n iteration. Thus, the number of the iteration required to converge is decreased significantly. Because the convergency depends on the closeness of the singularity of the surface Green's function, too small η gives rise to increase the number of iteration, on the other hand, too large η gives a result far away from the exact value which should be calculated when $\eta = 0$.

Semi-analytical method

The alternative approaches have been proposed by several groups, for example, transfer-matrix method [51], matrix Möbius transformation method [52, 53], semi-analytical method [54, 55]. All of these approaches use the solutions of the Kohn-Sham equation for bulk system to construct the surface Green's function. According to the pioneering work by Sanvito and his coworkers [54], who first applied this approach to the electron transport, I call these approaches as semi-analytical method in this thesis. The advantages of the semi-analytical method compared to the iterative methods are summarized as follows: (i) exact (semi-infinite) surface Green's function can be constructed analytically from the generalized Bloch states, (ii) computation of the surface Green's function is one-shot (non-self-consistent procedure), (iii) it is not required to add the infinitesimal η to the energy and the computational cost does not depend on the choice of η. On the other hand, one needs to solve the eigenvalue problem typically with ill condition instead of calculating the matrix inversion. However, it has been reported that the computational cost of the semi-analytical approach is faster than the quick decimation technique [48]. Because the semi-analytical method is closely related with the WFM method, I introduce the details of this method in the latter section.

Alternatives to self-energy matrices

The computation of the self-energy matrices must be done independently at each energy point and takes a large amount of time. To reduce this computational burden, several approximate expressions for self-energy matrices have been proposed. The simplest one might be the jellium model for metal electrode which approximates the spatially varying electron density by a uniformly distributed electron density. This model is often used with the Lippmann-Schwinger scattering method as an unperturbed system. Because the solutions of the Kohn-Sham equation for jellium model are plane waves, the self-energy matrices are constructed analytically. However, it is not difficult to imagine that the jellium model fails to describe the atomic contact between electrodes and central region, which leads an artificial scattering at the interface. In addition, it may give the poor description for the electronic structure of $3d$-transition metals where valence electrons are spatially localized. On the other hand, in the field of quantum computational chemistry, the wide-band limit (WBL) approximation [56] is the commonly used approach especially for the gold metal. The WBL approximation assumes that the real part and energy dependency of the of the self-energy matrices can be neglected, that is, $\hat{\Sigma}(\varepsilon) \approx i\hat{\Gamma}(\varepsilon_F)$.

Moreover, the imaginary part of the surface Green's function is approximated by the density of states of bulk electrode. This approximation might give reasonable results when the interaction between electrodes and central region is weak and density of states is nearly constant around the Fermi energy. The advantage of the WBL approximation is that the self-energy matrices are independent of the energy and determined by the self-consistent calculation for the bulk system, realizing the dramatic reduction of the computational cost. Recently, the complex adsorbing potential (CAP) method [57–59] has been developed to compute the Green's function and the self-energy matrices. The CAP method mimics the self-energy matrices via imaginary analytical functions such as Gaussian function. However, the use of the CAP requires several parameters to be tuned manually to remove spurious reflections at the left and right boundaries; this may restrict its applicability to complicated electrode materials. In addition, the CAP method can only be used for the linear response regime, that is, for non-self-consistent calculations.

3.3 Wavefunction-matching method: WFM

Transport properties in nanoscale conductor can be calculated directly from the scattering states that are the solutions of the Kohn-Sham equation under the open boundary condition. There have been proposed many numerical approaches to evaluate the scattering states, including the Lippmann-Schwinger scattering method [60–62], recursive transfer-matrix method [63, 64], and wavefunction-matching (WFM) method [65–67]. However, due to the restriction of the jellium approximation, applications of the Lippmann-Schwinger scattering method and recursive transfer-matrix method have been limited only for the systems connected with the structureless electrodes, e.g., molecular junctions or atomic chains suspended by the jellium electrodes. On the other hand, the WFM method can deal with the electrodes and central region on the equal footing. In this section, I here focus on the theoretical framework of the WFM method.

3.3.1 Basic concept

Again, let us consider the scattering problem of the two-probe system. From Eq. (2.54), the Kohn-Sham equation in the transition region can be written as below:

$$-\hat{B}^\dagger \boldsymbol{\psi}_{L1} + [\varepsilon \hat{I} - \hat{H}_T]\boldsymbol{\psi}_T - \hat{B}\boldsymbol{\psi}_{R1} = 0, \tag{3.44}$$

As already mentioned, I assume that the scattering state inside the electrodes are expressed by the linear combination of the solutions of the Kohn-Sham equation for the Hamiltonian of electrodes. If there is no incoming electron from the right electrode, scattering states will fulfill the following boundary matching condition:

$$\boldsymbol{\psi}_{Lk} = \boldsymbol{\phi}_{Lk,i}^+ + \sum_j r_{i,j} \boldsymbol{\phi}_{Lk,j}^-, \tag{3.45}$$

$$\boldsymbol{\psi}_{Rk} = \sum_j t_{i,j} \boldsymbol{\phi}_{Rk,j}^+, \tag{3.46}$$

for $k \geq 0$, where $\phi_{Lk,j}^{+}$ and $\phi_{Rk,j}^{-}$ represent the j-th right-going in Lk layer and left-going states in Rk layer, respectively. Right-going states are either evanescent waves that are decaying to the right direction or Bloch waves that are propagating to the right direction. On the other hand, left-going states are either decaying or propagating to the left direction. Here, $r_{i,j}$ and $t_{i,j}$ are reflection and transmission coefficients. In addition, the dual vectors $\{\widetilde{\phi}_j\}$ are determined so as to satisfy $(\phi_j, \widetilde{\phi}_{j'}) = \delta_{j,j'}$. In case that $\{\widetilde{\phi}_j\}$ spans the complete system in each electrode, the following identity is established

$$\sum_j \phi_{Lk,j}^{+} (\widetilde{\phi}_{Lk,j}^{+})^{\dagger} = \sum_j \phi_{Rk,j}^{-} (\widetilde{\phi}_{Rk,j}^{-})^{\dagger} = I, \tag{3.47}$$

and therefore scattering states in the left and right electrodes can be written as

$$\psi_{Lk} = \phi_{Lk,i}^{+} + \sum_j (\widetilde{\phi}_{Lk,j}^{-}, \psi_{Lk}) \phi_{Lk,j}^{-}, \tag{3.48}$$

$$\psi_{Rk} = \sum_j (\tilde{\phi}_{Rk,j}^{+}, \psi_{Rk}) \phi_{Rk,j}^{+}. \tag{3.49}$$

Comparing Eqs. (3.45) and (3.46) with Eqs. (3.48) and (3.49) yields

$$r_{i,j} = (\widetilde{\phi}_{Lk,j}^{-}, \psi_{Lk} - \phi_{Lk,i}^{+}), \tag{3.50}$$

$$t_{i,j} = (\widetilde{\phi}_{Rk,j}^{+}, \psi_{Rk}). \tag{3.51}$$

Especially when $k = 0$,

$$r_{i,j} = (\widetilde{\phi}_{L0,j}^{-}, \psi_{L0} - \phi_{L0,i}^{+}), \tag{3.52}$$

$$t_{i,j} = (\widetilde{\phi}_{R0,j}^{+}, \psi_{R0}). \tag{3.53}$$

By substituting Eq. (3.52) into Eq. (3.45), ψ_{L1} can be expressed by

$$\psi_{L1} = \Big[\sum_j \phi_{L1,j}^{-} (\widetilde{\phi}_{L0,j}^{-})^{\dagger} \Big] \psi_{L0} + \phi_{L1,i}^{+} - \Big[\sum_j \phi_{L1,j}^{-} (\widetilde{\phi}_{L0,j}^{-})^{\dagger} \Big] \phi_{L0,i}^{+}. \tag{3.54}$$

Similarly for ψ_{R1}, eliminating t_{ij} from Eq. (3.46) results in

$$\psi_{R1} = \Big[\sum_j \phi_{R1,j}^{+} (\widetilde{\phi}_{R0,j}^{+})^{\dagger} \Big] \psi_{R0}. \tag{3.55}$$

Finally substituting Eqs. (3.54) and (3.55) into Eq. (3.44) yields the closed form of the Kohn-Sham equation for the transition region,

$$[\varepsilon \hat{I} - \hat{H}_T - \hat{\Sigma}_L(\varepsilon) - \hat{\Sigma}_R(\varepsilon)] \psi_T = \Omega_i(\varepsilon), \tag{3.56}$$

where

$$\hat{\Sigma}_L = \hat{B}^{\dagger} \Big[\sum_j \phi_{L1,j}^{-} (\widetilde{\phi}_{L0,j}^{-})^{\dagger} \Big], \tag{3.57}$$

$$\hat{\Sigma}_R = \hat{B} \Big[\sum_j \phi_{R1,j}^{+} (\widetilde{\phi}_{R0,j}^{+})^{\dagger} \Big], \tag{3.58}$$

$$\Omega_i = \hat{B}^{\dagger} \phi_{L1,i}^{+} - \hat{\Sigma}_L \phi_{L0,i}^{+} \tag{3.59}$$

Note that $\hat{\Sigma}_L$ and $\hat{\Sigma}_R$ are same as that of Eqs. (3.27) and (3.28) defined by the surface Green's functions, although the relationship is not seemingly obvious. The equivalence of the self-energy matrices between the GF and WFM methods will be proved in the next section. Remembering that the transition region has N principal layers for central region and 2 additional layers L0 and R0 as wavefunction matching planes, $\varepsilon\hat{I} - \hat{H}_T$ can be written as the following matrix form:

$$
\varepsilon\hat{I} - \hat{H}_T = \begin{bmatrix} \overline{H}_{00} & \overline{H}_{01} & & 0 \\ \overline{H}_{10} & \overline{H}_{11} & \ddots & \\ & \ddots & \ddots & \overline{H}_{N,N+1} \\ 0 & & \overline{H}_{N+1,N} & \overline{H}_{N+1,N+1} \end{bmatrix}.
\tag{3.60}
$$

Thus, Kohn-Sham equation (3.44) is reduced to the $(N+2)$-dimensional linear equation

$$
\begin{bmatrix} \overline{H}_{00} - \hat{\Sigma}_L & \overline{H}_{01} & & & 0 \\ \overline{H}_{10} & \overline{H}_{11} & \hat{H}_{12} & & \\ & \ddots & \ddots & \ddots & \\ & & \overline{H}_{N,N-1} & \overline{H}_{N,N} & \overline{H}_{N,N+1} \\ 0 & & & \overline{H}_{N+1,N} & \overline{H}_{N+1,N+1} - \hat{\Sigma}_R \end{bmatrix} \begin{bmatrix} \boldsymbol{\psi}_{L0} \\ \boldsymbol{\psi}_{C1} \\ \vdots \\ \boldsymbol{\psi}_{CN} \\ \boldsymbol{\psi}_{R0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Omega}_i \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}.
\tag{3.61}
$$

The derivation of the above equation was firstly done by Ando [65] for the tight-binding approximation and later Khomyakov *et al.* [68,69] and Kong *et al.* [70,71] generalized it to the first-principles method within real-space finite-difference scheme.

After solving Eq. (3.61), the electric current is calculated by the integration of the current density $\boldsymbol{j}(\boldsymbol{r})$ over the arbitral surface region

$$
\begin{aligned}
I &= \int \boldsymbol{j}(\boldsymbol{r}) \cdot d\boldsymbol{S} \\
&= \frac{2e}{(2\pi)^3} \sum_{i,j} \int (f(\varepsilon - \mu_L) - f(\varepsilon - \mu_R)) v'_j |t_{i,j}|^2 S d\boldsymbol{k} \\
&= \sum_{i,j} \int \frac{S}{(2\pi)^2} dk_x dk_y \int_{-\infty}^{\infty} (f(\varepsilon - \mu_L) - f(\varepsilon - \mu_R)) \frac{2e}{h} \frac{v'_j}{v_i} |t_{i,j}|^2 d\varepsilon \\
&= \frac{e}{\pi\hbar} \int_{-\infty}^{\infty} (f(\varepsilon - \mu_L) - f(\varepsilon - \mu_R)) \sum_{i,j} \frac{v'_j}{v_i} |t_{i,j}|^2 d\varepsilon,
\end{aligned}
\tag{3.62}
$$

where v_i and v'_j are the group velocity of the incident i-th and transmitting j-th right-going states, respectively. As shown in appendix A, v_i is given by

$$
v_i = \frac{1}{\hbar} \frac{\partial E}{\partial k_i^+} = \frac{2a}{\hbar} \mathrm{Im}[(\lambda_i^+)^{-1} (\boldsymbol{\phi}_i^+)^\dagger B^\dagger \boldsymbol{\phi}_i^+].
\tag{3.63}
$$

Here a is the length of the unit cell in the z direction, k_i^+ is the wave vector for i-th right going states, and λ is the Bloch factor defined by $\lambda_i^+ = e^{ik_i^+ a}$.

Thus, transmission probability is obtained as

$$T(\varepsilon) = \sum_{i,j} \frac{v_j'}{v_i} |t_{i,j}|^2. \tag{3.64}$$

Note that the summation runs over right-going propagating states i and j for right and left electrodes, respectively, because the propagating states contribute to the physical transmissions.

3.3.2 Generalized Bloch states

Generalized Bloch states (i.e., solutions of the Kohn-Sham equation for the Hamiltonian of the electrode) play a significant role in the WFM method since the scattering state is expanded by them at the left and right electrode regions. To find the generalized Bloch states, it is necessary to seek the complex wave vectors that satisfy the Kohn-Sham equation under the periodic boundary condition. Because of the block tridiagonal structure of the Kohn-Sham Hamiltonian (3.26), Kohn-Sham equation can be written as the following matrix equation:

$$-\hat{B}^\dagger \boldsymbol{\psi}_{l-1} + (\varepsilon \hat{I} - \hat{A})\boldsymbol{\psi}_l - \hat{B}\boldsymbol{\psi}_{l+1} = 0, \tag{3.65}$$

where $\boldsymbol{\psi}_l$ is the solution of the Kohn-Sham equation for the l-th unit cell. By introducing the Bloch ansatz, $\boldsymbol{\psi}_{l\pm1} = e^{\pm ika}\boldsymbol{\phi}_n$, the above equation can be rewritten as

$$[-e^{-ik_n a}\hat{B}^\dagger + (\varepsilon \hat{I} - \hat{A}) - e^{ik_n a}\hat{B}]\boldsymbol{\phi}_n = 0, \tag{3.66}$$

where n-th complex wave vector k_n and a is the distance between adjacent unit cells. Note that the $\{k_n\}$ and $\{\boldsymbol{\phi}_n\}$ need to be determined for a given input energy ε. The above equation is solved most typically by converting it to the generalized eigenvalue problem with twice matrix dimension of the original problem [55, 72] because generalize eigenvalue problem is readily solved by the standard eigensolvers such as QZ method. It should be noted that the coupling matrix $\hat{B}$ is singular in most cases, resulting that solutions of Eq. (3.66) will have $\text{Im}(k_n) \to \pm\infty$. Fortunately, the contribution of the waves that decay infinitely fast is negligibly small, and therefore it is sufficient to compute $2N_r$ physically important solutions, with N_r being the rank number of the coupling matrix computed with the singular value decomposition technique [53, 55, 73].

Furthermore, $2N_r$ solutions are distinguished into N_r left-going waves $\{\boldsymbol{\phi}_n^-\}$ with either $\text{Im}(k_n) < 0$ or $\text{Im}(k_n) = 0$ and $v_n < 0$, and N_r right-going waves $\{\boldsymbol{\phi}_n^+\}$ with either $\text{Im}(k_n) > 0$ or $\text{Im}(k_n) = 0$ and $v_n > 0$. This is because if k_n is an eigenvalue, then $-k_n, k_n^*$, and $-k_n^*$ are also eigenvalues [72]. From the Bloch ansatz, the solutions with real k_n values are conventional Bloch states while those with imaginary or complex k_n are exponentially growing or decaying waves.

3.4 Relationship between GF and WFM

I have reviewed the NEGF and WFM methods in the sections above. The two seemingly different formalisms introduced in previous sections are closely related. This section shows the relationship between the self-energy matrices for electrodes, transmission probability, and charge density obtained by the NEGF and WFM methods.

3.4.1 Construction of Self-energy matrices from generalized Bloch states

According to Ref. [54], I show that surface Green's function can be constructed by the generalized Bloch states. I focus on the l-th line of the Green's function of the infinite system

$$-\hat{B}^{\dagger}\hat{G}_{l-2,l} + [\varepsilon\hat{I} - \hat{A}]\hat{G}_{l-1,l} - \hat{B}\hat{G}_{l,l} = 0, \tag{3.67}$$

$$-\hat{B}^{\dagger}\hat{G}_{l-1,l} + [\varepsilon\hat{I} - \hat{A}]\hat{G}_{l,l} - \hat{B}\hat{G}_{l+1,l} = \hat{I}, \tag{3.68}$$

$$-\hat{B}^{\dagger}\hat{G}_{l,l} + [\varepsilon\hat{I} - \hat{A}]\hat{G}_{l+1,l} - \hat{B}\hat{G}_{l+2,l} = 0, \tag{3.69}$$

where $\hat{G}_{l,m}$ is the matrix element of the Green's function of the infinite system. Comparing Eqs. (3.67)-(3.69) with Eq. (3.65), it is obvious that the Green's function is just generalized Bloch states except for the diagonal element $\hat{G}_{l,l}$. By the requirement that Green's function does not diverge and smoothly connect at $\hat{G}_{l,l}$, Green's function can be expressed by

$$\hat{G}_{l,m} = \begin{cases} \displaystyle\sum_{j} \phi_j^+ e^{ik_j^+(l-m)a}(\widetilde{\phi}_j^+)^{\dagger}\hat{G}_{l,l} & (l \geq m) \tag{3.70a} \\[2ex] \displaystyle\sum_{j} \phi_j^- e^{ik_j^-(l-m)a}(\widetilde{\phi}_j^-)^{\dagger}\hat{G}_{l,l} & (l \leq m). \tag{3.70b} \end{cases}$$

Substituting Eq. (3.70) into Eq. (3.67), one might obtain

$$-\hat{B}^{\dagger}\Big(\sum_{j}\phi_j^- e^{-ik_j^- a}(\widetilde{\phi}_j^-)^{\dagger}\Big)\hat{G}_{l,l} + [\varepsilon\hat{I} - \hat{A}]\Big(\sum_{j}\phi_j^+(\widetilde{\phi}_j^+)^{\dagger}\Big)\hat{G}_{l,l} - \hat{B}\Big(\sum_{j}\phi_j^+ e^{ik_j^+ a}(\widetilde{\phi}_j^+)^{\dagger}\Big)\hat{G}_{l,l} = \hat{I}. \tag{3.71}$$

On the other hand, multiplying $(\widetilde{\phi}_j^+)^{\dagger}\hat{G}_{l,l}$ from the left side of Eq. (3.66) and summing up for ϕ_j^+, one might obtain

$$-\hat{B}^{\dagger}\Big(\sum_{j}\phi_j^+ e^{-ik_j^+ a}(\widetilde{\phi}_j^+)^{\dagger}\Big)\hat{G}_{l,l} + [\varepsilon\hat{I} - \hat{A}]\Big(\sum_{j}\phi_j^+(\widetilde{\phi}_j^+)^{\dagger}\Big)\hat{G}_{l,l} - \hat{B}\Big(\sum_{j}\phi_j^+ e^{ik_j^+ a}(\widetilde{\phi}_j^+)^{\dagger}\Big)\hat{G}_{l,l} = 0. \tag{3.72}$$

From Eqs. (3.71) and (3.72), the diagonal term $\hat{G}_{l,l}$ is given by

$$\hat{G}_{l,l}^{-1} = \hat{B}^{\dagger}\Big[\sum_{j}\phi_j^+ e^{-ik_j^+ a}(\widetilde{\phi}_j^+)^{\dagger} - \sum_{j}\phi_j^- e^{-ik_j^- a}(\widetilde{\phi}_j^-)^{\dagger}\Big]^{-1}. \tag{3.73}$$

In the same way, arranging so as to leave $\hat{B}$, one might obtain

$$\hat{G}_{l,l}^{-1} = \hat{B}\Big[\sum_{j}\phi_j^- e^{ik_j^- a}(\widetilde{\phi}_j^-)^{\dagger} - \sum_{j}\phi_j^+ e^{ik_j^+ a}(\widetilde{\phi}_j^+)^{\dagger}\Big]^{-1}. \tag{3.74}$$

Next, I consider to make the surface Green's function $\hat{G}_S$ from that of infinite system. This is achieved by subtracting the surface term from the Green's function of infinite system,

$$\hat{G}_{S,n,m} = \hat{G}_{n,m} - \hat{G}_{n,0}\hat{G}_{0,0}^{-1}\hat{G}_{0,m}. \tag{3.75}$$

When putting $n \geq 0$ is the vacuum region and $n < 0$ is the bulk region, the left surface Green's function $\hat{g}_{00}^L (= \hat{G}_{S,-1,-1})$ is obtained as

$$\hat{g}_{0,0}^L = \left[\hat{I} - \left(\sum_j \phi_j^- e^{-ik_j^- a} (\widetilde{\phi}_j^-)^\dagger \right) \left(\sum_j \phi_j^+ e^{ik_j^+ a} (\widetilde{\phi}_j^+)^\dagger \right) \right] \hat{G}_{0,0}. \tag{3.76}$$

Similarly for right surface Green's function,

$$\hat{g}_{0,0}^R = \left[\hat{I} - \left(\sum_j \phi_j^+ e^{ik_j^+ a} (\widetilde{\phi}_j^+)^\dagger \right) \left(\sum_j \phi_j^- e^{-ik_j^- a} (\widetilde{\phi}_j^-)^\dagger \right) \right] \hat{G}_{0,0}. \tag{3.77}$$

Thus, once the solutions of Eq. (3.66) are obtained, Green's function of the infinite system and surface Green's functions are constructed by them analytically. This is so called semi-analytical method proposed by Sanvito and his coworkers [54].

Next, I clarify the relationship between the self-energy matrices defined by NEGF and WFM methods. For this purpose, I introduce the transfer matrices [55]

$$\hat{T}_L^- = \sum_j \phi_{L1,j}^- (\widetilde{\phi}_{L0,j}^-)^\dagger = \sum_j \phi_{L0,j}^- e^{-ik_j^- a} (\widetilde{\phi}_{L0,j}^-)^\dagger, \tag{3.78}$$

$$\hat{T}_R^+ = \sum_j \phi_{R1,j}^+ (\tilde{\phi}_{R0,j}^+)^\dagger = \sum_j \phi_{R0,j}^+ e^{ik_j^+ a} (\tilde{\phi}_{R0,j}^+)^\dagger. \tag{3.79}$$

The subscripts L and R indicate the left and right electrodes, respectively. Using the transfer matrices, one can rewrite Green's function of the infinite system in Eqs. (3.73) and (3.74) and surface Green's functions in Eqs. (3.76) and (3.81) as follows:

$$\begin{aligned}
\hat{G}_{0,0}^{-1} &= \hat{B}^\dagger ((\hat{T}_R^+)^{-1} - \hat{T}_L^-) \\
&= \hat{B}((\hat{T}_L^-)^{-1} - \hat{T}_R^+)),
\end{aligned} \tag{3.80}$$

$$\begin{aligned}
\hat{g}_{00}^L &= (\hat{I} - \hat{T}_L^- \hat{T}_R^+)[\hat{B}((\hat{T}_L^-)^{-1} - \hat{T}_R^+)]^{-1} \\
&= \hat{T}_L^- \hat{B}^{-1},
\end{aligned} \tag{3.81}$$

$$\begin{aligned}
\hat{g}_{00}^R &= (\hat{I} - \hat{T}_R^+ \hat{T}_L^-)[\hat{B}^\dagger ((\hat{T}_R^+)^{-1} - \hat{T}_L^-)]^{-1}, \\
&= \hat{T}_R^+ (\hat{B}^\dagger)^{-1}.
\end{aligned} \tag{3.82}$$

By substituting Eqs. (3.81) and (3.82) into Eqs. (3.27) and (3.28), one might obtain

$$\hat{\Sigma}_L(\varepsilon) = \hat{B}^\dagger \hat{T}_L^-, \tag{3.83}$$

$$\hat{\Sigma}_R(\varepsilon) = \hat{B}\hat{T}_R^+. \tag{3.84}$$

Compared with Eqs. (3.57) and (3.58), the self-energy matrices obtained by the NEGF method are identical to that obtained by the WFM method.

3.4.2   Transmission probability

Let us prove the equivalence of Eq. (3.25) and Eq. (3.64). In preparation, I introduce the Bloch matrices

$$\hat{Q}_{Lk} = [\phi^+_{Lk,1}, \phi^+_{Lk,2}, ..., \phi^+_{Lk,N_r}], \quad \hat{Q}_{Rk} = [\phi^+_{Rk,1}, \phi^+_{Rk,2}, ..., \phi^+_{Rk,N_r}], \tag{3.85}$$

$$\widetilde{Q}_{Lk} = [\widetilde{\phi}^+_{Lk,1}, \widetilde{\phi}^+_{Lk,2}, ..., \widetilde{\phi}^+_{Lk,N_r}], \quad \widetilde{Q}_{Rk} = [\widetilde{\phi}^+_{Rk,1}, \widetilde{\phi}^+_{Rk,2}, ..., \widetilde{\phi}^+_{Rk,N_r}]. \tag{3.86}$$

From the definition of the dual vectors, $\hat{Q}_{Lk}\widetilde{Q}^\dagger_{Lk} = \hat{Q}_{Rk}\widetilde{Q}^\dagger_{Rk} = \hat{I}$, and $\widetilde{Q}^\dagger = \hat{Q}^{-1}$ if $\hat{Q}$ is full rank. Furthermore, one can write

$$\hat{Q}_{L1} = \hat{T}^+_L \hat{Q}_{L0}, \tag{3.87}$$

where $\hat{T}^+_L$ is the transfer matrix for the right-going states in the left electrode defined by

$$\hat{T}^+_L = \sum_j \phi^+_{L1,j}(\tilde{\phi}^+_{L0,j})^\dagger = \sum_j \phi^+_{L0,j} e^{-ik^+_j a}(\widetilde{\phi}^+_{L0,j})^\dagger. \tag{3.88}$$

In the exactly same manner as deriving Eqs. (3.83) and (3.84), the advanced self-energy matrix for the left electrode is defined by

$$\hat{\Sigma}^\dagger_L(\varepsilon) = \hat{B}^\dagger \hat{T}^+_L. \tag{3.89}$$

Using Eq. (3.61), (3.87), and (3.89), the scattering state in the transition region can be rewritten as

$$\boldsymbol{\psi}_T = i\hat{G}_T \hat{\Gamma}_L \boldsymbol{\phi}_{L0,i}. \tag{3.90}$$

To derive the Fisher-Lee formula from the WFM method, I rewrite the total transmission probability of Eq. (3.64) to the following matrix form

$$T = \text{Tr}[\hat{\mathcal{T}}^\dagger \hat{\mathcal{V}}_R \hat{\mathcal{T}} \hat{\mathcal{V}}^{-1}_L], \tag{3.91}$$

where $\hat{\mathcal{T}}_{i,j} = t_{i,j}$ and $\hat{\mathcal{V}}_{L/R}$ is the velocity matrix for the left or right electrode (see Appendix A). Using Eq. (3.53) and (3.90),

$$\hat{\mathcal{T}} = i\widetilde{Q}^\dagger_{R0}\hat{G}_T \hat{\Gamma}_L \hat{Q}_{L0}. \tag{3.92}$$

In addition, it is easy to show that

$$\begin{aligned}
\hat{\mathcal{V}}_L &= i[\hat{Q}^\dagger_{L1}\hat{B}\hat{Q}_{L0} - \hat{Q}^\dagger_{L0}\hat{B}^\dagger\hat{Q}_{L1}] \\
&= i\hat{Q}^\dagger_{L0}[(\hat{B}^\dagger\hat{T}^+_L)^\dagger - \hat{B}^\dagger\hat{T}^+_L]\hat{Q}_{L0} \\
&= i\hat{Q}^\dagger_{L0}[\hat{\Sigma}_L(\varepsilon) - \hat{\Sigma}^\dagger_L(\varepsilon)]\hat{Q}_{L0} \\
&= \hat{Q}^\dagger_{L0}\hat{\Gamma}_L(\varepsilon)\hat{Q}_{L0},
\end{aligned} \tag{3.93}$$

and

$$\begin{aligned}
\hat{\mathcal{V}}_R &= i[\hat{Q}^\dagger_{R0}\hat{B}\hat{Q}_{R1} - \hat{Q}^\dagger_{R1}\hat{B}^\dagger\hat{Q}_{R0}] \\
&= i\hat{Q}^\dagger_{R0}[\hat{B}\hat{T}^+_R - (\hat{T}^+_R)^\dagger\hat{B}^\dagger]\hat{Q}_{R0} \\
&= i\hat{Q}^\dagger_{R0}[\hat{\Sigma}_R(\varepsilon) - \hat{\Sigma}^\dagger_R(\varepsilon)]\hat{Q}_{R0} \\
&= \hat{Q}^\dagger_{R0}\hat{\Gamma}_R(\varepsilon)\hat{Q}_{R0}.
\end{aligned} \tag{3.94}$$

Substituting Eqs. (3.92)-(3.94), into Eq. (3.91) leads the Fisher-Lee formula (3.25).

3.4.3 Charge density

It is possible to evaluate the charge density required for the self-consistent calculation directly from the wavefunctions. The charge density in the non-equilibrium condition can be computed by summing up the scattering states injected from the left and right electrodes

$$\tilde{n}_T(\mathbf{r}) = \int_{-\infty}^{\infty} |\psi_L^+(\varepsilon;\mathbf{r})|^2 f(\varepsilon - \mu_L)d\varepsilon + \int_{-\infty}^{\infty} |\psi_R^-(\varepsilon;\mathbf{r})|^2 f(\varepsilon - \mu_R)d\varepsilon \tag{3.95}$$

where $\psi_L^+(\varepsilon;\mathbf{r})$ is the right-going scattering states injected from the left electrode and $\psi_R^-(\varepsilon;\mathbf{r})$ is the left-going scattering states injected from the right electrode. To see the relationship between the above equation and Eq. (3.21), it is useful to introduce the density of states $\rho(\varepsilon;\mathbf{r})$, which is divided into the contributions from left and right electrodes,

$$\rho(\varepsilon;\mathbf{r}) = \rho_L(\varepsilon;\mathbf{r}) + \rho_R(\varepsilon;\mathbf{r}), \tag{3.96}$$

where

$$\rho_L(\varepsilon;\mathbf{r}) = |\psi_L^+(\varepsilon;\mathbf{r})|^2, \tag{3.97}$$

$$\rho_R(\varepsilon;\mathbf{r}) = |\psi_R^-(\varepsilon;\mathbf{r})|^2. \tag{3.98}$$

On the other hand, according to the Green's function theory, the density of states is given by

$$\rho(\varepsilon;\mathbf{r}) = -\frac{1}{\pi}\text{Im}\hat{G}(\varepsilon;\mathbf{r},\mathbf{r}). \tag{3.99}$$

Since $(\hat{G}^\dagger)^{-1} - \hat{G}^{-1} = \hat{\Sigma} - \hat{\Sigma}^\dagger = -i\hat{\Gamma} = -i[\hat{\Gamma}_L + \hat{\Gamma}_R]$,

$$\text{Im}\hat{G} = \frac{i}{2}[\hat{G} - \hat{G}^\dagger]$$

$$= \frac{1}{2}\hat{G}[\hat{\Gamma}_L + \hat{\Gamma}_R]\hat{G}^\dagger. \tag{3.100}$$

By substituting Eq. (3.100) to Eq. (3.99), one can confirm the equivalence of Eq. (3.95) and Eq. (3.21). It should be noted that the equivalence is valid only when the infinitesimal η can be neglected. Actually, as pointed out in Refs. [41,42,62], the expression of the charge density in Eq. (3.95) is not true because the bound states that do not couple to the electrodes contribute to the charge density. If bound states exist, total charge density in the transition region is given by

$$n_T(\mathbf{r}) = \tilde{n}_T(\mathbf{r}) + n_b(\mathbf{r}). \tag{3.101}$$

The second term is the contribution from the bound states

$$n_b(\mathbf{r}) = \sum_{\varepsilon_b} |\psi(\varepsilon_b;\mathbf{r})|^2 \mathcal{F}(\varepsilon_b), \tag{3.102}$$

where ε_b is the discrete energy levels of the bound states, $\mathcal{F}(\varepsilon_b)$ is the occupation number of the bound states, and $\psi(\varepsilon_b;\mathbf{r})$ is the bound state, which is the solutions of the nonlinear eigenvalue problem:

$$[\hat{H}_T + \hat{\Sigma}_L(\varepsilon_b) + \hat{\Sigma}_R(\varepsilon_b)]\psi_{\varepsilon_b} = \varepsilon_b\psi_{\varepsilon_b}, \tag{3.103}$$

where $\psi(\varepsilon_b; \mathbf{r}) = \langle \mathbf{r} | \boldsymbol{\psi}_{\varepsilon_b} \rangle$. The above equation can be derived from Eq. (3.61) by putting $\boldsymbol{\Omega}_i = 0$.

To evaluate the contribution from the bound states, it is required to give some representation for $\mathcal{F}(\varepsilon_b)$. Obviously, $\mathcal{F}(\varepsilon_b) = 1$ ($\mathcal{F}(\varepsilon_b) = 0$) when ε_b is lower (higher) than two chemical potentials μ_L and μ_R. On the other hand, if the bound states have energies within the bias window, $\mathcal{F}(\varepsilon_b)$ remains unknown and additional information on the filling is needed. Fortunately, these situations rarely arise in practice since a certain physical interaction exits between electrodes. Otherwise, electron transport behavior becomes quantum-dot like in the Coulomb blockade regime, which is out of our interest.

Fig. 3.3 illustrates the one-to-one correspondence between NEGF and WFM methods.
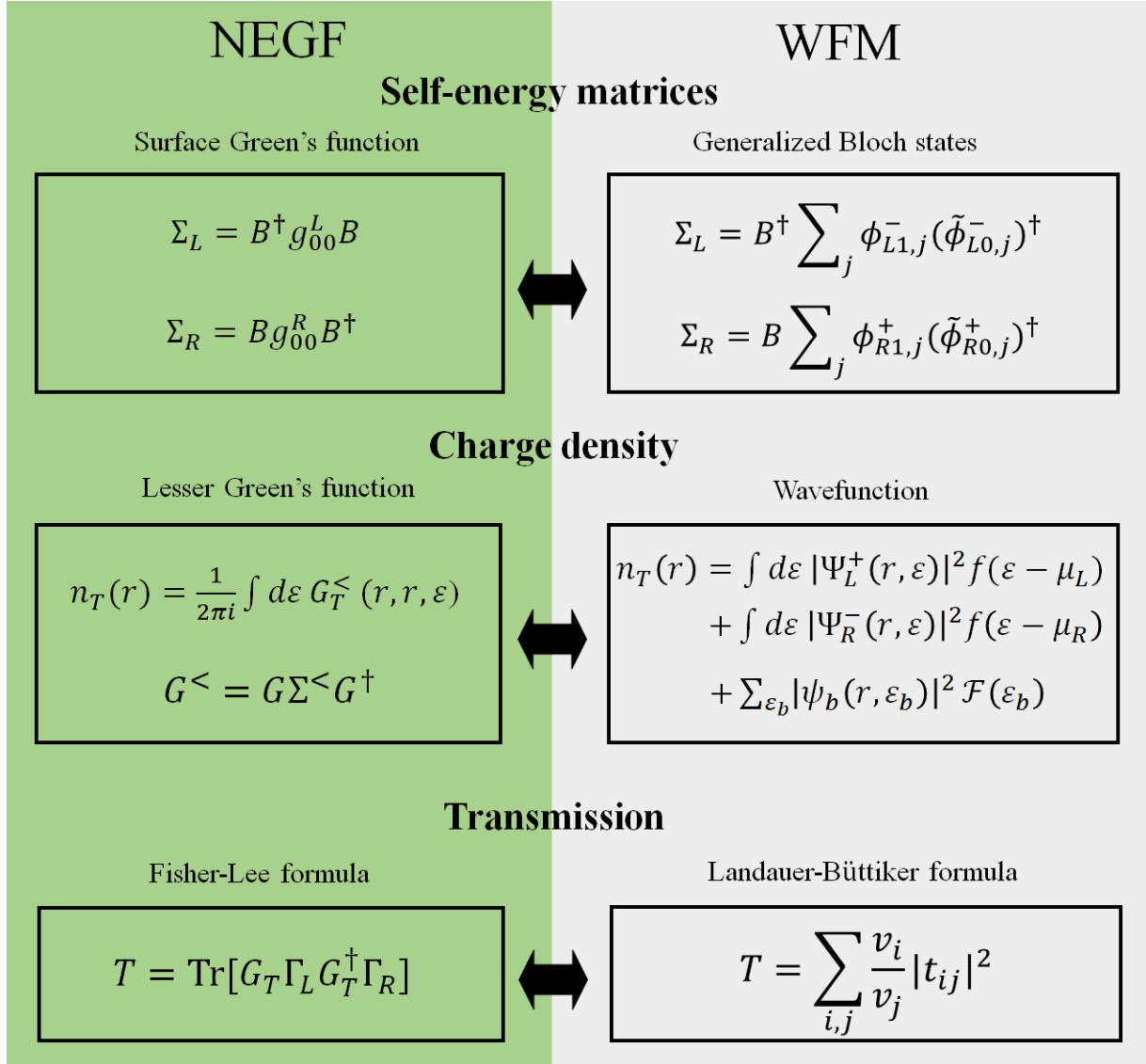
Figure 3.3: One-to-one correspondence between NEGF and WFM methods

Chapter 4

Evaluation of self-energy matrices on real-space finite-difference scheme

In Chap. 3, I reviewed the typical methods to calculate the self-energy matrices. These methods have been used commonly with the tight-binding approximation or localized basis set. While one can apply them for any spatially localized basis set, the calculation is still slow especially when the real-space finite-difference (RSFD) scheme [74, 75] is employed because the size of the Hamiltonian matrix in the unit cell becomes very large. To overcome this difficulty, efficient implementations for the RSFD scheme have been developed so far [67, 68, 73, 76]. However, the relationship between the typical methods such as RGF, quick decimation, and semi-analytical methods, and those based on the RSFD scheme is not seemingly obvious. In addition, in my knowledge, there is no study which investigates whether the quick decimation method is applicable for RSFD scheme or not. In this chapter, I first show that the Hamiltonian matrix of the size of the unit cell can be reduced to the contracted Hamiltonian matrix of the size of the electrode principal layer by introducing the partitioning technique in Ref. [68]. Next, the methods introduced in Chap. 3 can be used for a contracted Hamiltonian matrix without large modification. Finally, I introduce the several efficient implementations to compute the self-energy matrices using the singular value decomposition. The results and discussions in this chapter are unpublished.

4.1   Basic concept

The RSFD scheme leads to the structured, very sparse matrix representation of the Hamiltonian as well as other real-space methods such as finite elements, or wavelets. In this study, real-space calculations are performed on uniform Cartesian meshes, and the Laplacian operator is approximated by the higher-order finite-difference approximation. The unit cell consists of $M(= M_x \times M_y \times M_z)$ grid points, where $M_x, M_y,$ and M_z grid points in the $x, y,$ and z directions, respectively. In addition, the direction of current flow is assumed to be along z axis. Due to the sparsity of real-space Hamiltonian matrix, the unit cell can be decomposed into m electrode principal layers and interaction exists only

between the nearest neighboring layers, which allows us to rewrite $\hat{A}$ and $\hat{B}$ in Eq. (3.26) as

$$
\hat{A} = \begin{bmatrix} \hat{A}_1 & \hat{B}_1 & & 0 \\ \hat{B}_1^\dagger & \ddots & \ddots & \\ & \ddots & \ddots & \hat{B}_{m-1} \\ 0 & & \hat{B}_{m-1}^\dagger & \hat{A}_m \end{bmatrix}, \quad \hat{B} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \\ \hat{B}_0 & 0 & \cdots & 0 \end{bmatrix}, \tag{4.1}
$$

where $\hat{A}_i$ and $\hat{B}_i$ are block-matrix elements with the dimension of $\widetilde{M}(= M_x \times M_y \times M_{B0})$. M_{B0} corresponds to the number of grid points in the z direction, which is defined so as to cover the nonlocal region of the pseudopotentials. Note that $\hat{B}_0 = \hat{B}_m$ due to the periodicity. Because $\hat{B}$ has the nonzero term only at the left corner, self-energy matrix for the left electrode defined in Eq. (3.27) has the form

$$
\begin{aligned}
\hat{\Sigma}_L(\varepsilon) &= \hat{B}^\dagger \hat{g}_{00}^L \hat{B} \\
&= \begin{bmatrix} 0 & \cdots & 0 & \hat{B}_0^\dagger \\ 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{g}_{11}^{L,00} & \cdots & \cdots & \hat{g}_{1,m}^{L,00} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \hat{g}_{m,1}^{L,00} & \cdots & \cdots & \hat{g}_{m,m}^{L,00} \end{bmatrix} \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \\ \hat{B}_0 & 0 & \cdots & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & \hat{B}_0^\dagger \hat{g}_{m,m}^{L,00} \hat{B}_0 \end{bmatrix}, \tag{4.2}
\end{aligned}
$$

where $\hat{g}_{m,m}^{L,00}$ is the left surface Green's function of the outermost (m-th) electrode principal layer in L0 layer. Equation (4.2) means that it is unnecessary to compute $\hat{g}_{00}^L$ with the size of the unit cell, but need the only $\hat{g}_{m,m}^{L,00}$ with the size of the electrode principal layer. Similarly for the self-energy matrix for right electrode, one may obtain

$$
\begin{aligned}
\hat{\Sigma}_R(\varepsilon) &= \hat{B} \hat{g}_{00}^R \hat{B}^\dagger \\
&= \begin{bmatrix} \hat{B}_0 \hat{g}_{11}^{R,00} \hat{B}_0^\dagger & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \tag{4.3}
\end{aligned}
$$

where $\hat{g}_{11}^{R,00}$ is the right surface Green's function of the outermost (1st) electrode principal layer in R0 layer.

In order to calculate the self-energy matrices efficiently, I apply the partitioning technique in Ref. [68]. I first assume that $\hat{B}_0$ is full rank, that is, $N_r = \widetilde{M}$, and consider the case $N_r \neq \widetilde{M}$ in the section later. To eliminate trivial solutions of Eq. (3.65), I partition the Hamiltonian matrix into $\widetilde{M}$ and $(m-1)\widetilde{M}$ dimensional matrices, and then Eq. (3.65) can be rewritten as

$$
\begin{bmatrix}
\ddots & \ddots & & & & & \\
\ddots & \hat{\alpha}_1 & \hat{\beta}_1 & & & & \\
 & \hat{\beta}_1^\dagger & \hat{\alpha}_2 & \hat{\beta}_2 & & & \\
 & & \hat{\beta}_2^\dagger & \hat{\alpha}_1 & \hat{\beta}_1 & & \\
 & & & \hat{\beta}_1^\dagger & \hat{\alpha}_2 & \hat{\beta}_2 & \\
 & & & & \hat{\beta}_2^\dagger & \hat{\alpha}_1 & \hat{\beta}_1 \\
 & & & & & \hat{\beta}_1^\dagger & \hat{\alpha}_2 & \ddots \\
 & & & & & & \ddots & \ddots
\end{bmatrix}
\begin{bmatrix}
\vdots \\
\boldsymbol{\psi}_1^{l-1} \\
\boldsymbol{\psi}_{2:m}^{l-1} \\
\boldsymbol{\psi}_1^{l} \\
\boldsymbol{\psi}_{2:m}^{l} \\
\boldsymbol{\psi}_1^{l+1} \\
\boldsymbol{\psi}_{2:m}^{l+1} \\
\vdots
\end{bmatrix}
= \varepsilon
\begin{bmatrix}
\vdots \\
\boldsymbol{\psi}_1^{l-1} \\
\boldsymbol{\psi}_{2:m}^{l-1} \\
\boldsymbol{\psi}_1^{l} \\
\boldsymbol{\psi}_{2:m}^{l} \\
\boldsymbol{\psi}_1^{l+1} \\
\boldsymbol{\psi}_{2:m}^{l+1} \\
\vdots
\end{bmatrix},
\tag{4.4}
$$

where

$$
\hat{\alpha}_1 = \hat{A}_1, \quad \hat{\alpha}_2 =
\begin{bmatrix}
\hat{A}_2 & \hat{B}_2 & & 0 \\
\hat{B}_2^\dagger & \ddots & \ddots & \\
 & \ddots & \ddots & \hat{B}_{m-1} \\
0 & & \hat{B}_{m-1}^\dagger & \hat{A}_m
\end{bmatrix}, \quad
\hat{\beta}_1 = [\hat{B}_1, 0, ..., 0], \quad
\hat{\beta}_2 =
\begin{bmatrix}
0 \\
0 \\
\vdots \\
\hat{B}_0
\end{bmatrix}, \quad
\boldsymbol{\psi}_l =
\begin{bmatrix}
\boldsymbol{\psi}_1^l \\
\boldsymbol{\psi}_{2:m}^l
\end{bmatrix}.
\tag{4.5}
$$

The generalized Bloch states will satisfy that

$$
-\hat{\beta}_1^\dagger \boldsymbol{\psi}_1^{l-1} + (\varepsilon\hat{I} - \hat{\alpha}_2)\boldsymbol{\psi}_{2:m}^{l-1} - \hat{\beta}_2\boldsymbol{\psi}_1^l = 0, \tag{4.6}
$$

$$
-\hat{\beta}_2^\dagger \boldsymbol{\psi}_{2:m}^{l-1} + (\varepsilon\hat{I} - \hat{\alpha}_1)\boldsymbol{\psi}_1^l - \hat{\beta}_1\boldsymbol{\psi}_{2:m}^l = 0, \tag{4.7}
$$

$$
-\hat{\beta}_1^\dagger \boldsymbol{\psi}_1^l + (\varepsilon\hat{I} - \hat{\alpha}_2)\boldsymbol{\psi}_{2:m}^l - \hat{\beta}_2\boldsymbol{\psi}_1^{l+1} = 0. \tag{4.8}
$$

From Eqs. (4.6) and (4.8),

$$
\boldsymbol{\psi}_{2:m}^{l-1} = (\varepsilon\hat{I} - \hat{\alpha}_2)^{-1}(\hat{\beta}_1^\dagger \boldsymbol{\psi}_1^{l-1} + \hat{\beta}_2\boldsymbol{\psi}_1^l), \tag{4.9}
$$

$$
\boldsymbol{\psi}_{2:m}^{l} = (\varepsilon\hat{I} - \hat{\alpha}_2)^{-1}(\hat{\beta}_1^\dagger \boldsymbol{\psi}_1^{l} + \hat{\beta}_2\boldsymbol{\psi}_1^{l+1}). \tag{4.10}
$$

Substituting Eqs. (4.9) and (4.10) into Eq. (4.7) yields

$$
-\hat{\mathcal{B}}^\dagger \boldsymbol{\psi}_1^{l-1} + (\varepsilon\hat{I} - \hat{\mathcal{A}})\boldsymbol{\psi}_1^l - \hat{\mathcal{B}}\boldsymbol{\psi}_1^{l+1} = 0, \tag{4.11}
$$

where $\hat{g}_{11}^{R,k0}$ is the right surface Green's function of the 1st electrode principal layer in Lk layer, and

$$\hat{\mathcal{A}} = \hat{\alpha}_1 + \hat{\beta}_2^\dagger(\varepsilon\hat{I} - \hat{\alpha}_2)^{-1}\hat{\beta}_2 + \hat{\beta}_1(\varepsilon\hat{I} - \hat{\alpha}_2)^{-1}\hat{\beta}_1^\dagger, \tag{4.12}$$

$$\hat{\mathcal{B}} = \hat{\beta}_1(\varepsilon\hat{I} - \hat{\alpha}_2)^{-1}\hat{\beta}_2. \tag{4.13}$$

Using the Bloch ansatz, $\boldsymbol{\psi}_1^{l\pm1} = \lambda^{\pm1}\boldsymbol{\phi}_1^l$, one might obtain the quadratic eigenvalue problem for $\boldsymbol{\phi}_1^l$:

$$[-\lambda^{-1}\hat{\mathcal{B}}^\dagger + (\varepsilon\hat{I} - \hat{\mathcal{A}}) - \lambda\hat{\mathcal{B}}]\boldsymbol{\phi}_1^l = 0. \tag{4.14}$$

Once solving Eq. (4.14), $\boldsymbol{\phi}_{2:m}^l$ will be obtained from Eq. (4.10). Note that the above derivation was firstly done by Khomyakov *et al.* [68]. A contracted form of Kohn-Sham equation (4.14) is particularly attractive against Eq. (3.65) because the problem is reduced to the only $\widetilde{M}$-dimensional eigenvalue problem compared with $M(=m\widetilde{M})$-dimensional eigenvalue problem of the original equation.

Next I derive a contracted formula for the self-energy matrix. I first consider the right electrode and the extension to the left electrode will be presented in the future section. The 0-th column of the right surface Green's function which is the outermost one and needed for the transport calculation satisfies

$$\begin{bmatrix} z\hat{I} - \hat{A} & -\hat{B} & & 0 \\ -\hat{B}^\dagger & z\hat{I} - \hat{A} & -\hat{B} & \\ & -\hat{B}^\dagger & z\hat{I} - \hat{A} & \ddots \\ 0 & & \ddots & \ddots \end{bmatrix} \begin{bmatrix} \hat{g}_{00}^R \\ \hat{g}_{10}^R \\ \hat{g}_{20}^R \\ \vdots \end{bmatrix} = \begin{bmatrix} \hat{I} \\ 0 \\ 0 \\ \vdots \end{bmatrix}, \tag{4.15}$$

where $z = \varepsilon + i\eta$. By the exactly same manner as deriving Eq. (4.14), it is easy to obtain a constructed form of the right surface Green's function

$$\begin{bmatrix} z\hat{I} - \hat{\mathcal{A}}_\mathcal{S} & -\hat{\mathcal{B}} & & 0 \\ -\hat{\mathcal{B}}^T & z\hat{I} - \hat{\mathcal{A}} & -\hat{\mathcal{B}} & \\ & -\hat{\mathcal{B}}^T & z\hat{I} - \hat{\mathcal{A}} & \ddots \\ 0 & & \ddots & \ddots \end{bmatrix} \begin{bmatrix} \hat{g}_{11}^{R,00} \\ \hat{g}_{11}^{R,10} \\ \hat{g}_{11}^{R,20} \\ \vdots \end{bmatrix} = \begin{bmatrix} \hat{I} \\ 0 \\ 0 \\ \vdots \end{bmatrix}, \tag{4.16}$$

where

$$\hat{\mathcal{A}}_\mathcal{S} = \hat{\alpha}_1 + \hat{\beta}_1(z\hat{I} - \hat{\alpha}_2)^{-1}\hat{\beta}_1^\dagger. \tag{4.17}$$

Note that $\hat{\mathcal{B}}^T = \hat{\mathcal{B}}^\dagger$ when $\eta = 0$ [1]. Compared with Eqs. (4.15) and (4.16), one can see that the translational symmetry of the contracted Hamiltonian is broken only at the outermost principal layer

---

[1]It should be noted that $\hat{\mathcal{B}}^T = \hat{\beta}_2^\dagger(z\hat{I} - \hat{\alpha}_2)^{-1}\hat{\beta}_1^\dagger$ is established only when A_i and B_i are real matrices whose elements consist entirely of real numbers. Although we need to introduce the new matrix $\hat{\mathcal{C}} = \hat{\beta}_2^\dagger(z\hat{I} - \hat{\alpha}_2)^{-1}\hat{\beta}_1^\dagger$ if A_i and B_i are complex Hermite matrices, all formulations in this chapter are available with minor modification.

since $\hat{\mathcal{A}}_S \neq \hat{\mathcal{A}}$, leading the sophisticated iteration methods for computing the surface Green's function are not applicable directly.

Fortunately, this is a trivial problem. Using the matrix identity for the matrix inversion,

$$
\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{a_{11}-a_{12}a_{22}^{-1}a_{21}} & -a_{11}^{-1}a_{12}\frac{1}{a_{22}-a_{21}a_{11}^{-1}a_{12}} \\ -a_{22}^{-1}a_{21}\frac{1}{a_{11}-a_{12}a_{22}^{-1}a_{21}} & \frac{1}{a_{22}-a_{21}a_{11}^{-1}a_{12}} \end{bmatrix}, \tag{4.18}
$$

$\hat{g}_{11}^{R,00}$ is given by

$$
\hat{g}_{11}^{R,00} = [z\hat{I} - \hat{\mathcal{A}}_S - \hat{\mathcal{B}}(z\hat{I} - \hat{\mathcal{H}}_R^{bulk})^{-1}\hat{\mathcal{B}}^T]^{-1}, \tag{4.19}
$$

where

$$
\hat{\mathcal{H}}_R^{bulk} = \begin{bmatrix} \hat{\mathcal{A}} & \hat{\mathcal{B}} & & 0 \\ \hat{\mathcal{B}}^T & \hat{\mathcal{A}} & \hat{\mathcal{B}} & \\ & \hat{\mathcal{B}}^T & \hat{\mathcal{A}} & \ddots \\ 0 & & \ddots & \ddots \end{bmatrix}. \tag{4.20}
$$

It is obvious that $\hat{\mathcal{H}}_R^{bulk}$ has the translational symmetry for the transport direction and therefore iterative technique is applicable for computing the self-energy-like matrix, $\hat{\Sigma}_R^{bulk} = \hat{\mathcal{B}}(z\hat{I} - \hat{\mathcal{H}}_R^{bulk})^{-1}\hat{\mathcal{B}}^T$. For example, $\hat{\Sigma}_R^{bulk}$ can be computed by the RGF method of Eq. (3.31) as below

$$
\hat{\Sigma}_R^{(n+1)} = \hat{\mathcal{B}}[z\hat{I} - \hat{\mathcal{A}} - \hat{\Sigma}_R^{(n)}]^{-1}\hat{\mathcal{B}}^T, \tag{4.21}
$$

where n is the number of iteration. In the case of the semi-analytical method, one can easily extend Eq. (3.84) for $\hat{\Sigma}_R^{bulk}$, that is,

$$
\hat{\Sigma}_R^{bulk} = \hat{\mathcal{B}}\hat{\mathcal{T}}_R^+, \tag{4.22}
$$

where $\hat{\mathcal{T}}_R^+$ is defined by

$$
\hat{\mathcal{T}}_R^+ = \sum_j \phi_{R,j,1}^+ e^{ik_j^+ a}(\tilde{\phi}_{R,j,1}^+)^\dagger, \tag{4.23}
$$

where $\phi_{R,j,1}^+$ and $\tilde{\phi}_{R,j,1}^+$ are right-going generalized Bloch states on the 1st electrode principal layer and its dual vector, respectively.

4.2 Computational aspects

In this section, I briefly discuss the efficient implementation. One might think that all matrix elements of $(z\hat{I}-\hat{\alpha}_2)^{-1}$ are needed to construct $\mathcal{A}, \mathcal{B}$, and $\mathcal{A}_S$. This is not the case, but the corner block elements of the inverted matrix are necessary because of the sparsity of $\hat{\beta}_1$ and $\hat{\beta}_2$. The matrix elements of

$(z\hat{I} - \hat{\alpha}_2)^{-1}$ can be written as

$$(z\hat{I} - \hat{\alpha}_2)^{-1} = \begin{bmatrix} \hat{D}_{22} & \cdots & \hat{D}_{2,m} \\ \vdots & \ddots & \vdots \\ \hat{D}_{m,2} & \cdots & \hat{D}_{m,m} \end{bmatrix}. \tag{4.24}$$

From Eqs. (4.12), (4.13), and (4.17), the terms that involve $(z\hat{I} - \hat{\alpha}_2)^{-1}$ are reduced to the following forms

$$\hat{\beta}_2^\dagger (z\hat{I} - \hat{\alpha}_2)^{-1} \hat{\beta}_2 = \hat{B}_0^\dagger \hat{D}_{m,m} \hat{B}_0, \tag{4.25}$$

$$\hat{\beta}_1 (z\hat{I} - \hat{\alpha}_2)^{-1} \hat{\beta}_1^\dagger = \hat{B}_1 \hat{D}_{2,2} \hat{B}_1^\dagger, \tag{4.26}$$

$$\hat{\beta}_1 (z\hat{I} - \hat{\alpha}_2)^{-1} \hat{\beta}_2 = \hat{B}_1 \hat{D}_{2,m} \hat{B}_0. \tag{4.27}$$

It should be noted that the block matrix elements of $\hat{D}_{22}$, $\hat{D}_{2,m}$, and $\hat{D}_{m,m}$ are evaluated efficiently by use of the iterative solvers. The details of the iterative solver will be presented in the later chapter.

In case that the semi-analytical method is used, it is efficient to solve the contracted quadratic eigenvalue problem (4.14) directly by the contour integral eigenvalue solver, such as Sakurai-Sugiura method [77]. In this method, the most time-consuming part is inverting the matrix

$$\hat{\mathcal{P}}^{-1}(z_j) = [-z_j^{-1}\hat{\mathcal{B}}^\dagger + (\varepsilon\hat{I} - \hat{\mathcal{A}}) - z_j\hat{\mathcal{B}}]^{-1}, \tag{4.28}$$

on each quadrature point z_j. As noted before, the quadratic eigenvalue problem is usually solved by transforming the generalized eigenvalue with twice matrix size of $\hat{\mathcal{P}}$ [67, 68]. However, a significant reduction of the computational time will be achieved by directly solving the quadratic eigenvalue problem because the matrix inversion scales to $\mathcal{O}(N_{matrix}^3)$, where N_{matrix} is the dimension of the matrix to be inverted.

In addition, because $(\varepsilon\hat{I} - \hat{\mathcal{A}})^\dagger = (\varepsilon\hat{I} - \hat{\mathcal{A}})$, one might obtain the relationship

$$\hat{\mathcal{P}}^\dagger(z_j) = \hat{\mathcal{P}}(z_j^{*-1}). \tag{4.29}$$

Thus, if either $\hat{\mathcal{P}}^{-1}(z_j)$ or LU-decomposed matrix of $\hat{\mathcal{P}}(z_j)$ at outer quadrature points z_j is stored, one can immediately obtain the resolvent at the inner quadrature points z_j^{*-1}. It means that one can save the cost of the integration by almost half. From the discussion above, it is estimated that the time- and memory-saving of the proposed method are about 16 times and 4 times, respectively, against the standard eigenvalue approaches proposed for the RSFD scheme [67, 68].

4.3 Singular value decomposition

If the nonlocal region of pseudopotentials are larger than the order of the finite-difference, off-diagonal elements of the Hamiltonian, i.e., $\hat{B}_0$ and $\hat{B}_1$ become usually singular, which results in that $\hat{\mathcal{B}}$ is

also singular. In this case, the quadratic eigenvalue problem will be ill-condition and it is expected that the accuracy degradation will occur. This problem can be resolved by use of the singular value decomposition technique [55]:

$$\hat{\mathcal{B}} = \hat{U}\hat{S}\hat{V}^{\dagger}, \tag{4.30}$$

where $\hat{U}$ and $\hat{V}$ are unitary matrices and $\hat{S}$ is the diagonal matrix whose diagonal elements are singular values s_i. Suppose that the rank of $\hat{\mathcal{B}}$ is N_r, $s_i = 0$ for $i > N_r$. By applying the unitary transformation in terms of $\hat{U}$, $\hat{\mathcal{A}}' = \hat{U}^{\dagger}\hat{\mathcal{A}}\hat{U}$, $\hat{\mathcal{B}}' = \hat{U}^{\dagger}\hat{\mathcal{B}}\hat{U}$ and $\phi_1'^l = \hat{U}^{\dagger}\phi_1^l$ might have the forms

$$\hat{\mathcal{A}}' = \begin{bmatrix} \hat{\mathcal{A}}'_{11} & \hat{\mathcal{A}}'_{12} \\ \hat{\mathcal{A}}'_{21} & \hat{\mathcal{A}}'_{22} \end{bmatrix}, \quad \hat{\mathcal{B}}' = \begin{bmatrix} \hat{\mathcal{B}}'_{11} & \hat{\mathcal{B}}'_{12} \\ 0 & 0 \end{bmatrix}, \quad \phi_1'^l = \begin{bmatrix} \phi_{1,1}'^l \\ \phi_{1,2}'^l \end{bmatrix}, \tag{4.31}$$

where $\hat{\mathcal{A}}'_{11}$ and $\hat{\mathcal{B}}'_{11}$ are $N_r \times N_r$ matrices, $\hat{\mathcal{A}}'_{12}$ and $\hat{\mathcal{B}}'_{12}$ are $N_r \times (\widetilde{M} - N_r)$ matrices, $\hat{\mathcal{A}}'_{21}$ is $(\widetilde{M} - N_r) \times N_r$ matrix, $\hat{\mathcal{A}}'_{22}$ is $(\widetilde{M} - N_r) \times (\widetilde{M} - N_r)$ matrix, and $\phi_{1,1}'^l$ and $\phi_{1,2}'^l$ are N_r- and $(\widetilde{M} - N_r)$-dimensional vectors, respectively. The unitary transformation rewrites Eq. (4.14) to

$$[-\lambda^{-1}\hat{\mathcal{B}}'^{\dagger} + (\varepsilon\hat{I} - \hat{\mathcal{A}}') - \lambda\hat{\mathcal{B}}']\phi_1'^l = 0. \tag{4.32}$$

Because the number of the nontrivial solutions is $2N_r$, it is enough to calculate $\phi_{1,1}'^l$. Thus, eliminating $\phi_{1,2}'^l$ by substitution of Eq. (4.31) into Eq. (4.32) leads to the N_r-dimensional eigenvalue problem

$$[-\lambda^{-1}\hat{\mathcal{B}}_r^{\dagger} + (\varepsilon\hat{I} - \hat{\mathcal{A}}_r) - \lambda\hat{\mathcal{B}}_r]\phi_{1,1}'^l = 0, \tag{4.33}$$

where

$$\hat{\mathcal{A}}_r = \hat{\mathcal{A}}'_{11} + \hat{\mathcal{A}}'_{12}(\varepsilon\hat{I} - \hat{\mathcal{A}}'_{22})^{-1}\hat{\mathcal{A}}'_{21} + \hat{\mathcal{B}}'_{12}(\varepsilon\hat{I} - \hat{\mathcal{A}}'_{22})^{-1}(\hat{\mathcal{B}}'_{12})^{\dagger}, \tag{4.34}$$

$$\hat{\mathcal{B}}_r = \hat{\mathcal{B}}'_{11} + \hat{\mathcal{B}}'_{12}(\varepsilon\hat{I} - \hat{\mathcal{A}}'_{22})^{-1}\hat{\mathcal{A}}'_{21}. \tag{4.35}$$

Once $\phi_{1,1}'^l$ is obtained, rest terms are given by

$$\phi_{1,2}'^l = (\varepsilon\hat{I} - \hat{\mathcal{A}}'_{22})^{-1}(\lambda^{-1}(\hat{\mathcal{B}}'_{12})^{\dagger} + \hat{\mathcal{A}}'_{21})\phi_{1,1}'^l. \tag{4.36}$$

Finally, the generalized Bloch states of the original system are obtained by the inverse unitary transformation $\phi_1^l = \hat{U}\phi_1'^l$.

Next, I extend the above discussion for the surface Green's function. To avoid the complexity of the notation, I here put $z = \varepsilon$. Minor changes are needed for the general case (see Appendix B). Applying the unitary transformation, $\hat{\mathcal{A}}'_{\mathcal{S}} = \hat{U}^{\dagger}\hat{\mathcal{A}}_{\mathcal{S}}\hat{U}$, $\hat{\mathcal{A}}' = \hat{U}^{\dagger}\hat{\mathcal{A}}\hat{U}$, $\hat{\mathcal{B}}' = \hat{U}^{\dagger}\hat{\mathcal{B}}\hat{U}$, and $\hat{g}' = \hat{U}^{\dagger}\hat{g}\hat{U}$, they have the form

$$\hat{\mathcal{A}}'_{\mathcal{S}} = \begin{bmatrix} \hat{\mathcal{A}}'_{\mathcal{S},11} & \hat{\mathcal{A}}'_{\mathcal{S},12} \\ \hat{\mathcal{A}}'_{\mathcal{S},21} & \hat{\mathcal{A}}'_{\mathcal{S},22} \end{bmatrix}, \quad \hat{g}_{11}'^{R,00} = \begin{bmatrix} \hat{g}_{11,1}'^{R,00} \\ \hat{g}_{11,2}'^{R,00} \end{bmatrix}. \tag{4.37}$$

The unitary transformation rewrites Eq. (4.16) to

$$
\begin{bmatrix}
\varepsilon\hat{I} - \hat{\mathcal{A}}'_{\mathcal{S}} & -\hat{\mathcal{B}}' & & 0 \\
-\hat{\mathcal{B}}'^{\dagger} & \varepsilon\hat{I} - \hat{\mathcal{A}}' & -\hat{\mathcal{B}}' & \\
& -\hat{\mathcal{B}}'^{\dagger} & \varepsilon\hat{I} - \hat{\mathcal{A}}' & \ddots \\
0 & & \ddots & \ddots
\end{bmatrix}
\begin{bmatrix}
\hat{g}'^{R,00}_{11} \\
\hat{g}'^{R,10}_{11} \\
\hat{g}'^{R,20}_{11} \\
\vdots
\end{bmatrix}
=
\begin{bmatrix}
\hat{I} \\
0 \\
0 \\
\vdots
\end{bmatrix},
\tag{4.38}
$$

From Eqs. (4.37) and (4.38), eliminating $\hat{g}'^{i0}_{1,2}$ yields

$$
\begin{bmatrix}
\varepsilon\hat{I} - \hat{\mathcal{A}}_{S,r} & -\hat{\mathcal{B}}_r & & 0 \\
-\hat{\mathcal{B}}^{\dagger}_r & \varepsilon\hat{I} - \hat{\mathcal{A}}_r & -\hat{\mathcal{B}}_r & \\
& -\hat{\mathcal{B}}^{\dagger}_r & \varepsilon\hat{I} - \hat{\mathcal{A}}_r & \ddots \\
0 & & \ddots & \ddots
\end{bmatrix}
\begin{bmatrix}
\hat{g}'^{R,00}_{11,1} \\
\hat{g}'^{R,10}_{11,1} \\
\hat{g}'^{R,20}_{11,1} \\
\vdots
\end{bmatrix}
=
\begin{bmatrix}
\hat{I}_r \\
0 \\
0 \\
\vdots
\end{bmatrix},
\tag{4.39}
$$

where

$$
\hat{\mathcal{A}}_{S,r} = \hat{\mathcal{A}}'_{S,11} + \hat{\mathcal{A}}'_{S,12}(\varepsilon\hat{I} - \hat{\mathcal{A}}'_{S,22})^{-1}\hat{\mathcal{A}}'_{S,21} + \hat{\mathcal{B}}'_{12}(\varepsilon\hat{I} - \hat{\mathcal{A}}'_{22})^{-1}\hat{\mathcal{B}}'^{\dagger}_{12},
\tag{4.40}
$$

$$
\hat{I}_r = \hat{I}_1 + \hat{\mathcal{A}}'_{S,12}(\varepsilon\hat{I} - \hat{\mathcal{A}}'_{S,22})^{-1}\hat{I}_2.
\tag{4.41}
$$

Here, I split the identity matrix $\hat{I}$ into

$$
\hat{I} =
\begin{bmatrix}
\hat{I}_{N_r\times N_r} & 0 \\
0 & \hat{I}_{\widetilde{M}-N_r\times\widetilde{M}-N_r}
\end{bmatrix}
=
\begin{bmatrix}
\hat{I}_1 \\
\hat{I}_2
\end{bmatrix},
\tag{4.42}
$$

where $\hat{I}_{N_r\times N_r}$ and $\hat{I}_{\widetilde{M}-N_r\times\widetilde{M}-N_r}$ are identity matrices with N_r and $\widetilde{M} - N_r$ dimensions, respectively. The outermost element of the surface Green's function $\hat{g}'^{00}_{11,1}$ is obtained by

$$
\hat{g}'^{00}_{11,1} = [(\varepsilon\hat{I} - \hat{\mathcal{A}}_{S,r}) - \hat{\mathcal{B}}_r(\varepsilon\hat{I} - \hat{\mathcal{H}}^{bulk}_{R,r})^{-1}\hat{\mathcal{B}}^{\dagger}_r]^{-1}\hat{I}_r,
\tag{4.43}
$$

where $\hat{\mathcal{H}}^{bulk}_{R,r}$ is defined by

$$
\hat{\mathcal{H}}^{bulk}_{R,r} =
\begin{bmatrix}
\hat{\mathcal{A}}_r & \hat{\mathcal{B}}_r & & 0 \\
\hat{\mathcal{B}}^{\dagger}_r & \hat{\mathcal{A}}_r & \hat{\mathcal{B}}_r & \\
& \hat{\mathcal{B}}^{\dagger}_r & \hat{\mathcal{A}}_r & \ddots \\
0 & & \ddots & \ddots
\end{bmatrix}.
\tag{4.44}
$$

Because $\hat{\mathcal{H}}_{R,r}^{bulk}$ preserves the translational symmetry for transport direction, it might be possible to compute the self-energy-like matrix. Again, for example, recursive Green's function method (self-consistent equation for self-energy matrix) enables us to compute

$$\tilde{\Sigma}_{R,r}^{(n+1)} = \hat{\mathcal{B}}_r[\varepsilon\hat{I} - \hat{\mathcal{A}}_r - \tilde{\Sigma}_{R,r}^{(n)}]^{-1}\hat{\mathcal{B}}_r^{\dagger}, \tag{4.45}$$

where n is the number of iteration. Once $\hat{g}_{11,1}^{\prime R,00}$ is obtained, $\hat{g}_{11,2}^{\prime R,00}$ can be computed by

$$\hat{g}_{11,2}^{\prime R,00} = (\varepsilon\hat{I} - \hat{\mathcal{A}}_{S,22}')^{-1}(\hat{\mathcal{A}}_{S,21}'\hat{g}_{11,1}^{\prime R,00} + \hat{I}_2), \tag{4.46}$$

and the inverse unitary transformation leads to $\hat{g}_{11}^{R,00} = \hat{U}\hat{g}_{11}^{\prime R,00}\hat{U}^{\dagger}$.

4.4   Left surface Green's function

The formula of the self-energy matrix for the left electrode can be derived in the analogous manner as for the right electrode, that is, $\hat{g}_{m,m}^{L,00}$ can be obtained by partitioning the Hamiltonian matrix into $(m-1)\widetilde{M}$ and $\widetilde{M}$ dimensional matrices. In this case, the contracted matrices $\hat{\mathcal{A}}$ and $\hat{\mathcal{B}}$ must be recalculated because the partitioning position is different between left and right electrodes. However, if the left and right electrodes are identical, one can skip the computation of contracted matrices. In this section, I show that the self-energy matrix for the left electrode can be computed from $\hat{g}_{1,m}^{L,00}$ using the relationship between $\hat{g}_{m,m}^{L,00}$. The 0-th column of the left surface Green's function is

$$\begin{bmatrix} \ddots & \ddots & & 0 \\ \ddots & z\hat{I} - \hat{A} & -\hat{B} & \\ & -\hat{B}^{\dagger} & z\hat{I} - \hat{A} & -\hat{B} \\ 0 & & -\hat{B}^{\dagger} & z\hat{I} - \hat{A} \end{bmatrix} \begin{bmatrix} \vdots \\ \hat{g}_{20}^{L} \\ \hat{g}_{10}^{L} \\ \hat{g}_{00}^{L} \end{bmatrix} = \begin{bmatrix} \vdots \\ 0 \\ 0 \\ \hat{I} \end{bmatrix}. \tag{4.47}$$

Partitioning of Eq. (4.47) leads to

$$\begin{bmatrix} \ddots & \ddots & & 0 \\ \ddots & z\hat{I} - \hat{\mathcal{A}} & -\hat{\mathcal{B}} & \\ & -\hat{\mathcal{B}}^{T} & z\hat{I} - \hat{\mathcal{A}} & -\hat{\mathcal{B}} \\ 0 & & -\hat{\mathcal{B}}^{T} & z\hat{I} - \hat{\mathcal{A}} \end{bmatrix} \begin{bmatrix} \vdots \\ \hat{g}_{1,m}^{L,20} \\ \hat{g}_{1,m}^{L,10} \\ \hat{g}_{1,m}^{L,00} \end{bmatrix} = \begin{bmatrix} \vdots \\ 0 \\ 0 \\ \hat{J} \end{bmatrix}, \tag{4.48}$$

where

$$\hat{J} = \hat{\beta}_1(z\hat{I} - \hat{\alpha}_2)^{-1}\hat{I}'. \tag{4.49}$$

Here,

$$\hat{I}' = \begin{bmatrix} 0 \\ \hat{I}_{\widetilde{M}\times\widetilde{M}} \end{bmatrix}, \tag{4.50}$$

where $\hat{I}_{\widetilde{M} \times \widetilde{M}}$ is the $\widetilde{M}$ dimensional identity matrix. Thus $\hat{g}_{1,m}^{L,00}$ is

$$\hat{g}_{1,m}^{L,00} = [z\hat{I} - \hat{\mathcal{A}} - \hat{\mathcal{B}}^T(z\hat{I} - \hat{\mathcal{H}}_L^{bulk})^{-1}\hat{\mathcal{B}}]^{-1}\hat{J}, \tag{4.51}$$

where

$$\hat{\mathcal{H}}_L^{bulk} = \begin{bmatrix} \ddots & \ddots & & 0 \\ \ddots & \hat{\mathcal{A}} & \hat{\mathcal{B}} & \\ & \hat{\mathcal{B}}^T & \hat{\mathcal{A}} & \hat{\mathcal{B}} \\ 0 & & \hat{\mathcal{B}}^T & \hat{\mathcal{A}} \end{bmatrix}. \tag{4.52}$$

The self-energy-like matrix for left electrode, $\tilde{\Sigma}_L = \hat{\mathcal{B}}^T(z\hat{I} - \hat{\mathcal{H}}_L^{bulk})^{-1}\hat{\mathcal{B}}$, can be computed by the recursive Green's function method,

$$\tilde{\Sigma}_L^{(n+1)} = \hat{\mathcal{B}}^T[z\hat{I} - \hat{\mathcal{A}} - \tilde{\Sigma}_L^{(n)}]^{-1}\hat{\mathcal{B}}. \tag{4.53}$$

After obtaining $\hat{g}_{1,m}^{L,00}$, $\hat{g}_{m,m}^{L,00}$ is given by

$$\hat{g}_{m,m}^{L,00} = \hat{D}_{m,2}\hat{B}_1^\dagger \hat{g}_{1,m}^{L,00} + \hat{D}_{m,m}. \tag{4.54}$$

Finally substituting Eq. (4.54) into (4.2), self-energy matrices for left electrode can be calculated.

4.5   Matrix inversion

As in Eqs. (4.25)-(4.27), it is necessary to compute $\hat{D}_{2,2}, \hat{D}_{2,m}$ and $\hat{D}_{m,m}$ that are $\widetilde{M} \times \widetilde{M}$ matrices. If $\hat{\beta}_1$ and $\hat{\beta}_2$ are singular, it is possible to reduce the computational cost of matrix inversion. First I perform the singular value decomposition for $\hat{B}_0$ and $\hat{B}_1$. Because the rank of $\hat{B}_0$ is N_r, $\hat{B}_0$ can be truncated as

$$\begin{aligned} \hat{B}_0 &= \hat{U}^{B_0}\hat{S}^{B_0}(\hat{V}^{B_0})^\dagger \\ &= [\hat{U}_1^{B_0}, \hat{U}_2^{B_0}] \begin{bmatrix} \hat{S}_1^{B_0} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} (\hat{V}_1^{B_0})^\dagger \\ (\hat{V}_2^{B_0})^\dagger \end{bmatrix} \\ &= \hat{U}_1^{B_0}\hat{S}_1^{B_0}(\hat{V}_1^{B_0})^\dagger, \end{aligned} \tag{4.55}$$

where $\hat{U}_1^{B_0}$ and $\hat{V}_1^{B_0}$ are $\widetilde{M} \times N_r$ matrices and $\hat{S}_1^{B_0}$ is $N_r \times N_r$ matrix. Then, $\hat{\beta}_2$ might have the following form

$$\hat{\beta}_2 = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \hat{B}_0 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \hat{U}_1^{B_0} \end{bmatrix} \hat{S}_1^{B_0}(\hat{V}_1^{B_0})^\dagger. \tag{4.56}$$

In the exactly same manner, $\hat{\beta}_1^\dagger$ might have the following form

$$\hat{\beta}_1^\dagger = \begin{bmatrix} \hat{B}_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \hat{V}_1^{B1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \hat{S}_1^{B1} (\hat{U}_1^{B1})^\dagger. \tag{4.57}$$

Thus, to construct $\hat{\mathcal{A}}$ and $\hat{\mathcal{B}}$, it is enough to solve the following linear equation

$$[z\hat{I} - \hat{\alpha}_2]\hat{X} = \begin{bmatrix} \hat{V}_1^{B1} & 0 \\ 0 & \vdots \\ \vdots & 0 \\ 0 & \hat{U}_1^{B0} \end{bmatrix}. \tag{4.58}$$

Note that the number of right-hand side of Eq. (4.58) is not $2\widetilde{M}$ but $2N_r$.

4.6 Optimization of OBM method

In the previous sections, I have derived the contracted quadratic eigenvalue problem Eq. (4.14) based on the partitioning technique proposed by Khomyakov *et al.* [68], and discuss the advantage of solving quadratic eigenvalue problem rather than solving the generalized eigenvalue problem with the twice size. An alternative method to calculate the generalized Bloch states on the RSFD scheme is the overbridging boundary-matching (OBM) method [67, 73, 76]. The OBM method also calculates the generalized Bloch states by solving the generalized eigenvalue problem. In this section, I show that the generalized eigenvalue problem in the OBM method can be transformed into the similar quadratic eigenvalue problem with high symmetric matrix form. The OBM method computes the generalized Bloch states by solving

$$\hat{\Pi}_1 \begin{bmatrix} \phi_m^{l-1} \\ \phi_1^{l+1} \end{bmatrix} = \lambda \hat{\Pi}_2 \begin{bmatrix} \phi_m^{l-1} \\ \phi_1^{l+1} \end{bmatrix}, \tag{4.59}$$

where

$$\hat{\Pi}_1 = \begin{bmatrix} \hat{\mathcal{G}}_{m,1}^l \hat{B}_0^\dagger & \hat{\mathcal{G}}_{m,m}^l \hat{B}_0 \\ 0 & \hat{I} \end{bmatrix}, \quad \hat{\Pi}_2 = \begin{bmatrix} \hat{I} & 0 \\ \hat{\mathcal{G}}_{1,1}^l \hat{B}_0^\dagger & \hat{\mathcal{G}}_{1,m}^l \hat{B}_0 \end{bmatrix}. \tag{4.60}$$

Here, $\hat{\mathcal{G}}^l(\varepsilon) = [\varepsilon\hat{I} - \hat{A}]^{-1}$. Eliminating ϕ_m^{l-1} from Eq. (4.59) leads to

$$\lambda^{-1}\hat{B}_0^{-\dagger}(\hat{\mathcal{G}}_{1,1}^l)^{-1}\hat{\mathcal{G}}_{1,m}^l\hat{B}_0\phi_1^{l+1} + [\hat{\mathcal{G}}_{m,m}^l\hat{B}_0 - \hat{\mathcal{G}}_{m,1}^l(\hat{\mathcal{G}}_{1,1}^l)^{-1}\hat{\mathcal{G}}_{1,m}^l\hat{B}_0 - \hat{B}_0^{-\dagger}(\hat{\mathcal{G}}_{1,1}^l)^{-1}]\phi_1^{l+1} + \lambda\hat{\mathcal{G}}_{m,1}^l(\hat{\mathcal{G}}_{1,1}^l)^{-1}\phi_1^{l+1} = 0. \tag{4.61}$$

To recover the symmetry, multiplying $\hat{B}_0^\dagger$ from the left-hand side yields

$$\lambda^{-1}(\hat{\mathcal{G}}_{1,1}^l)^{-1}\hat{\mathcal{G}}_{1,m}^l\hat{B}_0\phi_1^{l+1}+[\hat{B}_0^\dagger\hat{\mathcal{G}}_{m,m}^l\hat{B}_0-\hat{B}_0^\dagger\hat{\mathcal{G}}_{m,1}^l(\hat{\mathcal{G}}_{1,1}^l)^{-1}\hat{\mathcal{G}}_{1,m}^l\hat{B}_0-(\hat{\mathcal{G}}_{1,1}^l)^{-1}]\phi_1^{l+1}+\lambda\hat{B}_0^\dagger\hat{\mathcal{G}}_{m,1}^l(\hat{\mathcal{G}}_{1,1}^l)^{-1}\phi_1^{l+1} = 0.$$
$$(4.62)$$

If ε is a real number then $\hat{\mathcal{G}}^l = (\hat{\mathcal{G}}^l)^\dagger$, that is, $\hat{\mathcal{G}}_{1,1}^l = (\hat{\mathcal{G}}_{1,1}^l)^\dagger$, $\hat{\mathcal{G}}_{1,m}^l = (\hat{\mathcal{G}}_{m,1}^l)^\dagger$, $\hat{\mathcal{G}}_{m,1}^l = (\hat{\mathcal{G}}_{1,m}^l)^\dagger$, and $\hat{\mathcal{G}}_{m,m}^l = (\hat{\mathcal{G}}_{m,m}^l)^\dagger$. Therefore, Eq. (4.62) can be rewritten as

$$[\lambda^{-1}\hat{\mathcal{K}}_1^\dagger + \hat{\mathcal{K}}_0 + \lambda\hat{\mathcal{K}}_1]\phi_1^l = 0, \tag{4.63}$$

where

$$\hat{\mathcal{K}}_0 = \hat{B}_0^\dagger\hat{\mathcal{G}}_{m,m}^l\hat{B}_0 - \hat{B}_0^\dagger\hat{\mathcal{G}}_{m,1}^l(\hat{\mathcal{G}}_{1,1}^l)^{-1}\hat{\mathcal{G}}_{1,m}^l\hat{B}_0 - (\hat{\mathcal{G}}_{1,1}^l)^{-1}, \tag{4.64}$$

$$\hat{\mathcal{K}}_1 = \hat{B}_0^\dagger\hat{\mathcal{G}}_{m,1}^l(\hat{\mathcal{G}}_{1,1}^l)^{-1}. \tag{4.65}$$

Since $\hat{\mathcal{K}}_0 = \hat{\mathcal{K}}_0^\dagger$, the similar relationship of Eq. (4.29) is established. After solving Eq. (4.63), one can obtain the ϕ_m^l by

$$\phi_m^l = (\hat{\mathcal{G}}_{1,1}^l\hat{B}_0^\dagger)^{-1}[\hat{I} - \lambda\hat{\mathcal{G}}_{1,m}^l\hat{B}_0]\phi_1^l. \tag{4.66}$$

4.7   Reduction of matrix size

Even if $\hat{B}_0$ and $\hat{B}_1$ are regular matrices, $\hat{\mathcal{B}}$ is close to singular or more generally its condition number $\kappa(\hat{\mathcal{B}})$ becomes extremely large. To see this, numerical calculations of Au atomic chain are presented here as a simple example. I employ the central finite-difference approximation ($N = 1$ in Ref. [75]) for the Laplacian operator and local pseudopotential by Troullier-Martins [24]. Local density approximation is used as an exchange-correlation functional [12]. The size of the supercell is $22.68 \times 22.68 \times 10.96$ bohr^3 and two Au atoms are contained in the supercell. The number of grid points is $56 \times 56 \times 28$. The transport direction is set along z direction. Only gamma point is sampled in the two-dimensional Brillouin zone. In this case,

$$\hat{B}_0 = \hat{B}_1 = -\frac{1}{2h_z^2}\hat{I}, \tag{4.67}$$

where h_z is the grid spacing in z direction. From Eqs. (4.13) and (4.27), $\hat{\mathcal{B}}$ is reduced to

$$\hat{\mathcal{B}} = \frac{1}{4h_z^4}\hat{D}_{2,m}, \tag{4.68}$$

which indicates that $\hat{\mathcal{B}}$ seems to be regular because $(\varepsilon\hat{I} - \hat{\alpha}_2)$ is not singular.

In order to obtain $\kappa(\hat{\mathcal{B}})$, I first perform the singular value decomposition of $\hat{\mathcal{B}}$. Figure 4.1 shows the singular values of $\hat{\mathcal{B}}$. Note that the singular values are divided by the largest singular value $s_{\max}$, and input energy is set as the Fermi energy. $\hat{D}_{2,m}$ is computed by CG method and the convergence criteria is set by 10^{-15}. If the smallest singular value is denoted as $s_{\min}$, the condition number is defined by $\kappa(\hat{\mathcal{B}}) = s_{\max}/s_{\min}$. Then, in this case, $\kappa(\hat{\mathcal{B}}) = 6.9 \times 10^{22}$. Because the numerical error due to the round-off error is proportional to the condition number, the precision of the significand

is not guaranteed when $\kappa(\hat{\mathcal{B}}) \geq 10^{16}$ and double-precision floating-point operation is used. Although a quadruple-precision floating-point operation is effective to treat such an ill-conditioned problem, a significant increase of computational cost will be unacceptable in practical calculations. Instead, I introduce the tolerance parameter of singular value decomposition δ_{SVD} and replace the singular values less than $s_{\max}\delta_{\mathrm{SVD}}$ to 0. Then, by eliminating the degrees of null space, one may obtain the contracted matrix $\hat{\mathcal{B}}_r$ with $\kappa(\hat{\mathcal{B}}_r) = \delta_{\mathrm{SVD}}^{-1}$ $(\ll \kappa(\hat{\mathcal{B}}))$. In this section, I perform a detailed analysis of numerical error of self-energy matrix arising from the singular value decomposition.
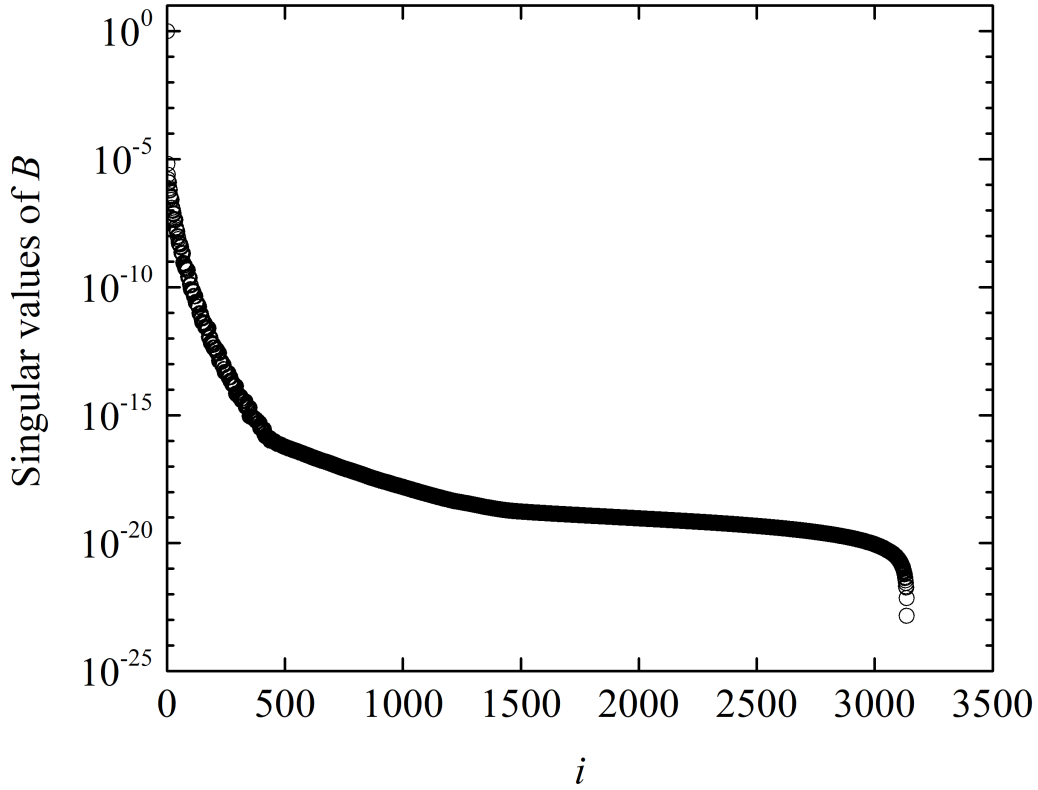
Figure 4.1: Singular values of $\hat{\mathcal{B}}$ normalized by the largest singular value $s_{\max}$.

4.7.1 Error analysis of generalized Bloch states

In order to see the effect of introducing δ_{SVD} on the accuracy of generalized Bloch states, I have calculated the residual 2-norms $||[-\lambda^{-1}\hat{\mathcal{B}}^{\dagger} + (\varepsilon\hat{I} - \hat{A}) - \lambda\hat{\mathcal{B}}]\phi_1^l||_2$ for a set of δ_{SVD}. To obtain the solutions of quadratic eigenvalue problems (4.14) and (4.33), they are transformed into the generalized

eigenvalue problem and solved by QZ method. For example, Eq. (4.14) is transformed by

$$
\begin{bmatrix} \varepsilon \hat{I} - \hat{\mathcal{A}} & -\hat{\mathcal{B}}^\dagger \\ \hat{I} & 0 \end{bmatrix} \begin{bmatrix} \phi_1^l \\ \phi_1^{l-1} \end{bmatrix} = \lambda \begin{bmatrix} \hat{\mathcal{B}} & 0 \\ 0 & \hat{I} \end{bmatrix} \begin{bmatrix} \phi_1^l \\ \phi_1^{l-1} \end{bmatrix}.
\tag{4.69}
$$

It is experimentally known that the accuracy of the solutions with $|\lambda| \ll 1$ is worse than that with $|\lambda| \gg 1$. According to Ref. [55], in order to improve the accuracy of solutions with $|\lambda| \ll 1$, solutions with $|\lambda| \ll 1$ are obtained by solving the equivalent equation

$$
\begin{bmatrix} \varepsilon \hat{I} - \hat{\mathcal{A}} & -\hat{\mathcal{B}} \\ \hat{I} & 0 \end{bmatrix} \begin{bmatrix} \phi_1^l \\ \phi_1^{l+1} \end{bmatrix} = \lambda^{-1} \begin{bmatrix} \hat{\mathcal{B}}^\dagger & 0 \\ 0 & \hat{I} \end{bmatrix} \begin{bmatrix} \phi_1^l \\ \phi_1^{l+1} \end{bmatrix}.
\tag{4.70}
$$

Figure 4.2 shows the residual norms for a set of δ_{SVD} in the range of $(0, 10^{-16}, 10^{-14}, 10^{-12}, 10^{-10})$. In case that $\delta_{\mathrm{SVD}} = 0$, I directly solve Eq. (4.14) without unitary transformation. Note that the solutions whose eigenvalues become not a number (NaN) or infinity are excluded from plots. In all cases, the accuracy of the solutions become wrong as $|\lambda|$ becomes greater or lesser than 1. This common feature may come from the numerical error arising from the QZ method. In addition, it is interesting to see that in spite of omitting the small singular values, better or same level accuracy is obtained when $\delta_{\mathrm{SVD}} = 10^{-16}$ or 10^{-14}. This results indicates that reducing the condition number enhances the numerical stability of the calculation. Furthermore, I plot the distribution of eigenvalues in Fig. 4.3 to investigate the relationship between the generalized Bloch states and δ_{SVD}. Figure 4.3 clearly shows that eigenvalues with $|\lambda| \ll 1$ and $|\lambda| \gg 1$ are omitted by increasing the values of δ_{SVD}. This indicates that the vector space with small singular values of $\hat{\mathcal{B}}$ corresponds to the rapidly decaying evanescent states with $|\lambda| \ll 1$ and $|\lambda| \gg 1$. In other words, information of important generalized Bloch states is handed over in the contracted system whose dimension is much smaller than that of original system.

It is useful to discuss why important generalized Bloch states are expressed by the vectors with singular values s_i which satisfy $s_i \geq s_{\max}\delta_{\mathrm{SVD}}$. Because $\hat{\alpha}_2$ is made by truncating the bulk Hamiltonian, $(\varepsilon \hat{I} - \hat{\alpha}_2)^{-1}$ can be regarded as the Green's function under the isolated boundary condition for z direction and periodic boundary condition for xy direction. In free electron approximation, Green's function satisfies [2]

$$
(\varepsilon + \nabla^2)G(\boldsymbol{r}, \boldsymbol{r}') = \delta(\boldsymbol{r}, \boldsymbol{r}').
\tag{4.71}
$$

In the 2D Laue representation, it becomes

$$
(\varepsilon_z + \frac{d^2}{dz^2})G(\varepsilon_z, z, z') = \delta(z - z'),
\tag{4.72}
$$

where $\varepsilon_z = \varepsilon - |\boldsymbol{G}_{||}|^2$. Note that the boundary condition is imposed for $G(\varepsilon_z, z, z')$ such that

$$
G(\varepsilon_z, z_1, z') = G(\varepsilon_z, z_{m+1}, z') = 0.
\tag{4.73}
$$

---

[2]Rydberg atomic unit is used to simplify the notation.
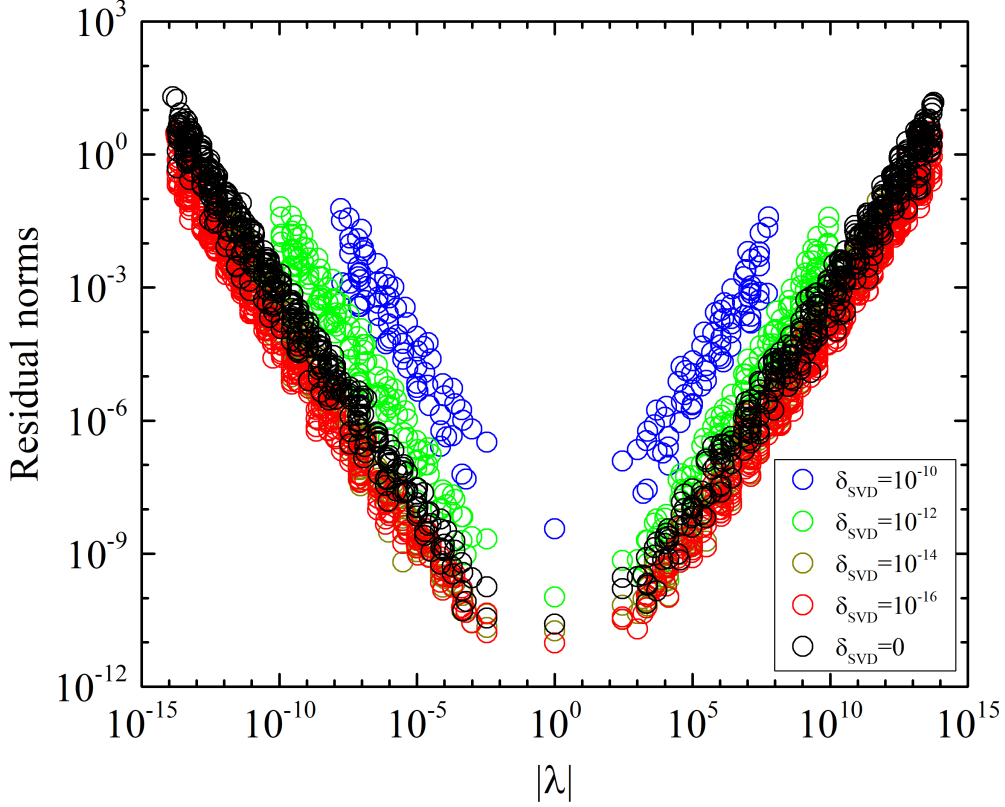
Figure 4.2: Residual 2-norms $||[-\lambda^{-1}\hat{\mathcal{B}}^\dagger + (\varepsilon\hat{I} - \hat{\mathcal{A}}) - \lambda\hat{\mathcal{B}}]\phi_1^l||_2$ for a set of δ_{SVD} in the range of $(0, 10^{-16}, 10^{-14}, 10^{-12}, 10^{-10})$. Note that the solutions are normalized by $||\phi_1^l||_2 = 1$.

For $\varepsilon_z < 0$, the analytic form of the Green's function is obtained as below [64]

$$G(\varepsilon_z, z, z') = \frac{e^{-\kappa_z|z-z'|} + e^{-\kappa_z(2(z_{m+1}-z_1)-|z-z'|)} + e^{-\kappa_z(z+z'-2z_1)} - e^{-\kappa_z(2z_{m+1}-z-z')}}{2\kappa_z(e^{-2\kappa_z(z_{m+1}-z_0)} - 1)}, \qquad (4.74)$$

where $\kappa_z = \sqrt{|\boldsymbol{G}_{||}|^2 - \varepsilon}$. On the other hand, $\varepsilon_z > 0$, $G(\varepsilon_z, z, z')$ is given by

$$G(\varepsilon_z, z, z') = \frac{e^{ik_z|z-z'|} + e^{ik_z(2(z_{m+1}-z_1)-|z-z'|)} + e^{ik_z(z+z'-2z_1)} - e^{ik_z(2z_{m+1}-z-z')}}{2ik_z(e^{2ik_z(z_{m+1}-z_0)} - 1)}, \qquad (4.75)$$

where $k_z = \sqrt{\varepsilon - |\boldsymbol{G}_{||}|^2}$. Eqs. (4.74) and (4.75) apparently indicate that $G(\varepsilon_z, z, z')$ is constructed by evanescent states if kinetic energy $|\boldsymbol{G}_{||}|^2$ of xy direction is greater than input energy ε, vice versa, if $\varepsilon > |\boldsymbol{G}_{||}|^2$, $G(\varepsilon_z, z, z')$ is constructed by the propagating states. Remembering that $\hat{\mathcal{B}} \approx G(x, y, z_2, x', y', z_m)$, steep evanescent waves with large κ_z contribute little to $\hat{\mathcal{B}}$ and corresponding singular values will be much smaller than propagating states with k_z. This might be the reason that the information of the generalized Bloch states with $|\lambda| \ll 1$ or $|\lambda| \gg 1$ drops from $\hat{\mathcal{B}}$.

55
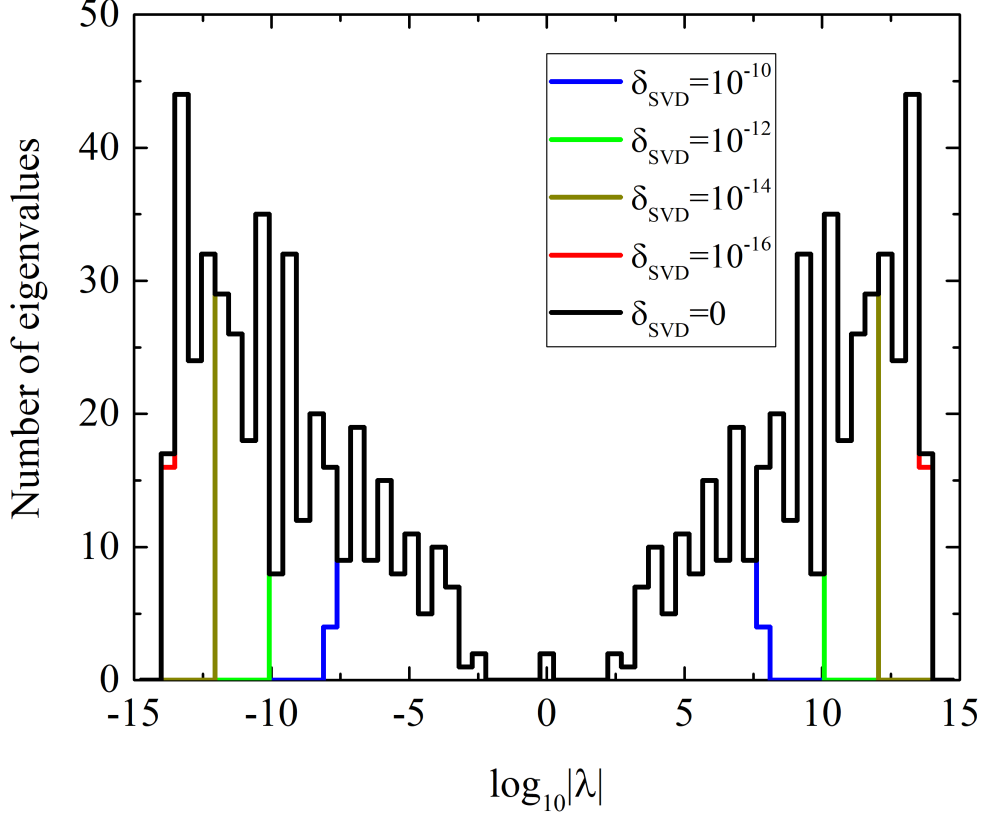
Figure 4.3: Distribution of eigenvalues of quadratic eigenvalue problem for a set of δ_{SVD} in the range of $(0, 10^{-16}, 10^{-14}, 10^{-12}, 10^{-10})$. The eigenvalues or residual norms becomes NaN are omitted from counts.

4.7.2　Computational cost

I here discuss the computational cost of the proposed algorithm. The proposed algorithm consists of three parts: (i) partitioning of Hamiltonian matrix in bulk system and setup contracted matrices $\hat{\mathcal{A}}$ and $\hat{\mathcal{B}}$, (ii) singular value decomposition of $\hat{\mathcal{B}}$ and setup the maximally contracted matrices $\hat{\mathcal{A}}_r$ and $\hat{\mathcal{B}}_r$, (iii) computing the self-energy matrices by quick iterative or semi-analytical method. The cost of partitioning scales as $\mathcal{O}((M - \widetilde{M})^2)$, which is the cost of matrix inversion by CG method. The cost of the singular value decomposition of $\hat{\mathcal{B}}$ scales as $\mathcal{O}(\widetilde{M}^3)$, while the cost of computing the self-energy matrices scales as $\mathcal{O}(N_r^3)$, which is the cost of solving quadratic eigenvalue problem (4.33) or inverting the matrix iteratively. Dominant part of the computational cost depends on the size of M, $\widetilde{M}$, and N_r. Table 4.1 shows the breakdown of the computational details to solve Eq. (4.14) and (4.33) by QZ method. Note that computational time to perform the singular value decomposition of $\hat{\mathcal{B}}$ and setup $\hat{\mathcal{A}}_r$ and $\hat{\mathcal{B}}_r$ are included in the calculation time. When $\delta_{\mathrm{SVD}} = 0$, the size of the matrix is not

Table 4.1: Breakdown of the method in Sec. 4.3. δ_{SVD} is the tolerance of the singular value decomposition. N_r is the number of singular values s_i such that $s_i \geq s_{\max}\delta_{\text{SVD}}$, i.e., rank of $\hat{B}_r$. $2N_r - N_{\text{NaN}}$ is the number of solutions except ones whose eigenvalues with NaN. Here, N_{NaN} is the number of solutions whose eigenvalues or residuals become NaN.

δ_{SVD}	N_r	$2N_r - N_{\text{NaN}}$	Elapsed time [sec.]
0	3136	820	2046.95
10^{-16}	443	818	18.92
10^{-14}	293	586	15.43
10^{-12}	185	370	14.33
10^{-10}	101	202	13.42

reduced, that is, $N_r = \widetilde{M}$. However, a vast majority of the solutions are numerically untractable due to a large condition number. The physical interpretation of such solutions are evanescent states with $|\lambda| \to \pm\infty$ and therefore they cause overflow. By increasing the value of δ_{SVD}, not only extremely fast decaying waves are removed but also a significant reduction of the computational cost is achieved. The reduction of computational cost is saturated when $\delta_{\text{SVD}} \geq 10^{-16}$ because N_r is sufficiently small and then cost of solving quadratic eigenvalue problem (4.33) is also small and instead the cost of singular value decomposition is dominant.

4.7.3 Error analysis of self-energy matrix

In order to analyze the effect of limiting the small singular values of $\hat{\mathcal{B}}$ on the accuracy of self-energy matrices, I define the matrix as

$$\hat{\Sigma}_{R,\text{SVD}}^{bulk} = \hat{\mathcal{B}}_{\text{SVD}}[\varepsilon\hat{I} - \hat{\mathcal{H}}_{R,SVD}^{bulk}]^{-1}\hat{\mathcal{B}}_{\text{SVD}}^{\dagger}, \tag{4.76}$$

where

$$\hat{\mathcal{B}}_{\text{SVD}} = \hat{U}\hat{S}_{\text{SVD}}\hat{V}^{\dagger}, \tag{4.77}$$

$$\hat{\mathcal{H}}_{R,SVD}^{bulk} = \begin{bmatrix} \hat{\mathcal{A}} & \hat{\mathcal{B}}_{\text{SVD}} & & 0 \\ \hat{\mathcal{B}}_{\text{SVD}}^{\dagger} & \hat{\mathcal{A}} & \hat{\mathcal{B}}_{\text{SVD}} & \\ & \hat{\mathcal{B}}_{\text{SVD}}^{\dagger} & \hat{\mathcal{A}} & \ddots \\ 0 & & \ddots & \ddots \end{bmatrix}. \tag{4.78}$$

Here, $\hat{S}_{\mathrm{SVD}}$ is the diagonal matrix whose diagonal elements $s_{n,\mathrm{SVD}}$ are $s_{n,\mathrm{SVD}} = s_n$ with $s_{n,\mathrm{SVD}} \geq s_{\max}\delta_{\mathrm{SVD}}$ other than that $s_{n,\mathrm{SVD}} = 0$. Taking the limit of $\delta_{\mathrm{SVD}} \to 0$, $\hat{\mathcal{B}}_{\mathrm{SVD}} \to \hat{\mathcal{B}}$ and $\hat{\Sigma}_{R,\mathrm{SVD}}^{bulk} \to \hat{\Sigma}_R^{bulk}$ (Eq. (4.22)). Therefore, it is possible to estimate the numerical accuracy by means of the error matrix

$$\hat{\Delta}_\Sigma = \hat{\Sigma}_R^{bulk} - \hat{\Sigma}_{R,\mathrm{SVD}}^{bulk}. \tag{4.79}$$

Using the sparsity of $\hat{S}_{\mathrm{SVD}}$, $\hat{\Sigma}_{R,\mathrm{SVD}}^{bulk}$ can be rewritten as

$$
\begin{aligned}
\hat{\Sigma}_{R,\mathrm{SVD}}^{bulk} &= \hat{U}\hat{S}_{\mathrm{SVD}}\hat{V}^\dagger[\varepsilon\hat{I} - \hat{\mathcal{H}}_{R,\mathrm{SVD}}^{bulk}]^{-1}\hat{V}\hat{S}_{\mathrm{SVD}}\hat{U}^\dagger \\
&= \hat{U}
\begin{bmatrix} \hat{s}_1 & 0 \\ 0 & 0 \end{bmatrix}
\begin{bmatrix} \hat{g}_{11} & \hat{g}_{12} \\ \hat{g}_{21} & \hat{g}_{22} \end{bmatrix}
\begin{bmatrix} \hat{s}_1 & 0 \\ 0 & 0 \end{bmatrix}
\hat{U}^\dagger \\
&= \hat{U}
\begin{bmatrix} \hat{s}_1\hat{g}_{11}\hat{s}_1 & 0 \\ 0 & 0 \end{bmatrix}
\hat{U}^\dagger,
\end{aligned}
\tag{4.80}
$$

where $\hat{s}_1$ is N_r-dimensional diagonal matrix whose non-zero elements are same as $\hat{S}$ and $\hat{g}_{11}, \hat{g}_{12}, \hat{g}_{21}$, and $\hat{g}_{22}$ are $N_r \times N_r$, $N_r \times (\widetilde{M} - N_r)$, $(\widetilde{M} - N_r) \times N_r$, and $(\widetilde{M} - N_r) \times (\widetilde{M} - N_r)$ matrices, respectively. Thus, the Frobenius norm of $\hat{\Delta}_\Sigma$ can be estimated by

$$
\begin{aligned}
||\hat{\Delta}_\Sigma|| &= ||\hat{U}^\dagger\hat{\Delta}_\Sigma\hat{U}|| \\
&\approx \left\| \begin{bmatrix} 0 & \hat{s}_1\hat{g}_{12}\hat{s}_2 \\ \hat{s}_2\hat{g}_{21}\hat{s}_1 & \hat{s}_2\hat{g}_{22}\hat{s}_2 \end{bmatrix} \right\| \\
&> \left\| \begin{bmatrix} 0 & s_{\max}\delta_{\mathrm{SVD}}\hat{g}_{12} \\ s_{\max}\delta_{\mathrm{SVD}}\hat{g}_{21} & \delta_{\mathrm{SVD}}^2\hat{g}_{22} \end{bmatrix} \right\| \\
&\approx s_{\max}\delta_{\mathrm{SVD}} \left\| \begin{bmatrix} 0 & \hat{g}_{12} \\ \hat{g}_{21} & 0 \end{bmatrix} \right\|,
\end{aligned}
\tag{4.81}
$$

which means that the error arising from the singular value decomposition is an order of $s_{\max}\delta_{\mathrm{SVD}}$.

Figure 4.4 shows the maximum norm $||\hat{\Delta}_\Sigma||_{\max}$ and mean absolute error $||\hat{\Delta}_\Sigma||_{\mathrm{mean}}$ as a function of δ_{SVD}. Both $||\hat{\Delta}_\Sigma||_{\max}$ and $||\hat{\Delta}_\Sigma||_{\mathrm{mean}}$ show the similar in shape to each other, which means that the error arises uniformly on all matrix elements. Although the errors are in accordance roughly with the estimated curve (dotted line in Fig. 4.4), the discrepancy is prominent at $\delta_{\mathrm{SVD}} = 10^{-6}$ and $\delta_{\mathrm{SVD}} = 10^{-16}$. This may come from the additional two errors not considered in Eq. (4.81). First error is the numerical error by the QZ method. Even when δ_{SVD} is forced to be zero, the accuracy of the calculated generalized Bloch states is at most 10^{-11} as shown in Fig. 4.2. Thus, it is reasonable to consider that the numerical error of the QZ method is not negligible when δ_{SVD} is sufficiently small.

Actually, error curves in Fig. 4.4 are saturated in the limit of $\delta_{\mathrm{SVD}} \to 0$. On the other hand, errors in the relatively large δ_{SVD} may come from that $\hat{\mathcal{H}}_R^{bulk}$ approximates $\hat{\mathcal{H}}_{R,\mathrm{SVD}}^{bulk}$. However, it is difficult to estimate the second error, but in the actual calculation, one can avoid it by monitoring the accuracy of the self-energy matrices by decreasing δ_{SVD}.
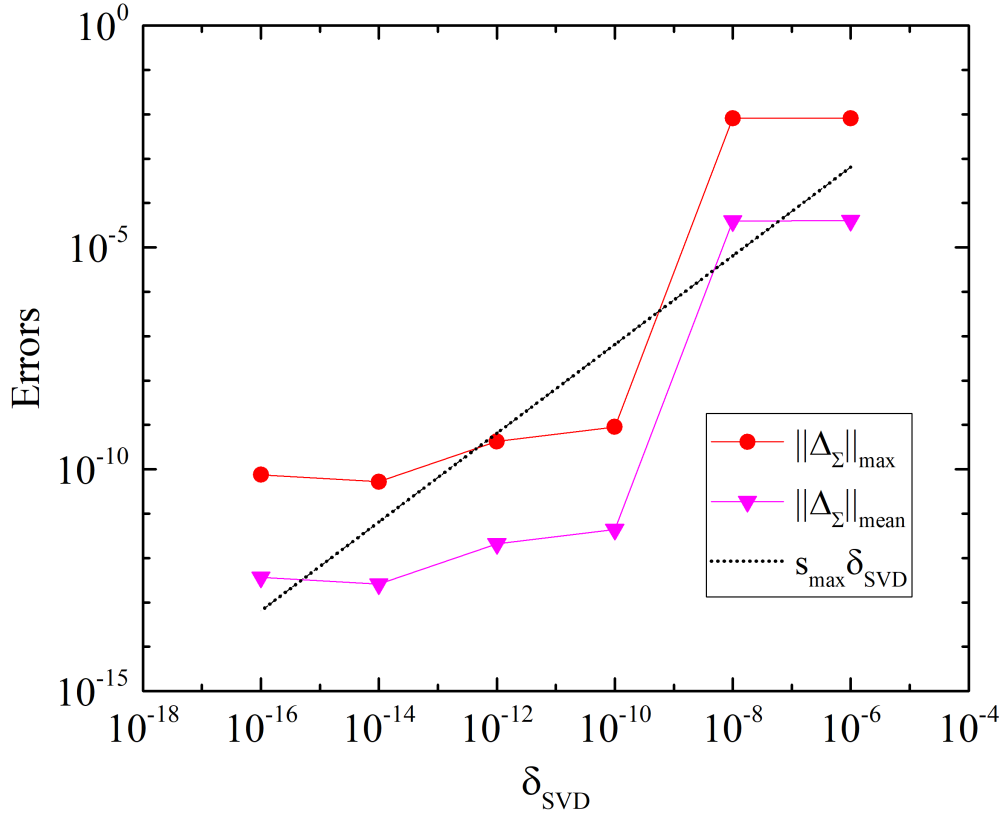
Figure 4.4: Errors in the calculation of self-energy matrix as a function of δ_{SVD}.

Chapter 5

Evaluation of Green's function on RSFD scheme

One of the expensive computational tasks of the NEGF method is inverting the Hamiltonian matrix for the transition region to obtain the retarded and non-equilibrium Green's functions. On the other hand, the WFM method, alternative to the NEGF method, can avoid any inversion of the Hamiltonian matrix by transforming Kohm-Sham equation under the open boundary condition into a set of linear equation (3.61). Since the real-space Hamiltonian matrix is very sparse and the self-energy matrices hold nonzero elements at the outermost principal layer on L0 and R0, the scattering wavefunction can be obtained efficiently by use of the iterative method such the bi-conjugate CG (BiCG) method. This achieves a great reduction both on computational cost and memory usage. However, it is very difficult to calculate the charge density by the WFM method due to the presence of the bound states. As mentioned in Sec. 3.4.3, the bound states are the solution of the highly nonlinear eigenvalue problem which is numerically difficult to solve. Furthermore, one does not know how many bound states exist *a priori*. Thus, it is well accepted that the self-consistent calculation which is the crucial step toward finite-bias simulation is numerically difficult except for the atomic chains where bound states do not exist. For this reason, I herein focus on the numerical implementation of the NEGF method based on RSFD scheme. The results and discussions in this chapter are based on my published papers [78, 79].

5.1   Charge density revisit

In Chap. 3, I have introduced the NEGF method for computing the non-equilibrium charge density; however, its implementation on the RSFD scheme is challenging both in computational cost and memory usage. This difficulty mainly comes from taking the inversion of $[z\hat{I}-\hat{H}_T-\hat{\Sigma}_L(z)-\hat{\Sigma}_R(z)]$. The simplest way is the direct inversion by Gaussian elimination method, but it is impractical when the size of Hamiltonian matrix for the transition region increases due to its cubic scaling of the computational cost. However, if the physical quantities can be evaluated by the block matrix element of the inverted matrix, computational cost and memory usage will be saved significantly. It is fortunately so, that

the charge density and transmission can be evaluated by not whole but the only several elements of the retarded Green's function in the transition region. I first present the efficient implementation to compute the charge density in the non-equilibrium state.

Because self-energy matrices have non-zero value only at the left and right matching planes, $\hat{\Gamma}_{L/R}$, the imaginary part of the self-energy matrix, has the form

$$\hat{\Gamma}_L = \begin{bmatrix} \hat{\Gamma}_{L0} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \quad \hat{\Gamma}_R = \begin{bmatrix} 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & \hat{\Gamma}_{R0} \end{bmatrix}. \tag{5.1}$$

Owing to the sparsity of $\hat{\Gamma}_{L/R}$, one may write

$$\hat{G}_T \hat{\Gamma}_L \hat{G}_T^\dagger = \begin{bmatrix} \hat{G}_{00} \\ \hat{G}_{10} \\ \vdots \\ \hat{G}_{N+1,0} \end{bmatrix} \times \hat{\Gamma}_{L0} \times \begin{bmatrix} \hat{G}_{00}^\dagger & \hat{G}_{10}^\dagger & \cdots & \hat{G}_{N+1,0}^\dagger \end{bmatrix}, \tag{5.2}$$

$$\hat{G}_T \hat{\Gamma}_R \hat{G}_T^\dagger = \begin{bmatrix} \hat{G}_{0,N+1} \\ \hat{G}_{1,N+1} \\ \vdots \\ \hat{G}_{N+1,N+1} \end{bmatrix} \times \hat{\Gamma}_{R0} \times \begin{bmatrix} \hat{G}_{0,N+1}^\dagger & \hat{G}_{1,N+1}^\dagger & \cdots & \hat{G}_{N+1,N+1}^\dagger \end{bmatrix}. \tag{5.3}$$

It is evident from Eqs. (3.19) and (3.100) that the non-equilibrium Green's function and the imaginary part of the retarded Green's function can be evaluated by the first and last columns of the retarded Green's function. To compute the columns of the retarded Green's function, standard iterative solvers such as BiCG method are not always efficient, since $[z\hat{I} - \hat{H} - \hat{\Sigma}_L(z) - \hat{\Sigma}_R(z)]$ is non-Hermitian, for which the slow convergence is frequently a serious problem. To overcome this issue, I present the efficient numerical technique to compute the retarded Green's function.

From Eq. (3.15), one can derive the Dyson equation

$$\hat{G}_T = \hat{\mathcal{G}}_T + \hat{\mathcal{G}}_T(\hat{\Sigma}_L + \hat{\Sigma}_R)\hat{G}_T, \tag{5.4}$$

where $\hat{\mathcal{G}}_T$ is the unperturbed Green's function defined as

$$\hat{\mathcal{G}}_T(\varepsilon) = [(\varepsilon + i\eta)\hat{I} - \hat{H}_T]^{-1}. \tag{5.5}$$

Focusing on the first and last columns of Eq. (5.4), the following equations are obtained

$$
\begin{bmatrix} \hat{G}_{00} \\ \hat{G}_{10} \\ \vdots \\ \hat{G}_{N+1,0} \end{bmatrix} = \begin{bmatrix} \hat{\mathcal{G}}_{00} \\ \hat{\mathcal{G}}_{10} \\ \vdots \\ \hat{\mathcal{G}}_{N+1,0} \end{bmatrix} + \begin{bmatrix} \hat{\mathcal{G}}_{00}\hat{\Sigma}_L\hat{G}_{00} \\ \hat{\mathcal{G}}_{10}\hat{\Sigma}_L\hat{G}_{00} \\ \vdots \\ \hat{\mathcal{G}}_{N+1,0}\hat{\Sigma}_L\hat{G}_{00} \end{bmatrix} + \begin{bmatrix} \hat{\mathcal{G}}_{0,N+1}\hat{\Sigma}_R\hat{G}_{N+1,0} \\ \hat{\mathcal{G}}_{1,N+1}\hat{\Sigma}_R\hat{G}_{N+1,0} \\ \vdots \\ \hat{\mathcal{G}}_{N+1,N+1}\hat{\Sigma}_R\hat{G}_{N+1,0} \end{bmatrix}, \tag{5.6}
$$

$$
\begin{bmatrix} \hat{G}_{0,N+1} \\ \hat{G}_{1,N+1} \\ \vdots \\ \hat{G}_{N+1,N+1} \end{bmatrix} = \begin{bmatrix} \hat{\mathcal{G}}_{0,N+1} \\ \hat{\mathcal{G}}_{1,N+1} \\ \vdots \\ \hat{\mathcal{G}}_{N+1,N+1} \end{bmatrix} + \begin{bmatrix} \hat{\mathcal{G}}_{00}\hat{\Sigma}_L\hat{G}_{0,N+1} \\ \hat{\mathcal{G}}_{10}\hat{\Sigma}_L\hat{G}_{0,N+1} \\ \vdots \\ \hat{\mathcal{G}}_{N+1,0}\hat{\Sigma}_L\hat{G}_{0,N+1} \end{bmatrix} + \begin{bmatrix} \hat{\mathcal{G}}_{0,N+1}\hat{\Sigma}_R\hat{G}_{N+1,N+1} \\ \hat{\mathcal{G}}_{1,N+1}\hat{\Sigma}_R\hat{G}_{N+1,N+1} \\ \vdots \\ \hat{\mathcal{G}}_{N+1,N+1}\hat{\Sigma}_R\hat{G}_{N+1,N+1} \end{bmatrix} \tag{5.7}
$$

From Eqs (5.6) and (5.7), corner matrix elements, $\hat{G}_{00}, \hat{G}_{0,N+1}, \hat{G}_{N+1,0}$ and $\hat{G}_{N+1,N+1}$, are obtained by solving linear equations [76]

$$
\begin{bmatrix} \hat{I} - \hat{\mathcal{G}}_{00}\hat{\Sigma}_L & -\hat{\mathcal{G}}_{0,N+1}\hat{\Sigma}_R \\ -\hat{\mathcal{G}}_{N+1,0}\hat{\Sigma}_L & \hat{I} - \hat{\mathcal{G}}_{N+1,N+1}\hat{\Sigma}_R \end{bmatrix} \begin{bmatrix} \hat{G}_{00} & \hat{G}_{0,N+1} \\ \hat{G}_{N+1,0} & \hat{G}_{N+1,N+1} \end{bmatrix} = \begin{bmatrix} \hat{\mathcal{G}}_{00} & \hat{\mathcal{G}}_{0,N+1} \\ \hat{\mathcal{G}}_{N+1,0} & \hat{\mathcal{G}}_{N+1,N+1} \end{bmatrix}. \tag{5.8}
$$

Once $\hat{G}_{00}, \hat{G}_{0,N+1}, \hat{G}_{N+1,0}$ and $\hat{G}_{N+1,N+1}$ are obtained, the rest matrix elements can be computed by

$$
\hat{G}_{m,0} = \hat{\mathcal{G}}_{m,0} + \hat{\mathcal{G}}_{m,0}\hat{\Sigma}_L\hat{G}_{00} + \hat{\mathcal{G}}_{m,N+1}\hat{\Sigma}_R\hat{G}_{N+1,0}, \tag{5.9}
$$

$$
\hat{G}_{m,N+1} = \hat{\mathcal{G}}_{m,N+1} + \hat{\mathcal{G}}_{m,0}\hat{\Sigma}_L\hat{G}_{0,N+1} + \hat{\mathcal{G}}_{m,N+1}\hat{\Sigma}_R\hat{G}_{N+1,N+1}. \tag{5.10}
$$

for $m = 1, ..., N$. Computation of Eq. (5.8) can be performed with relatively low cost. Therefore, the non-equilibrium charge density can be computed easily from the first and last columns of the unperturbed Green's function. Important point here is that the unperturbed Green's function can be calculated by making use of the conjugate-orthogonal CG (COCG) method with rapid convergence thanks to the complex symmetric property of $[\varepsilon\hat{I} - \hat{H}_T]$.

Strictly speaking, Eqs. (3.19) and (3.100) are established only when $\eta = 0$, i.e., the energy is on real axis. It means that the overall results based on Eqs. (3.19) and (3.100) are invalid for complex energy. In addition, calculating the density matrix by performing the integral of the Green's function on real axis is difficult. The reason is that the density of states diverges at the band edge where the group velocity becomes zero. This well-known problem can be overcome by the contour integration of the retarded Green's function on complex plane. In this case, from Dyson equation (5.4), the diagonal elements of the retarded Green's function at complex energy z is obtained by

$$
\hat{G}_{m,m} = \hat{\mathcal{G}}_{m,m} + \hat{\mathcal{G}}_{m,0}\hat{\Sigma}_L\hat{G}_{0,m} + \hat{\mathcal{G}}_{m,N+1}\hat{\Sigma}_R\hat{G}_{N+1,m}, \tag{5.11}
$$

for $m = 1, 2, ..., N$. $\hat{G}_{0,m}$ and $\hat{G}_{N+1,m}$ are determined by

$$
\begin{bmatrix}
\hat{I} - \hat{\mathcal{G}}_{00}\hat{\Sigma}_L & -\hat{\mathcal{G}}_{0,N+1}\hat{\Sigma}_R \\
-\hat{\mathcal{G}}_{N+1,0}\hat{\Sigma}_L & \hat{I} - \hat{\mathcal{G}}_{N+1,N+1}\hat{\Sigma}_R
\end{bmatrix}
\begin{bmatrix}
\hat{G}_{0,m} \\
\hat{G}_{N+1,m}
\end{bmatrix}
=
\begin{bmatrix}
\hat{\mathcal{G}}_{0,m} \\
\hat{\mathcal{G}}_{N+1,m}
\end{bmatrix}.
\tag{5.12}
$$

Note that it is not required to compute $\hat{\mathcal{G}}_{0,m}$ and $\hat{\mathcal{G}}_{N+1,m}$ because $\hat{\mathcal{G}}_{0,m} = \hat{\mathcal{G}}_{m,0}^T$ and $\hat{\mathcal{G}}_{N+1,m} = \hat{\mathcal{G}}_{m,N+1}^T$ for a real-symmetric Hamiltonian. In addition, one does not need the whole matrix element of $\hat{\mathcal{G}}_{m,m}$ but its diagonal elements to compute the charge density.

5.2   Transmission revisit

I next present the computational technique to obtain the transmission probability $T(\varepsilon)$ from Eq. (3.25) which indicates that matrices $\hat{\Sigma}_L, \hat{\Sigma}_R$, and $\hat{G}_T$ are required to determine $T(\varepsilon)$ at a given energy ε. Since $\hat{\Sigma}_L$ and $\hat{\Sigma}_R$ can be determined in an efficient manner in Chap. 4, I here discuss how to compute $\hat{G}_T$ efficiently. Due to the sparsity of the broadening matrices $\hat{\Gamma}_L$ and $\hat{\Gamma}_R$, the Fisher-Lee formula is simplified as below [76]:

$$
\begin{aligned}
T(\varepsilon) &= \mathrm{Tr}\Bigg\{\hat{G}_T(\varepsilon) \times
\begin{bmatrix}
\hat{\Gamma}_{L0} & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0
\end{bmatrix}
\times (\hat{G}_T(\varepsilon))^\dagger \times
\begin{bmatrix}
0 & & & 0 \\
& \ddots & & \vdots \\
& & 0 & 0 \\
0 & \cdots & 0 & \hat{\Gamma}_{R0}
\end{bmatrix}
\Bigg\} \\
&= \mathrm{Tr}[\hat{G}_{N+1,0}\hat{\Gamma}_{L0}\hat{G}_{N+1,0}^\dagger\hat{\Gamma}_{R0}],
\end{aligned}
\tag{5.13}
$$

which means that left-hand block matrix element of $\hat{G}_T$ has all information to determine $T(\varepsilon)$ when the couplings between electrodes via self-energy matrices are already known. From Eq. (5.8), $\hat{G}_{N+1,0}$ can be obtained easily by

$$
\hat{G}_{N+1,0} = [\hat{d} - \hat{c}\hat{a}^{-1}\hat{b}]^{-1}[\hat{\mathcal{G}}_{N+1,0} - \hat{c}\hat{a}^{-1}\hat{\mathcal{G}}_{0,0}],
\tag{5.14}
$$

where

$$
\begin{aligned}
\hat{a} &= \hat{I} - \hat{\mathcal{G}}_{00}\hat{\Sigma}_L, \tag{5.15} \\
\hat{b} &= -\hat{\mathcal{G}}_{0,N+1}\hat{\Sigma}_R, \tag{5.16} \\
\hat{c} &= -\hat{\mathcal{G}}_{N+1,0}\hat{\Sigma}_L, \tag{5.17} \\
\hat{d} &= \hat{I} - \hat{\mathcal{G}}_{N+1,N+1}\hat{\Sigma}_R. \tag{5.18}
\end{aligned}
$$

Thus, the transmission calculation requires the corner elements of the unperturbed Green's function, $\hat{\mathcal{G}}_{00}$, $\hat{\mathcal{G}}_{0,N+1}$, $\hat{\mathcal{G}}_{N+1,0}$, and $\hat{\mathcal{G}}_{N+1,N+1}$.

5.3 Shifted Krylov solvers

As shown in the previous two sections, the block matrix elements of the unperturbed Green's function is necessary to determine the physical quantities, e.g., charge density and transmission, via the relationship between the retarded Green's function. For large and sparse matrix such as real-space Hamiltonian matrix, the iterative solvers are suitable rather than the LU-factorization. Especially, the COCG method might be the best choice for our problem because $[z\hat{I} - \hat{H}_T]$ is very sparse and complex-symmetric when $\hat{H}_T$ is a real-symmetric matrix. Moreover, $[z_j\hat{I} - \hat{H}_T]$ is a shifted matrix for a given set of complex energy z_j $(j = 1, 2, ..., N_s)$, which is in contrast to $[z_j\hat{I} - \hat{H}_T - \hat{\Sigma}_L(z_j) - \hat{\Sigma}_R(z_j)]$ for the retarded Green's function due to the presence of the energy dependent self-energy matrices. In this section, I present the efficient numerical solver to compute the block matrix elements of the unperturbed Green's function.

5.3.1 Krylov solver for computing Green's function

Because the direct inversion of $[z\hat{I} - \hat{H}_T]$ is prohibitively expensive in many cases, it is efficient to compute the matrix elements of the unperturbed Green's function by solving a set of linear equations

$$\begin{aligned} \hat{\mathcal{G}}_{i,j}(z) &= \langle i | [z\hat{I} - \hat{H}_T]^{-1} | j \rangle \\ &= \langle i | \boldsymbol{x}_j(z) \rangle, \end{aligned} \tag{5.19}$$

where $|\boldsymbol{x}_j(z)\rangle$ is a solution vector of a linear equation,

$$[z\hat{I} - \hat{H}_T] | \boldsymbol{x}_j(z) \rangle = | j \rangle. \tag{5.20}$$

Here $|j\rangle$ indicates a site vector on the j-th principal layer in the transition region. Because $[z\hat{I} - \hat{H}_T]$ is a sparse and complex-symmetric matrix, the COCG method might be the best choice to solve the above equation. To begin with this section, I briefly explain the algorithmic details of the COCG method. Throughout this section, the linear algebraic expression will be used to avoid the misunderstanding.

5.3.2 COCG method

Let us consider a linear equation,

$$\hat{A}\boldsymbol{x} = \boldsymbol{b}, \tag{5.21}$$

where $\boldsymbol{x}, \boldsymbol{b} \in \mathbb{C}^M$, and $\hat{A} \in \mathbb{C}^{M \times M}$. Here, I assume that $\hat{A}$ is the complex symmetric matrix ($\hat{A} = \hat{A}^T \neq \hat{A}^\dagger$). In the COCG method, which is a family of the CG method, the approximated solution vectors are searched within the Krylov subspace. The n-th Krylov subspace is defined as

$$\mathcal{K}_n(\hat{A}, \boldsymbol{v}) = \mathrm{span}\{\boldsymbol{v}, \hat{A}\boldsymbol{v}, \hat{A}^2\boldsymbol{v}, ..., \hat{A}^{n-1}\boldsymbol{v}\}. \tag{5.22}$$

It is important to note that the whole space will be spanned by $\mathcal{K}_n(\hat{A}, \boldsymbol{v})$ when $n = M$. The correction scheme to the approximated solution vectors is derived such that the n-th residual vector, $\boldsymbol{r}_n = \boldsymbol{b} - \hat{A}\boldsymbol{x}_n$,

is conjugate orthogonal to the Krylov subspace. Under the initial condition $\boldsymbol{x}_0 = 0$, the conjugate orthogonality of the residual vector is expressed as below,

$$\boldsymbol{r}_n \in \mathcal{K}_{n+1}(\hat{A}, \boldsymbol{b}) \text{ and } \boldsymbol{r}_n \perp \mathcal{K}_n(\overline{\hat{A}}, \overline{\boldsymbol{b}}), \tag{5.23}$$

where the overline denotes the conjugate. Because the condition (5.23) is similar with the Ritz-Galerkin condition $\boldsymbol{r}_n \perp \mathcal{K}_n(\hat{A}, \boldsymbol{b})$, the COCG algorithm is quite similar with the well-known CG one. Specially, the difference appears only for construction of α_n and β_n. The COCG algorithm is written in Table 5.1.

Table 5.1: COCG algorithm for $\hat{A}\boldsymbol{x} = \boldsymbol{b}$, with a complex symmetric matrix $\hat{A} \in \mathbb{C}^{M \times M}, \boldsymbol{x}_n, \boldsymbol{p}_n, \boldsymbol{r}_n \in \mathbb{C}^M$, and $\alpha_n, \beta_n \in \mathbb{C}$. Note that the inner product is defined as $(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u}^T \boldsymbol{v}$.

1: Set $\boldsymbol{x}_0 = \boldsymbol{p}_{-1} = 0$, $\boldsymbol{r}_0 = \boldsymbol{b}$, $\alpha_{-1} = 1$, and $\beta_{-1} = 0$
2: For $n = 0, 1, 2, ...,$ until convergence do:
3:     $\alpha_n = (\boldsymbol{r}_n, \boldsymbol{r}_n)/(\boldsymbol{p}_n, \hat{A}\boldsymbol{p}_n)$
4:     $\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \alpha_n \boldsymbol{p}_n$
5:     $\boldsymbol{r}_{n+1} = \boldsymbol{r}_n - \alpha_n \hat{A}\boldsymbol{p}_n$
6:     $\beta_n = (\boldsymbol{r}_{n+1}, \boldsymbol{r}_{n+1})/(\boldsymbol{r}_n, \boldsymbol{r}_n)$
7:     $\boldsymbol{p}_{n+1} = \boldsymbol{r}_n + \beta_n \boldsymbol{p}_n$
8: End do

5.3.3   Shifted COCG method

Next, I consider to solve a series of linear equations whose matrix elements are connected with each other just by the scalar shift σ,

$$[\hat{A} + \sigma\hat{I}]\boldsymbol{x}(\sigma) = \boldsymbol{b}. \tag{5.24}$$

In the transport calculation, the number of σ is as many as $10^2 - 10^4$. If one apply the COCG method to Eq. (5.24), calculation cost increases linearly with the number of the shift, which results in that large-scale transport calculation is not feasible. However, the most-time consuming task, matrix-vector operation, is actually needed at the seed system ($\sigma = 0$). This trick is based on that the Krylov subspace is invariant against σ,

$$\mathcal{K}_n(\hat{A}, \boldsymbol{b}) = \mathcal{K}_n(\hat{A} + \sigma\hat{I}, \boldsymbol{b}), \tag{5.25}$$

which is easily confirmed from the fact that the power of $(\hat{A} + \sigma\hat{I})$ is decomposed into that of $\hat{A}$. Since the Krylov subspace is common among all shifts, residual vectors at arbitrary shifts and seed are collinear, that is,

$$\boldsymbol{r}_n(\sigma) = \frac{1}{\pi_n(\sigma)}\boldsymbol{r}_n, \tag{5.26}$$

where $\pi_n(\sigma)$ is the scalar function of σ. Eliminating $\boldsymbol{p}_n$ from fifth and seventh lines in Table 5.1, the new recurrence for $\boldsymbol{r}_n$ is obtained as below,

$$\boldsymbol{r}_{n+1} = \Big(1 + \frac{\beta_{n-1}\alpha_n}{\alpha_{n-1}} - \alpha_n \hat{A}\Big)\boldsymbol{r}_n - \frac{\beta_{n-1}\alpha_n}{\alpha_{n-1}}\boldsymbol{r}_{n-1}. \tag{5.27}$$

The above equation is also valid for the shifted linear equations,

$$\boldsymbol{r}_{n+1}(\sigma) = \Big(1 + \frac{\beta_{n-1}(\sigma)\alpha_n(\sigma)}{\alpha_{n-1}(\sigma)} - \alpha_n(\sigma)(\hat{A} + \sigma\hat{I})\Big)\boldsymbol{r}_n(\sigma) - \frac{\beta_{n-1}(\sigma)\alpha_n(\sigma)}{\alpha_{n-1}(\sigma)}\boldsymbol{r}_{n-1}(\sigma). \tag{5.28}$$

From Eqs. (5.24)-(5.26), one can derive the recurrence equations that determine the scalar functions $\pi_n(\sigma), \alpha_n(\sigma)$, and $\beta_n(\sigma)$:

$$\pi_{n+1}(\sigma) = \Big(1 + \frac{\beta_{n-1}\alpha_n}{\alpha_{n-1}} - \alpha_n\sigma\Big)\pi_n(\sigma) - \frac{\beta_{n-1}\alpha_n}{\alpha_{n-1}}\pi_{n-1}(\sigma), \tag{5.29}$$

$$\alpha_n(\sigma) = \frac{\pi_n(\sigma)}{\pi_{n+1}(\sigma)}\alpha_n, \tag{5.30}$$

$$\beta_n(\sigma) = \Big(\frac{\pi_n(\sigma)}{\pi_{n+1}(\sigma)}\Big)^2\beta_n. \tag{5.31}$$

It should be noted that these recurrences are evolved without time-consuming matrix vector operations. Thus, updating vectors $\boldsymbol{x}(\sigma)$ and $\boldsymbol{p}(\sigma)$ using Eqs. (5.30) and (5.31) will lead to a significant reduction of the computational cost. This is the shifted COCG method [80] and its algorithm is written in Table 5.2.

Next, I briefly comment on two numerical techniques about the shifted Krylov solvers. The shifted COCG method is not applicable to Eq. (5.11) when $[z\hat{I} - \hat{H}_T]$ is not complex symmetric matrix. This situation occurs when z is complex number and $\hat{H}_T$ is complex Hermitian matrix, $\hat{H}_T = \hat{H}_T^\dagger \neq \hat{H}_T^T$. In this case, the shifted BiCG method [81] might be suitable. On the other hand, if $z = \varepsilon$, $[\varepsilon\hat{I} - \hat{H}_T]$ is always Hermite, and one should use the shifted CG method [82]. The shifted CG, COCG, and BiCG methods are quite similar to each other, but the only shifted BiCG method requires two matrix-vector operations per iteration. Therefore, the use of shifted BiCG will increase the computational cost of matrix-vector operation as twice as larger than the shifted CG or COCG method. However, the shifted CG method can be applied to the problem with complex shifts if the reference energy is set on the real axis, leading to a half reduction of the matrix-vector operation which is the most time-consuming part. Furthermore, if $\hat{H}_T$ is a real-symmetric matrix, one should choose seed system as $z = \varepsilon$ and employ the shifted CG method instead of shifted COCG method because the shifted COCG method requires complex matrix-vector operation for a complex seed while the shifted CG method does real-matrix vector operations for a real seed.

The second technique is the seed switching [83]. There is no guarantee that the residual norms of all shifts satisfy the convergence criterion when that of the seed is sufficiently small. Because the iterations are performed until all the residual norms satisfy the convergence criterion, the total number of iterations increases, resulting in the slight increase of CPU time in the shifted COCG method with respect to the number of shifts. In addition, when the norm of the residual vector of

Table 5.2: Shifted COCG algorithm for $[\hat{A} + \sigma\hat{I}]\boldsymbol{x}_n(\sigma) = \boldsymbol{b}$, with a complex symmetric matrix $\hat{A} \in \mathbb{C}^{M \times M}, \sigma \in \mathbb{C}, \boldsymbol{x}_n, \boldsymbol{p}_n, \boldsymbol{r}_n, \boldsymbol{x}_n(\sigma), \boldsymbol{p}_n(\sigma) \in \mathbb{C}^M$, and $\alpha_n, \beta_n, \alpha_n(\sigma), \beta_n(\sigma), \pi_n(\sigma) \in \mathbb{C}$. Note that the inner product is defined as $(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u}^T \boldsymbol{v}$.

1: Input σ_j $(j = 1, 2, ..., N_s)$
2: Set $\boldsymbol{x}_0 = \boldsymbol{p}_{-1} = 0$, $\boldsymbol{r}_0 = \boldsymbol{b}$, $\alpha_{-1} = 1$, and $\beta_{-1} = 0$
3: Set $\boldsymbol{x}_0(\sigma_j) = \boldsymbol{p}_{-1}(\sigma_j) = 0$, $\boldsymbol{r}_0(\sigma_j) = \boldsymbol{b}$, $\alpha_{-1}(\sigma_j) = 1$, and $\beta_{-1}(\sigma_j) = 0, \pi_{-1}(\sigma_j) = \pi_0(\sigma_j) = 1$ $(j = 1, 2, ..., N_s)$
4: For $n = 0, 1, 2, ...$, until convergence do:
5:　　　$\alpha_n = (\boldsymbol{r}_n, \boldsymbol{r}_n)/(\boldsymbol{p}_n, \hat{A}\boldsymbol{p}_n)$
6:　　　$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \alpha_n \boldsymbol{p}_n$
7:　　　$\boldsymbol{r}_{n+1} = \boldsymbol{r}_n - \alpha_n \hat{A}\boldsymbol{p}_n$
8:　　　$\beta_n = (\boldsymbol{r}_{n+1}, \boldsymbol{r}_{n+1})/(\boldsymbol{r}_n, \boldsymbol{r}_n)$
9:　　　$\boldsymbol{p}_{n+1} = \boldsymbol{r}_n + \beta_n \boldsymbol{p}_n$
10:　　　For $j = 1, 2, ..., N_s$ do:
11:　　　　$\pi_{n+1}(\sigma) = \left(1 + \frac{\beta_{n-1}\alpha_n}{\alpha_{n-1}} - \alpha_n\sigma\right)\pi_n(\sigma) - \frac{\beta_{n-1}\alpha_n}{\alpha_{n-1}}\pi_{n-1}(\sigma)$
12:　　　　$\alpha_n(\sigma_j) = \frac{\pi_n(\sigma_j)}{\pi_{n+1}(\sigma_j)}\alpha_n$
13:　　　　$\beta_n(\sigma_j) = \left(\frac{\pi_n(\sigma_j)}{\pi_{n+1}(\sigma_j)}\right)^2\beta_n$
14:　　　　$\boldsymbol{x}_{n+1}(\sigma_j) = \boldsymbol{x}_n(\sigma_j) + \alpha_n(\sigma_j)\boldsymbol{p}_n(\sigma_j)$
15:　　　　$\boldsymbol{p}_{n+1}(\sigma_j) = 1/\pi_n(\sigma_j)\boldsymbol{r}_n + \beta_n(\sigma_j)\boldsymbol{p}_n(\sigma_j)$
16:　　　End do:
17: End do

the seed system becomes small, the numerical precision of the residual vectors for the shifts degrades. The seed switching technique, which switches the seed to the point where the residual norm is largest among the whole shifts after the solution of the seed system converges, enables us to build up the residual vectors for the sampling energy points without loss of numerical precision and significant increase of the computational cost.

5.3.4　Efficient implementation of the shifted COCG method

Although a significant reduction of the computational cost will be achieved by using shifted COCG method for solving shifted linear equations, its benefit fades out as increasing the number of shifts. This is because that the computational complexity of the (sparse) matrix-vector operation is $\mathcal{O}(cM)$, where cM is the nonzero element of $M \times M$ Hamiltonian matrix, while the scalar-vector product of the shifted COCG method scales as $\mathcal{O}(N_s M)$, where N_s is the number of shifts. Thus, if $N_s M \geq cM$, a cost to update the vectors $\boldsymbol{x}(\sigma)$ and $\boldsymbol{p}(\sigma)$, a large amount of scalar-vector products, becomes comparable to a cost of constructing the Krylov subspace on a seed. Moreover, the shifted COCG method does not contribute to reduce the memory consumption. Because many energy points are required to evaluate the transmission, large amount of memory consumption secondarily limits us to obtain the unperturbed

Green's function. To overcome these difficulties, I introduce the reduced vectors $\boldsymbol{\xi}(\sigma), \boldsymbol{\theta}(\sigma)$, and $\boldsymbol{\rho}$,

$$\boldsymbol{\xi}(\sigma) = \hat{V}^T \boldsymbol{x}(\sigma), \quad \boldsymbol{\theta}(\sigma) = \hat{V}^T \boldsymbol{p}(\sigma), \quad \boldsymbol{\rho} = \hat{V}^T \boldsymbol{r}, \tag{5.32}$$

where $\hat{V} \in \mathbb{C}^{M \times L}$ and $\boldsymbol{\xi}(\sigma), \boldsymbol{p}(\sigma), \boldsymbol{\rho} \in \mathbb{C}^L$. I assume that $L \ll M$. The reduced vectors can be updated without explicit computation of $\boldsymbol{x}_n(\sigma)$ and $\boldsymbol{p}_n(\sigma)$ instead by using the following recurrences

$$\boldsymbol{\xi}_{n+1}(\sigma) = \boldsymbol{\xi}_n(\sigma) + \alpha_n(\sigma)\boldsymbol{\theta}_n(\sigma), \tag{5.33}$$
$$\boldsymbol{\theta}_{n+1}(\sigma) = \boldsymbol{\rho}_n(\sigma) + \beta_n(\sigma)\boldsymbol{\theta}_n(\sigma). \tag{5.34}$$

The modified algorithm based on the shifted COCG method is written in Table 5.3. The algorithm is mathematically trivial but quite efficient because the scalar-vector operation in Eqs (5.33) and (5.34) scales $\mathcal{O}(N_s L)$ while that in the standard shifted COCG method scales $\mathcal{O}(N_s M)$.

Table 5.3: Modified shifted COCG algorithm for $[\hat{A} + \sigma \hat{I}]\boldsymbol{x}_n(\sigma) = \boldsymbol{b}$, with a complex symmetric matrix $\hat{A} \in \mathbb{C}^{M \times M}, \hat{V} \in \mathbb{C}^{M \times L}, \sigma \in \mathbb{C}, \boldsymbol{x}_n, \boldsymbol{p}_n, \boldsymbol{r}_n \in \mathbb{C}^M, \boldsymbol{\xi}_n(\sigma), \boldsymbol{\theta}_n(\sigma), \boldsymbol{\rho}_n \in \mathbb{C}^L$, and $\alpha_n, \beta_n, \alpha_n(\sigma), \beta_n(\sigma), \pi_n(\sigma) \in \mathbb{C}$. Note that the inner product is defined as $(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u}^T \boldsymbol{v}$.

1: Input $\sigma_j \ (j = 1, 2, ..., N_s)$
2: Set $\boldsymbol{x}_0 = \boldsymbol{p}_{-1} = 0, \ \boldsymbol{r}_0 = \boldsymbol{b}, \ \boldsymbol{\rho}_0 = \hat{V}^T \boldsymbol{b}, \ \alpha_{-1} = 1$, and $\beta_{-1} = 0$
3: Set $\boldsymbol{\xi}_0(\sigma_j) = \boldsymbol{\theta}_{-1}(\sigma_j) = 0, \ \alpha_{-1}(\sigma_j) = \pi_{-1}(\sigma_j) = \pi_0(\sigma_j) = 1$, and $\beta_{-1}(\sigma_j) = 0 \ (j = 1, 2, ..., N_s)$
4: For $n = 0, 1, 2, ...$, until convergence do:
5:      $\alpha_n = (\boldsymbol{r}_n, \boldsymbol{r}_n)/(\boldsymbol{p}_n, \hat{A}\boldsymbol{p}_n)$
6:      $\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \alpha_n \boldsymbol{p}_n$
7:      $\boldsymbol{r}_{n+1} = \boldsymbol{r}_n - \alpha_n \hat{A}\boldsymbol{p}_n$
8:      $\beta_n = (\boldsymbol{r}_{n+1}, \boldsymbol{r}_{n+1})/(\boldsymbol{r}_n, \boldsymbol{r}_n)$
9:      $\boldsymbol{p}_{n+1} = \boldsymbol{r}_n + \beta_n \boldsymbol{p}_n$
10:      $\boldsymbol{\rho}_{n+1} = \hat{V}^T \boldsymbol{r}_{n+1}$
11:      For $j = 1, 2, ..., N_s$ do:
12:        $\pi_{n+1}(\sigma) = \left(1 + \frac{\beta_{n-1}\alpha_n}{\alpha_{n-1}} - \alpha_n \sigma\right)\pi_n(\sigma) - \frac{\beta_{n-1}\alpha_n}{\alpha_{n-1}}\pi_{n-1}(\sigma)$
13:        $\alpha_n(\sigma_j) = \frac{\pi_n(\sigma_j)}{\pi_{n+1}(\sigma_j)}\alpha_n$
14:        $\beta_n(\sigma_j) = \left(\frac{\pi_n(\sigma_j)}{\pi_{n+1}(\sigma_j)}\right)^2 \beta_n$
15:        $\boldsymbol{\xi}_{n+1}(\sigma_j) = \boldsymbol{\xi}_n(\sigma_j) + \alpha_n(\sigma_j)\boldsymbol{\theta}_n(\sigma_j)$
16:        $\boldsymbol{\theta}_{n+1}(\sigma_j) = 1/\pi_n(\sigma_j)\boldsymbol{\rho}_n + \beta_n(\sigma_j)\boldsymbol{\theta}_n(\sigma_j)$
17:      End do:
18: End do

The modified algorithm is useful when a part of the unperturbed Green's function, $\hat{\mathcal{G}}(z)$, is needed. This situation frequently occurs in physics. For example, the four corner matrix elements, $\hat{\mathcal{G}}_{00}, \ \hat{\mathcal{G}}_{0,N+1} \ \hat{\mathcal{G}}_{N+1,0} \ \hat{\mathcal{G}}_{N+1,N+1}$, are needed to calculate the transmission. As another example, the charge density calculation requires the diagonal elements of the retarded Green's function $\hat{G}_{m,m}$ which is computed using $\hat{\mathcal{G}}_{m,0}, \ \hat{\mathcal{G}}_{m,m}, \ \hat{\mathcal{G}}_{m,N+1}$. Finally, the efficiency of the proposed algorithm is discussed.

I consider N_s different complex energy as shifts $\sigma_j = z_j$ and set $\hat{A} = \varepsilon_F \hat{I} - \hat{H}_T$, where ε_F is the Fermi level. In the transmission calculation, four corner matrix elements are obtained by setting $\boldsymbol{b} = |0\rangle$ and $|N+1\rangle$, and $\hat{V}^T = (\langle 0|, \langle N+1|)$, where $|0\rangle$ and $|N+1\rangle$ correspond to vectors on left and right electrodes principal layer, respectively. As discussed in the previous chapter, the number of grid points in the electrode is $\widetilde{M}(= M_x \times M_y \times \widetilde{M}_z)$, where M_x and M_y are the number of grid points in x and y direction and $\widetilde{M}_z$ is the number of grid points defined so as to cover the nonlocal part of the pseudopotential in z direction. Since $\widetilde{M}_z \ll M_z$ with M_z being the number of grid points in z direction, the dimension of the reduced vectors is significantly smaller than the dimension of the linear equation $M(= M_x \times M_y \times M_z)$. In the charge density calculation, $\hat{\mathcal{G}}_{m,0}$, $\hat{\mathcal{G}}_{m,m}$, $\hat{\mathcal{G}}_{m,N+1}$ are obtained by setting $\boldsymbol{b} = |0\rangle, |m\rangle$, and $|N+1\rangle$, and $\hat{V}^T = \langle m|$.

5.3.5 Extension to the generalized shifted linear equations

Let us consider the solution of the generalized shifted linear equation with complex symmetric matrices of the form:

$$[\hat{A} + \sigma\hat{B}]\boldsymbol{x}(\sigma) = \boldsymbol{b}, \tag{5.35}$$

where $\hat{A}$ and $\hat{B}$ are non-singular $M \times M$ matrices. When $\hat{B} = \hat{I}$, Eq. (5.35) is reduced to Eq. (5.24) and thus the shifted COCG method is applicable. However, if $\hat{B} \neq \hat{I}$, the shift-invariant property of the Krylov subspace does not hold. This situation occurs when the norm-conservation condition does not apply or the tight-binding Hamiltonian based on the non-orthogonal atomic basis set is used. Note that $\hat{B}$ is often called an overlap matrix in the electronic structure calculation. Recently, Teng *et al.* [84] proposed the extended algorithm of the shifted COCG method for solving the generalized shifted linear equation (5.35). The extension is based on transforming Eq. (5.35) into the standard shifted linear equation

$$[\hat{B}^{-1}\hat{A} + \sigma\hat{I}]\boldsymbol{x}(\sigma) = \hat{B}^{-1}\boldsymbol{b}. \tag{5.36}$$

Following the procedure of the shifted COCG method, the residual vector $\boldsymbol{r}'_n = \hat{B}^{-1}\boldsymbol{b} - \hat{B}^{-1}\hat{A}\boldsymbol{x}_n$ holds the orthogonality relation,

$$\boldsymbol{r}'_n \in \mathcal{K}_{n+1}(\hat{B}^{-1}\hat{A} + \sigma\hat{I}, \hat{B}^{-1}\boldsymbol{b}) \perp \mathcal{K}_n(\overline{\hat{B}^{-1}\hat{A} + \sigma\hat{I}}, \bar{\boldsymbol{b}}). \tag{5.37}$$

It means that the orthogonality of the vectors in the COCG method is extended by the following bilinear form:

$$(\boldsymbol{r}'_{n+1}, \boldsymbol{r}'_n)_B = \boldsymbol{r}'^T_{n+1}\hat{B}\boldsymbol{r}'_n = 0. \tag{5.38}$$

The resulting algorithm, the generalized shifted COCG method, is written in Table 5.4. Note that in the actual calculation $\boldsymbol{r}'_{n+1} = \hat{B}^{-1}\boldsymbol{r}_{n+1}$ is computed by solving the linear equation

$$\hat{B}\boldsymbol{r}'_{n+1} = \boldsymbol{r}_{n+1}, \tag{5.39}$$

at each step. Since $\hat{B}$ is usually real-symmetric positive definite, equation can be solved by the CG method.

Table 5.4: Generalized shifted COCG algorithm for $[\hat{A} + \sigma\hat{B}]\boldsymbol{x}_n(\sigma) = \boldsymbol{b}$, with a complex symmetric matrices $\hat{A}, \hat{B} \in \mathbb{C}^{M \times M}, \sigma \in \mathbb{C}, \boldsymbol{x}_n, \boldsymbol{p}_n, \boldsymbol{r}_n, \boldsymbol{r}_n' \in \mathbb{C}^M$, and $\alpha_n, \beta_n, \alpha_n(\sigma), \beta_n(\sigma), \pi_n(\sigma) \in \mathbb{C}$. Note that the inner product is defined as $(\boldsymbol{u}, \boldsymbol{v})_B = \boldsymbol{u}^T\hat{B}\boldsymbol{v}$.

1: Input σ_j $(j = 1, 2, ..., N_s)$
2: Set $\boldsymbol{x}_0 = \boldsymbol{p}_{-1} = 0$, $\boldsymbol{r}_0 = \boldsymbol{b}$, $\boldsymbol{r}_0' = \hat{S}^{-1}\boldsymbol{b}$, $\alpha_{-1} = 1$, and $\beta_{-1} = 0$
3: Set $\boldsymbol{x}_0(\sigma_j) = \boldsymbol{p}_{-1}(\sigma_j) = 0$, $\alpha_{-1}(\sigma_j) = \pi_{-1}(\sigma_j) = \pi_0(\sigma_j) = 1$, and $\beta_{-1}(\sigma_j) = 0$ $(j = 1, 2, ..., N_s)$
4: For $n = 0, 1, 2, ...$, until convergence do:
5:    $\alpha_n = (\boldsymbol{r}_n', \boldsymbol{r}_n')_B / (\boldsymbol{p}_n, \hat{B}^{-1}\hat{A}\boldsymbol{p}_n)_B$
6:    $\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \alpha_n\boldsymbol{p}_n$
7:    $\boldsymbol{r}_{n+1} = \boldsymbol{r}_n - \alpha_n\hat{A}\boldsymbol{p}_n$
8:    $\boldsymbol{r}_{n+1}' = \hat{B}^{-1}\boldsymbol{r}_{n+1}$
9:    $\beta_n = (\boldsymbol{r}_{n+1}', \boldsymbol{r}_{n+1}')_B / (\boldsymbol{r}_n', \boldsymbol{r}_n')_B$
10:    $\boldsymbol{p}_{n+1} = \boldsymbol{r}_n' + \beta_n\boldsymbol{p}_n$
11:    For $j = 1, 2, ..., N_s$ do:
12:      $\pi_{n+1}(\sigma) = \left(1 + \frac{\beta_{n-1}\alpha_n}{\alpha_{n-1}} - \alpha_n\sigma\right)\pi_n(\sigma) - \frac{\beta_{n-1}\alpha_n}{\alpha_{n-1}}\pi_{n-1}(\sigma)$
13:      $\alpha_n(\sigma_j) = \frac{\pi_n(\sigma_j)}{\pi_{n+1}(\sigma_j)}\alpha_n$
14:      $\beta_n(\sigma_j) = \left(\frac{\pi_n(\sigma_j)}{\pi_{n+1}(\sigma_j)}\right)^2\beta_n$
15:      $\boldsymbol{x}_{n+1}(\sigma_j) = \boldsymbol{x}_n(\sigma_j) + \alpha_n(\sigma_j)\boldsymbol{p}_n(\sigma_j)$
16:      $\boldsymbol{p}_{n+1}(\sigma_j) = 1/\pi_n(\sigma_j)\boldsymbol{r}_n' + \beta_n(\sigma_j)\boldsymbol{p}_n(\sigma_j)$
17:    End do:
18: End do

In compensation for obtaining the shift-invariant property in the generalized shifted COCG method, one need to solve linear equation (5.39) at each step, which might lead to increase the computational cost. Here I present the numerical technique to compute $\hat{B}^{-1}$. In the ultrasoft pseudopotentials and projector augmented wave method, $\hat{B}$ has a form

$$\hat{B} = \hat{I} + \hat{P}\hat{D}_B\hat{P}^T, \tag{5.40}$$

with $\hat{P} \in \mathbb{C}^{M \times N_B}$ and $\hat{D}_B \in \mathbb{C}^{N_B \times N_B}$. Here N_B is the number of projectors, which is as many as the number of atoms. Since $N_B \ll M$, the Sherman-Morrison-Woodbury formula yields

$$\hat{B}^{-1} = \hat{I} - \hat{P}(\hat{D}_B + \hat{P}^T\hat{P})^{-1}\hat{P}^T. \tag{5.41}$$

Thus, the matrix inversion $\hat{B}^{-1}$ is reduced to the $N_B \times N_B$ dimensional inversion $(\hat{D}_B + \hat{P}^T\hat{P})^{-1}$ which can be directly computed by LU-decomposition method in advance.

5.4   Benchmark test

In this section, I will demonstrate the power of the proposed method through transmission calculations of Na atomic wire and C_{60}@(10,10)CNT peapod. The Hamiltonian matrices are obtained from the

real-space pseudopotential DFT code RSPACE [85, 86]. All calculations in this section are performed by using the local density approximation [12] and norm-conserving pseudopotentials proposed by Troullier-Martins [24]. Unless noted otherwise, I employ the central finite-difference approximation ($N = 1$ in Ref. [75]) for the Laplacian operator.

5.4.1   Na atomic wire

To demonstrate the efficiency of the present scheme, the shifted COCG method with the seed switching technique is applied to the calculation of the unperturbed Green's functions of the transition region for a Na atomic wire. The size of the supercell is $20.0 \times 20.0 \times 7.0$ bohr^3 and a single Na atom is contained in the supercell. The number of grid points is $40 \times 40 \times 16$. The transport direction is set along z direction.

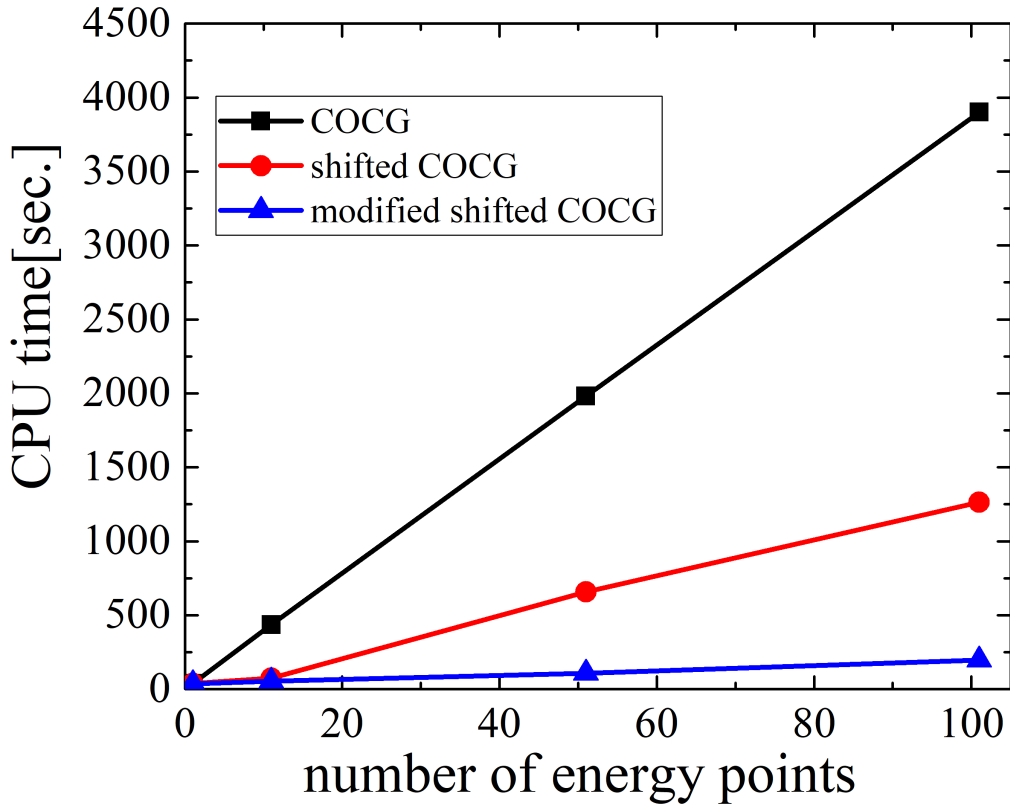Figure 5.1: CPU time required to obtain unperturbed Green's functions of Na atomic wire. The black square, red circle, and blue triangle are the results obtained by the COCG (Table 5.1), shifted COCG (Table 5.2), and the modified shifted COCG (Table 5.3), respectively. The energy points are set are chosen so as to be equidistance in the interval $\varepsilon - \varepsilon_F \in [-1, 1]$ eV, where ε_F is the Fermi level.

Figure 5.1 shows the CPU time versus the number of sampling energy points for the COCG (Table 5.1), shifted COCG (Table 5.2 [80]), and the modified shifted COCG (Table 5.3 [80]). The energy points are set so as to be equidistance in the interval $\varepsilon - \varepsilon_F \in [-1, 1]$ eV, where ε_F is the Fermi level. The CPU time of the COCG method is almost proportional to the number of energy point. In the shifted COCG method, although the CPU time is substantially reduced compared with the result of the COCG method, the CPU time linearly increases over 11 energy points. It means that the computational cost of updating vectors in shifted system becomes dominant in the whole computation. Since, in this calculation, the (Hamiltonian) matrix-vector operation can be decomposed into the 7 times scalar-vector operations originated from the central finite-difference approximation and 4 times vector-vector and scalar-vector operations from the evaluation of Na pseudopotential contribution per iteration, the computational complexity of the matrix-vector operation is about $\mathcal{O}(11M)$. On the other hand, the computational complexity of the scalar-vector operation in shifted system is $\mathcal{O}(N_s M)$, and therefore the speed-up owing to using the shifted COCG method saturates at around $N_s \approx 11$, which is consistent with the result of the shifted COCG method in Fig. 5.1. This simple analysis is valid for understanding the result of the modified shifted COCG method. In this calculation, the dimension of the reduced vectors is 8 times smaller than that of full vectors in the conventional shifted COCG method, which implies that the cost of the scaler-vector operations in shifts reduced by almost 8 times. Actually, the increase of the CPU time using the modified shifted COCG method is substantially suppressed against the result of the shifted COCG method over 11 energy points.

The above discussion suggests that the modified shifted COCG method is particularly efficient in the following two situations: (i) the matrix-vector operation is time-consuming, e.g., higher-order finite-difference approximation is employed or many atoms are contained in the supercell, (ii) the distance of the transition region along current flow direction is sufficiently large. Because both situations are usually satisfied in the large-scale transport calculations, the modified shifted COCG method might be suitable to investigate the transport properties of large-scale conductors.

5.4.2 C_{60}@(10,10)CNT peapod

I next compute a transmission of C_{60}@(10,10)CNT peapod, in which a (10,10)CNT encapsulates a single C_{60} molecule, as an application of the proposed method to a large system. This system has been observed by transmission electron microscopy [87] and its electronic structure has been intensively studied by DFT calculations [88,89]. The transport properties of peapods have mostly been investigated using the tight-binding approximation [90] because of the large system size required for calculations. However, the tight-binding Hamiltonian which is constructed by carbon $2s$ and $2p$ atomic orbitals does not take into account the effect of the interaction between the molecular state of C_{60} and nearly free electron state of the carbon nanotube correctly, making it difficult the qualitative prediction of the transport properties. Comparing the results between DFT and tight-binding calculations is one of our interests.

Figure 3.2 shows the computational model. The size of the transition region is taken to be $42.2 \times$

42.2×32.6 bohr and the transition region contains 340 carbon atoms. The initial carbon-carbon distance is set to 2.68 bohr and structural relaxation is carried out. To determine the Kohn-Sham effective potential for electrode and transition regions, the conventional supercell is used, where the periodic boundary condition in imposed in all directions, and the grid spacing is ~ 0.45 bohr^3, which gives a total number of grid points $774144 (= 96 \times 96 \times 84)$.

The electronic band structures of the (10,10)CNT and C_{60}@(10,10)CNT peapod are shown in Figs. 5.2(a) and 5.2(b), respectively. In an isolated C_{60} molecule, there is the threefold degenerated lowest unoccupied state t_{1u} which is characterized by quantum number $m = -1, 0, 1$ [91,92]. Because the structural symmetry of the C_{60} is broken, degenerated t_{1u} levels are split into three levels which are located just above the Fermi level in Fig. 5.2(b). Since the difference of the band structures between (10,10)CNT and C_{60}@(10,10)CNT peapod is significant only at just above the Fermi level, it might be expected that the transport properties of the (10,10)CNT is not disordered except around the Fermi energy.
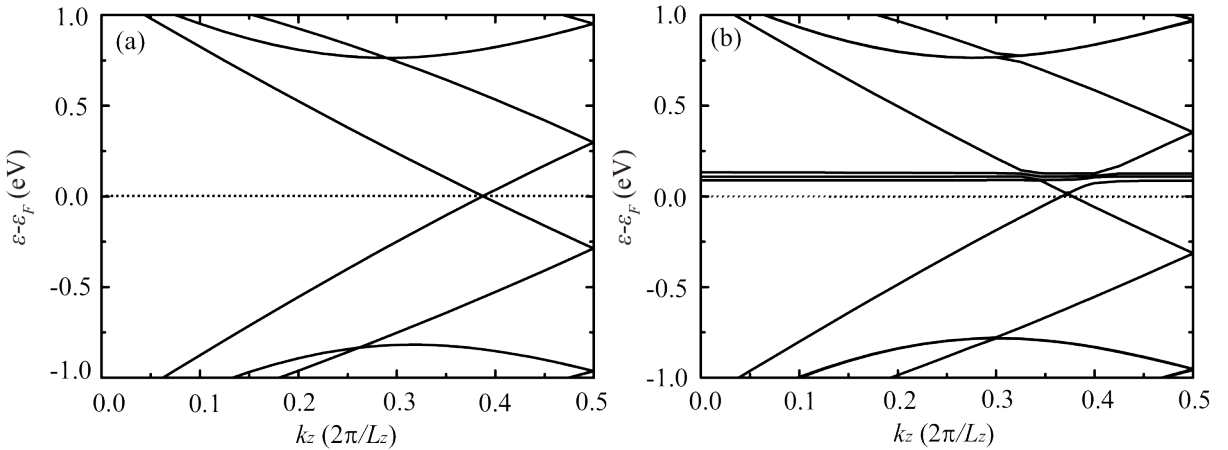
Figure 5.2: Electronic band structures of (a) (10,10)CNT and (b) C_{60}@(10,10)CNT. The Fermi level is marked by the dotted line.

The conductance spectrum of the C_{60}@(10,10)CNT is plotted in Fig. 5.3. Owing to the use of the shifted COCG method, a large number of energy points can be sampled to detect the spiky dips in the spectrum. It is found that the reductions of the conductance with maximally one quantum unit occurs at energy levels of t_{1u} states. Figure 5.4 shows a real-space picture of the charge densities of scattering wavefunctions, which is useful for understanding the spatial behaviors of the transport phenomenon. The charge density distribution with the energy at $T(E) \approx 2$ spreads around the (10,10)CNT, while that with the energy at the dip of the conductance spectrum accumulates in the vicinity of the C_{60} molecule. These results imply that the transport behavior of (10,10)CNT is only modified at the resonant scattering of the t_{1u} states of the encapsulated C_{60} molecule. Note that the resonant scattering is observable if the Fermi level of the (10,10)CNT is manipulated by impurity

doping or gate-voltage. The obtained result is qualitatively similar to the tight-binding calculation [90]. However, the tight-binding approximation shows that the positions of the resonant scattering are at below the Fermi level, while the DFT calculations predict the resonant scattering occurs above the Fermi level. This discrepancy might come from the C_{60}-CNT interaction is difficult to be taken into account by the tight-binding approximation, which limits us to predict the quantitative prediction of the transport properties of peapods.

Narrow conduction dips with maximally one quantum unit can be understood by the one-dimensional tight-binding model attached with single impurity [90]. This model indicates that the quite narrow resonance peaks corresponds to the weak interaction between the (10,10)CNT and encapsulated fullerene. Furthermore, from this model, there is a possibility of controlling the position of the resonant scattering of the C_{60}@(10,10)CNT by modifying the molecular level of the t_{1u} states of the fullerene. For example, endohedral doping is one of the interesting aspects of fullerene to modulate the transport properties of the peapod since it modifies the electronic structure of the fullerene without altering the shape of the fullerene. In last decades, many attempts have been made on the doping on the C_{60} using alkali [93], transition [94], rare-earth metals [95], noble gas [96], and molecules [97]. Here I focus on alkali metal endohedral fullerene M@C_{60} (M=Li, Na, K) because M@C_{60} shows the interesting physical properties not observed in the C_{60} due to the charge transfer from encapsulated metal to fullerene. Owing to the weak C_{60}-CNT interaction, the M@C_{60}@(10,10)CNT may exhibit the similar resonant scattering at the energy levels of the t_{1u} states as same as the fullerene case. However, because the t_{1u} state of M@C_{60} is partially occupied by valence electron of alkali metal, the reduction of the conductance at the Fermi level will be expected. Since peapods and M@C_{60} have recently become experimentally accessible structures, the obtained results should be helpful in understanding and predicting the resonant scattering in experimental transport measurements.

5.5   Application

5.5.1   Silicon Carbide MOSFETs

Silicon carbide (SiC) is a promising material for power electronics because of its wide band gap, high breakdown electric field, high thermal conductivity, and ability to form a native gate insulator, namely SiO_2. One of the most serious problems of SiC based metal-oxide-semiconductor field-effect transistors (MOSFETs), primarily those of the n-channel type, is their low channel mobility caused by excessive electron scattering at the SiC/SiO_2 interface [98, 99]. It is widely accepted that the degradation of channel mobility originates from electron trapping, Coulomb scattering, surface phonon scattering, and surface roughness scattering [100]. The Hall measurement suggests that the amount of electron trapping is reasonably low after NO annealing and that it does not play an important role in the channel mobility under a heavy inversion condition [101]. Coulomb scattering can be suppressed by reducing the interface state density. However, even after the interface state density is significantly reduced by passivation treatment, the peak channel mobility is still less than 10% of the bulk mobility [102–104].
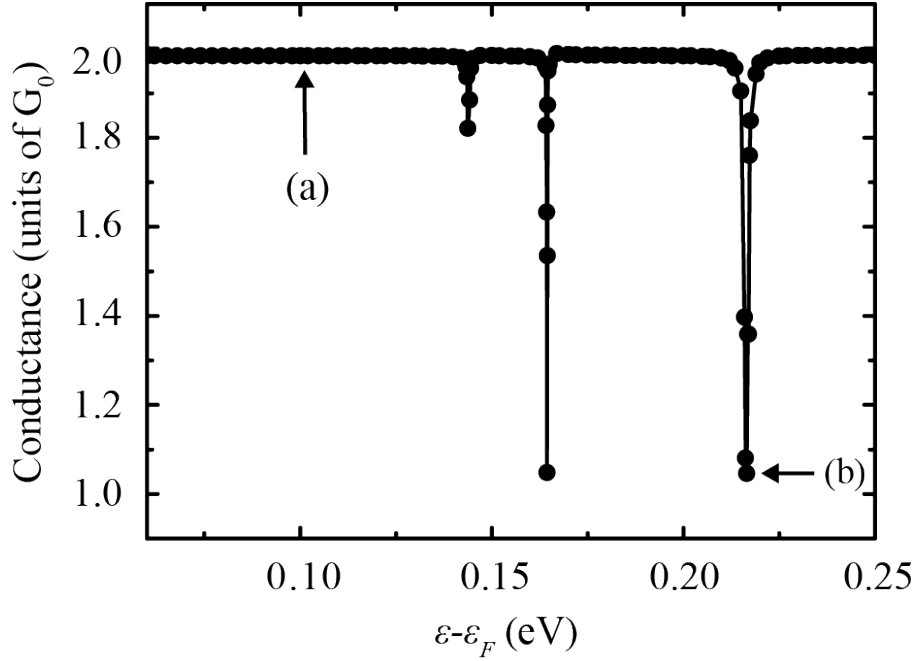
Figure 5.3: Conductance spectrum of C_{60}@(10,10)CNT.

On the basis of mobility models parametrized by the fitting of experimental mobility data, it has been reported that surface phonon scattering is not a factor limiting channel mobility [101, 105]. Although the surface roughness scattering is dominant at a high effective field, the channel mobility is low at a low effective field. The above results imply that an additional scattering mechanism that does not appear in a conventional Si/SiO_2 interface is needed for the accurate modeling of mobility. A more comprehensive understanding of the electron-scattering mechanism at the SiC/SiO_2 interface will be indispensable for further improving the channel mobility in SiC-MOSFETs.

SiC has numerous polytypes, which are characterized by the stacking sequence along the [0001] direction. Recently, Matsushita *et al.* performed first-principles electronic-structure calculations for SiC polytypes and revealed that the wave functions at the conduction-band edge (CBE) of SiC are distributed not near atomic sites but in the internal space [106, 107]. The shape of the internal-space states (ISSs) determines the band gap and electron mobility of SiC polytypes. On the other hand, for the SiC/SiO_2 interface, the behavior of the ISSs is affected by the surface orientation of the SiC substrate because the wave functions of the ISSs are distributed along a specific crystal direction. Kirkham *et al.* previously investigated the 4H-SiC(0001)/SiO_2 interface, which is commonly employed for MOSFETs, and found that the spatial distribution of the ISSs near the interface varies between two types of interface structures denoted by the *h* and *k* types, which have cubic and hexagonal stacking sequences from the top of the SiC bilayer, respectively [108]. Indeed, for the SiC surface, the existence of such inequivalent structures is supported theoretically [109, 110] and experimentally [111]. Moreover,
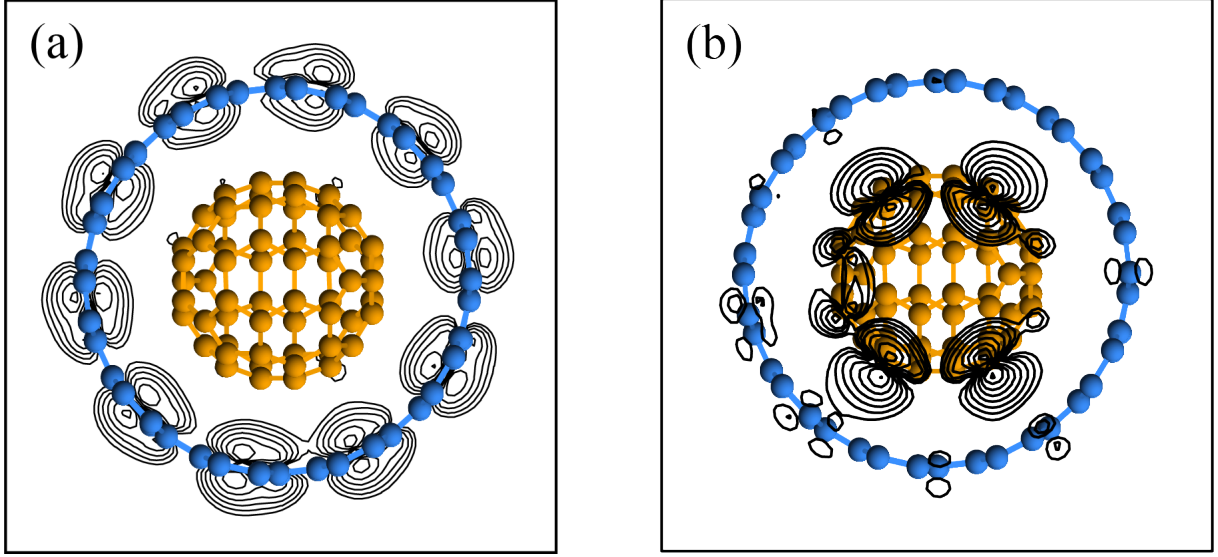
Figure 5.4: Charge density distributions of scattering wavefunctions of C_{60}@(10,10)CNT. (a) and (b) correspond to the energies indicated by the arrows in Fig. 5.3. The spheres represent the positions of carbon atoms. Each contour represents twice or half the charge density of the adjacent contour lines. The lowest-density contour represents a density of 5.0×10^{-4} $e/\text{Å}^3$

a vicinal 4H-SiC(0001)/SiO$_2$ interface containing the h and k types was observed by transmission electron microscopy [112]. Since carrier electrons pass through the CBE states, i.e., the ISSs, in n-channel MOSFETs, it is important to examine the relationship between the behavior of the ISSs at both interfaces and the transport property through the ISSs.

For the modeling of carrier mobility in SiC devices, technology computer-aided design tools using empirical scattering parameters are usually employed. Although technology computer-aided design simulations are useful for obtaining a rough estimation of the scattering mechanism, it is unclear in many cases how the experimental data should be interpreted in terms of the microscopic behavior. More rigorous and accurate theoretical processes based on first-principles are required to take into account the effect of the ISSs on the transport property. Recently, Iskandarova *et al*. performed first-principles electron-transport calculations for the 4H-SiC(0001)/SiO$_2$ interface using localized basis sets [113]. However, since localized basis sets cannot reproduce the ISSs correctly [106], the contribution of the ISSs has not been properly investigated. Therefore, to my knowledge, there have been no first-principles electron-transport calculations examining the contribution of the ISSs to the scattering property of defects at the SiC/SiO$_2$ interface.

In this study, first-principles calculations on the electron-scattering property of the oxygen-related structures at the 4H-SiC(0001)/SiO$_2$ interface, which appear during dry oxidation, are performed. Note that the oxygen-related structures at the h type, which do not have defect states at the interface, give rise to considerable electron scattering. The large scattering at the h type is direct evidence

that the difference in the behavior of the ISSs between the h and k types plays a decisive role in the electron-transport property at the SiC/SiO$_2$ interface. I also examine the electron transmission when carbon-related defects exist in the interface. It is intriguing that the effect of the ISSs on the electron scattering is more significant than that of the defects. Since electron scattering by electrically inactive oxygen-related structures does not generally occur in conventional Si-MOSFETs, the finding provides a new paradigm for researchers interested in SiC-MOSFETs research. Because the 4H-SiC(0001)/SiO$_2$ interface inevitably contains both interface types, electron scattering at the SiC/SiO$_2$ interface generated by dry oxidation is intrinsic and appears to be one of the limiting factors for obtaining high channel mobility in n-channel SiC-MOSFETs.

5.5.2 Computational details
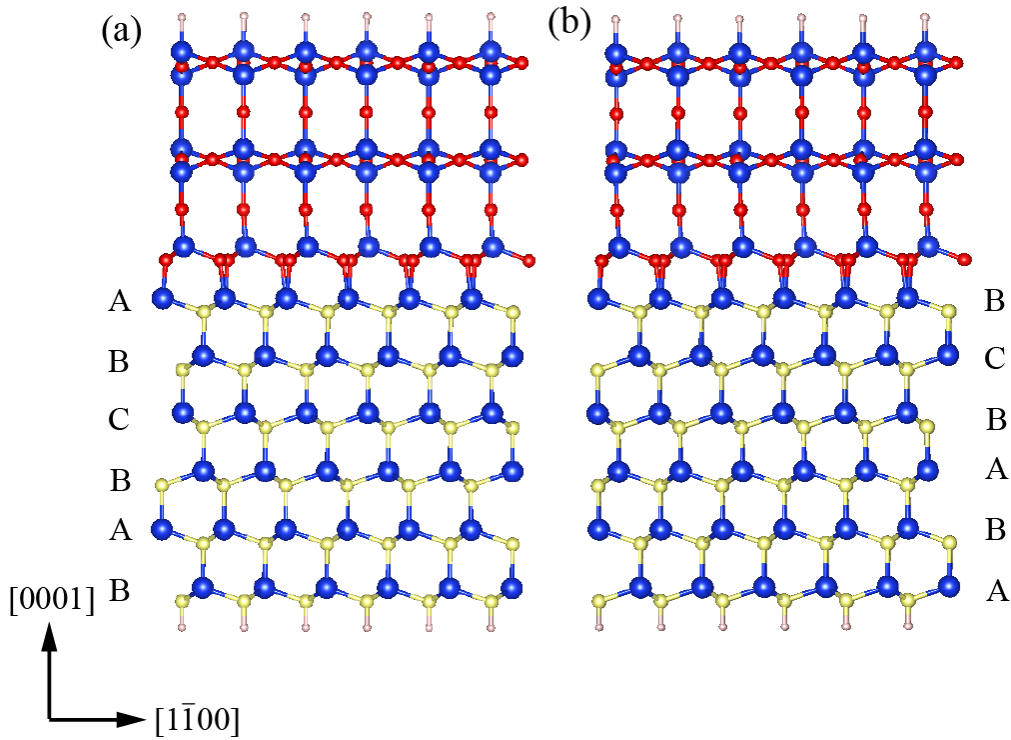
Figure 5.5: Atomic structures of 4H-SiC(0001)/SiO$_2$ interface. (a) h type and (b) k type. Blue, yellow, red, and white spheres are Si, C, O, and H atoms, respectively.

Figure 5.5 shows the atomic structure of the interface. Since the transport calculation for the interface between crystalline SiC and amorphous SiO$_2$ is computationally difficult, the crystalline interface model generated in previous works [108, 114–117] is used to examine the electron-transport property of the 4H-SiC(0001)/SiO$_2$ interface. The two-dimensional slab model with a 21 bohr vacuum region contains a crystalline substrate with 6 planes of SiC bilayers connected to β-tridymite SiO$_2$

with a thickness of 17 bohr. This model is referred to as the initial interface. Since it has been reported that dry oxidation occurs via the reaction of O_2 sequentially arriving at the interface with the CO emission [108,116,118], the following oxygen-related structures will be present at the SiC/SiO$_2$ interface: single oxygen interstitials at the interface, O_{if}, double oxygen interstitials at the interface and subsurface, O_{if+sub}, and carbon vacancies at the interface saturated by two O atoms, $V_C O_2$. In addition, single carbon interstitials, $(C\text{-}C_i)_C$, and carbonyl complexes, CC, at the interface, which are referred to as carbon-related defects, are investigated since they are among the strongest candidates for the interface defects according to other theoretical and experimental results [118–120]. The atomic structures of these defects are illustrated in Fig. 5.6. The optimized atomic structures are obtained in the same manner as in Ref. [108].
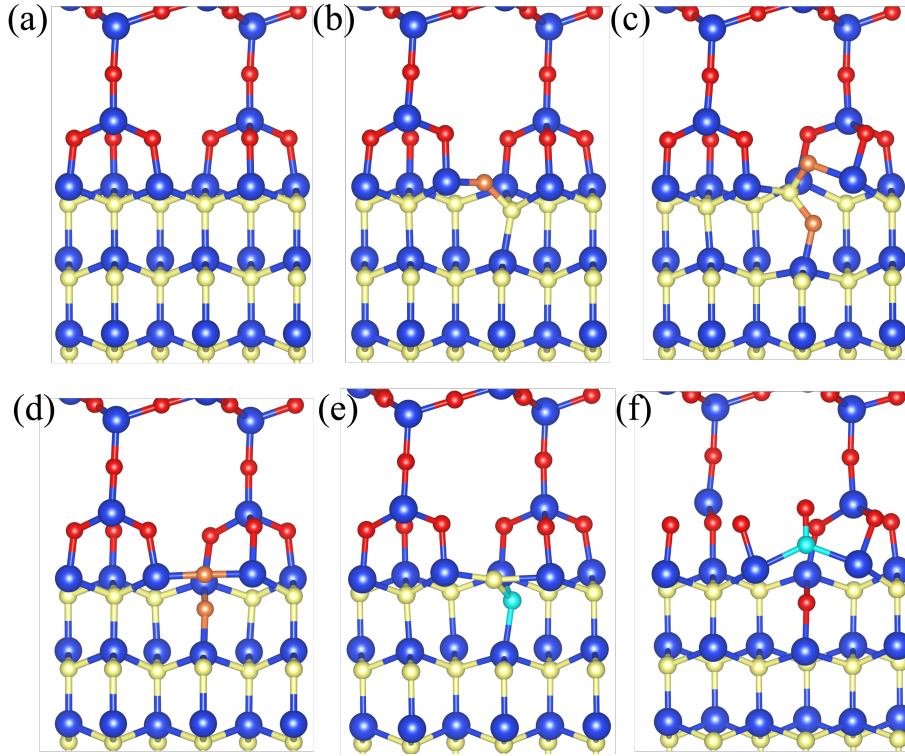
Figure 5.6: Atomic structures of (a) clean, (b) O_{if}, (c) O_{if+sub}, (d) $V_C O_2$, (e) $(C\text{-}C_i)_C$, and (f) CC at SiC/SiO$_2$ interface. Green spheres are additional O atoms in O_{if}, O_{if+sub}, and $V_C O_2$. Light-blue spheres are C atoms in $(C\text{-}C_i)_C$ and CC. Other colors are the same as those in Fig. 5.5.

Figure 5.7 illustrates a schematics of the computational models used for the transport calculations, in which the whole system is divided into three parts: a left electrode, a transition region, and a right electrode. The transition region is composed of the SiC/SiO$_2$ interface including the oxygen-related structures or carbon-related defects. The left (right) electrode is a semi-infinite slab along the $[11\bar{2}0]$ ($[\bar{1}\bar{1}20]$) direction and its atomic structures correspond to those of the initial interface. Periodic

boundary conditions are imposed in the $\langle 1\bar{1}00 \rangle$ and $\langle 0001 \rangle$ directions. The transmission from the left electrode to the right electrode can be derived using the Fisher-Lee formula. Transport calculations are performed using the real-space finite-difference method implemented in the RSPACE code. The exchange-correlation energy among the electrons is treated in the local density approximation. Norm-conserving pseudopotentials generated by the Troullier-Martins scheme are adopted to describe the electron-ion interaction. A grid spacing of 0.42 bohr in the real space and the Γ-point approximation in the two-dimensional Brillouin zone are used.
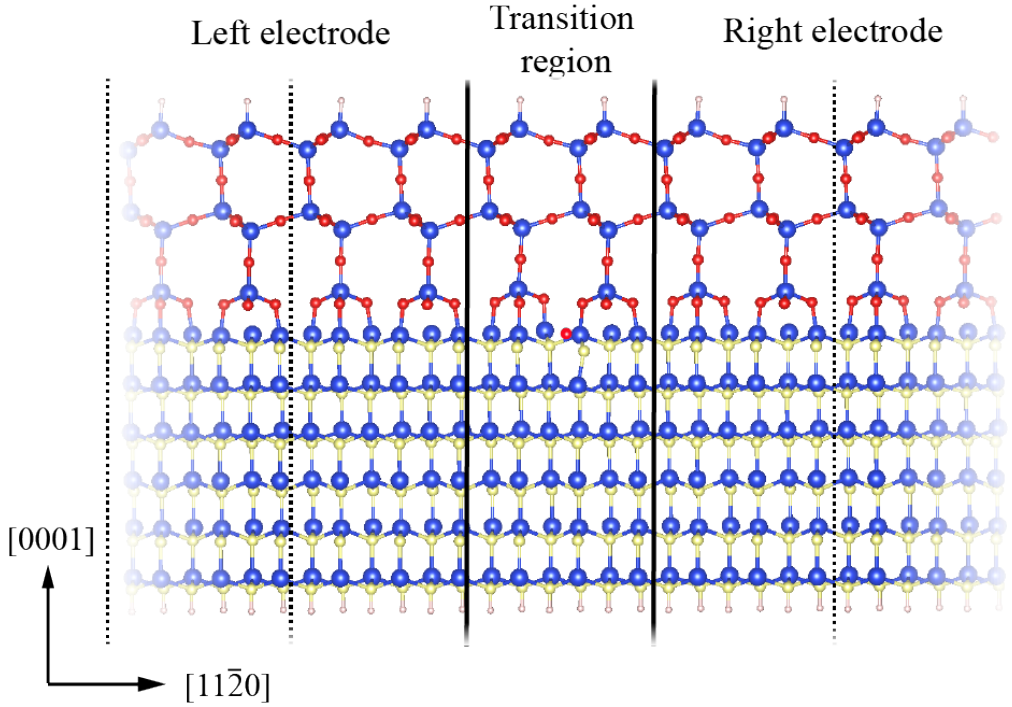
Figure 5.7: Schematic image of transport-calculation model. The boundary between the scattering region and the semi-infinite electrodes is distinguished by solid lines. Supercells of electrodes are bounded by dotted lines. The illustrated transition region contains O_{if} within the h type. Colors are the same as those in Fig. 5.5.

5.5.3   Results and discussion

I first discuss the effect of the oxygen-related structures on the electron-transport property at the SiC/SiO_2 interface. The transmission spectra defined by Eq. (3.25) are shown in Figs. 5.8(a) and 5.8(b). The transmission of the initial interface is depicted as a black line for comparison. Since it has been reported that p-channel SiC devices cannot compete with Si devices [99], the majority of SiC-MOSFETs are n-channel ones. Thus, I focus on the electron transport through the CBE states. I find that the transmission at the CBE strongly depends on the type of interface, decreasing at the h type

but almost remaining unchanged at the k type. This result indicates that the electron transmission through the h type is sensitive to the structural deformation caused by the insertion of oxygen. It is surprising that the oxygen-related structures, which are naturally generated at the SiC/SiO$_2$ interface during dry oxidation, cause the electron scattering because the oxygen-related structures considered here have been reported to be electrically inactive [118].

To obtain further insight into the origin of the electron scattering at the h type, I perform the eigenchannel decomposition [121] of scattering wave functions, where the scattering wave functions are obtained using G_T, Γ_L, and incident waves from the left electrode. I show the channel transmission and the spatial distributions of the square of the scattering wave function only for the case of O$_{\mathrm{if}}$ in Fig. 5.9 since no significant differences are observed among O$_{\mathrm{if}}$, O$_{\mathrm{if+sub}}$, and V$_{\mathrm{C}}$O$_2$. It is found that there are three channels in the cubic-stacking regions of SiC (ABC or CBA). Although these channels have the same energy level in the bulk case, they have slightly different energy levels owing to the existence of the interface for the present slab models. The transmission in the first and second channels rapidly saturates to unity, which means that there is no electron scattering. On the other hand, the transmission through the third channel, which is located slightly below the SiC/SiO$_2$ interface, is low. To consider the scattering property of the third channel in more detail, I calculate the barrier height V of the scattering potential using a one-dimensional free-electron-like model. V is fitted so as to reproduce the transmission probability of the third channel $T_{\mathrm{3rd}}(\varepsilon)$ obtained by first-principles calculations,

$$T_{\mathrm{3rd}}(\varepsilon) = \begin{cases} \frac{4\varepsilon(V-\varepsilon)}{4\varepsilon(V-\varepsilon)+V^2\sinh^2\kappa b} & (\varepsilon < V) \\ \frac{4\varepsilon(\varepsilon-V)}{4\varepsilon(\varepsilon-V)+V^2\sin^2 Kb} & (\varepsilon \geq V). \end{cases} \tag{5.42}$$

Here, $\varepsilon = mv^2/2$, $\kappa = \sqrt{2m(V-\varepsilon)}/\hbar$, and $K = \sqrt{2m(\varepsilon-V)}/\hbar$, where $\hbar$ is the reduced Planck's constant, m is the electron mass, and v is the group velocity of the incident electrons through the third channel. b is the barrier length of the scattering potential, which is chosen to be one-third of the supercell because the length of the Wigner-Seitz cell of Si and C atoms in SiC bulk is one-sixth of the supercell along the $\langle 11\bar{2}0 \rangle$ direction. The calculated barrier heights V are listed in Table 5.5. It is found that all oxygen-related structures behave as a potential barrier.

The scattering mechanism can be understood in terms of the modulation of the CBE of SiC near the interface. As mentioned in Ref. [107], the electrostatic potential at the tetrahedral interstitial site surrounded by four Si atoms (Si tetrahedral structure) is low because of the electron transfer from Si to C. However, when an O atom is inserted between the Si-C bond at the interface, the electrostatic potential at the Si tetrahedral structure is shifted upward owing to the strong electronegativity of O. Since the Si tetrahedral structure appears in the cubic-stacking region, the band gap of the h type, where the ISSs appear slightly below the interface, is widened locally around the oxygen interstitial. For the k type, where the cubic-stacking region starts from the second SiC bilayer, the ISSs are insensitive to the insertion of oxygen in the interface. Therefore, the transmission through the h type is decreased, while that through the k type is almost unchanged.

The transmissions through the h and k types with carbon-related defects are illustrated in Figs. 5.8(c)

Table 5.5: Transmission probabilities of the third channel and barrier heights calculated using Eq. (5.42). The energies of the incident wave are $\varepsilon_{CBE} + 0.7$ eV and $\varepsilon_{CBE} + 1.0$ eV, where ε_{CBE} is the energy of the CBE.

	Model	Transmission probability		Barrier height (V)	
		$\varepsilon_{CBE} + 0.7$ eV	$\varepsilon_{CBE} + 1.0$ eV	$\varepsilon_{CBE} + 0.7$ eV	$\varepsilon_{CBE} + 1.0$ eV
h-type	O_{if}	0.30	0.52	1.48	1.30
	O_{if+sub}	0.48	0.72	1.15	0.97
	$V_C O_2$	0.33	0.34	1.41	1.64
	$(C\text{-}C_i)_C$	0.11	0.14	2.16	2.30
	CC	0.14	0.69	1.99	1.02

and 5.8(d), respectively. Similarly to the case of the oxygen-related structures, the transmission at the h type is markedly decreased. The scattering at the h type occurs in the third channel, which can be explained by the negatively charged C atoms from the carbon-related defects and the existence of the ISS distributed near the interface. Note that a decrease in the transmission of $(C\text{-}C_i)_C$ at the k type is also observed, while the scattering of the CC is marginal. From the charge density distributions of the CBE state at the k type (not shown here), hybridization between the C=C π^* state and the topmost ISSs of the k type is found, indicating that the scattering at the k type originates from the hybridization effect. On the other hand, from the charge density distributions, I know that hybridization between the C=C π^* state and the ISSs of the h type is small. However, the transmission of the h type is smaller than that of the k type due to the effect of the ISSs.

To compare the density of the ISSs at the interface with the sheet electron density in the inversion layer, I solve the Poisson equation using typical conditions, i.e., an accepter density of 1.0×10^{-16} cm^{-3} and a temperature of 500 K. The sheet electron density is on the order of 10^{13} cm^{-2}, which is larger than the defect density ($\sim 10^{11}$ cm^{-2}) calculated using the experimental interface state density and the dispersion of the density in energy. On the other hand, the density of atoms at the SiC(0001) face is 2.44×10^{15} cm^{-2}, indicating that the density of the ISSs, which scatter electrons, is on the order of 10^{14} cm^{-2} when I assume that the oxygen-related structures investigated here evenly appear during dry oxidation. Therefore, I can conclude that the ISSs play a prominent role in the scattering at the SiC-MOSFET when the interface state density is sufficiently reduced because the oxygen-related structures are usually formed by dry oxidation.

5.5.4   Short summary

In summary, I investigated the electron transmission through the CBE state at the two types of 4H-SiC/SiO$_2$ interface, i.e., the h and k types, to clarify the atomistic origin of the mobility degradation in n-channel MOSFETs. The results show that oxygen-related structures at the h type lead to electron scattering, which is counterintuitive because these structures are believed to be electrically inactive.
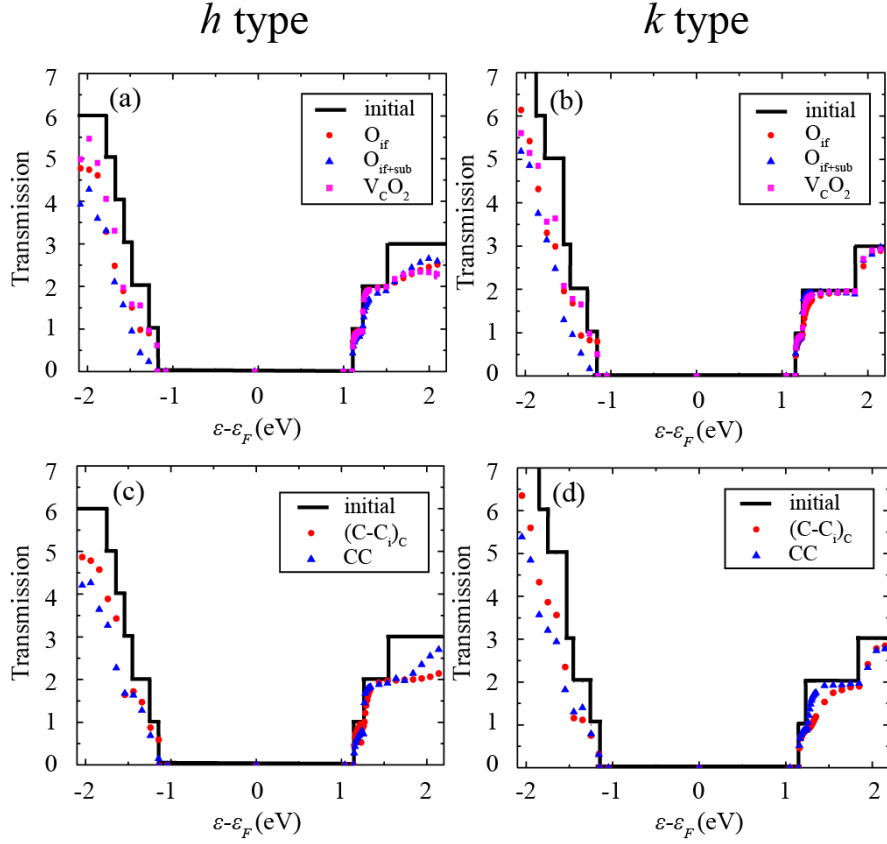
Figure 5.8: Transmission spectra for (a) oxygen-related structures in h type, (b) oxygen-related structures in k type, (c) carbon-related defects in h type, (d) carbon-related defects in k type. The horizontal axis is the energy relative to the Fermi energy ε_F defined as the center of the band gap. The vertical axis is the total transmission probability.

Two physical phenomena combine to prevent electron transmission in the h type. First, the ISSs appear from the top of the interface in the h type. Second, the energy level of the ISSs is shifted upward by the Coulomb interaction with inserted O atoms or defects. The electron scattering by carbon-related defects was also examined. Interestingly, the contribution of the ISSs to the electron scattering is greater than that of the electrically active states of carbon-related defects. Since the existence of both interfaces has been proven by transmission electron microscopy, these phenomena likely occur at the 4H-SiC(0001)/SiO$_2$ interface, resulting in low channel mobility. Further improvement of the SiC-MOSFET will require the consideration of the relationship between the ISSs at the interface and the crystal orientation of SiC as well as the decrease in the interface state density.
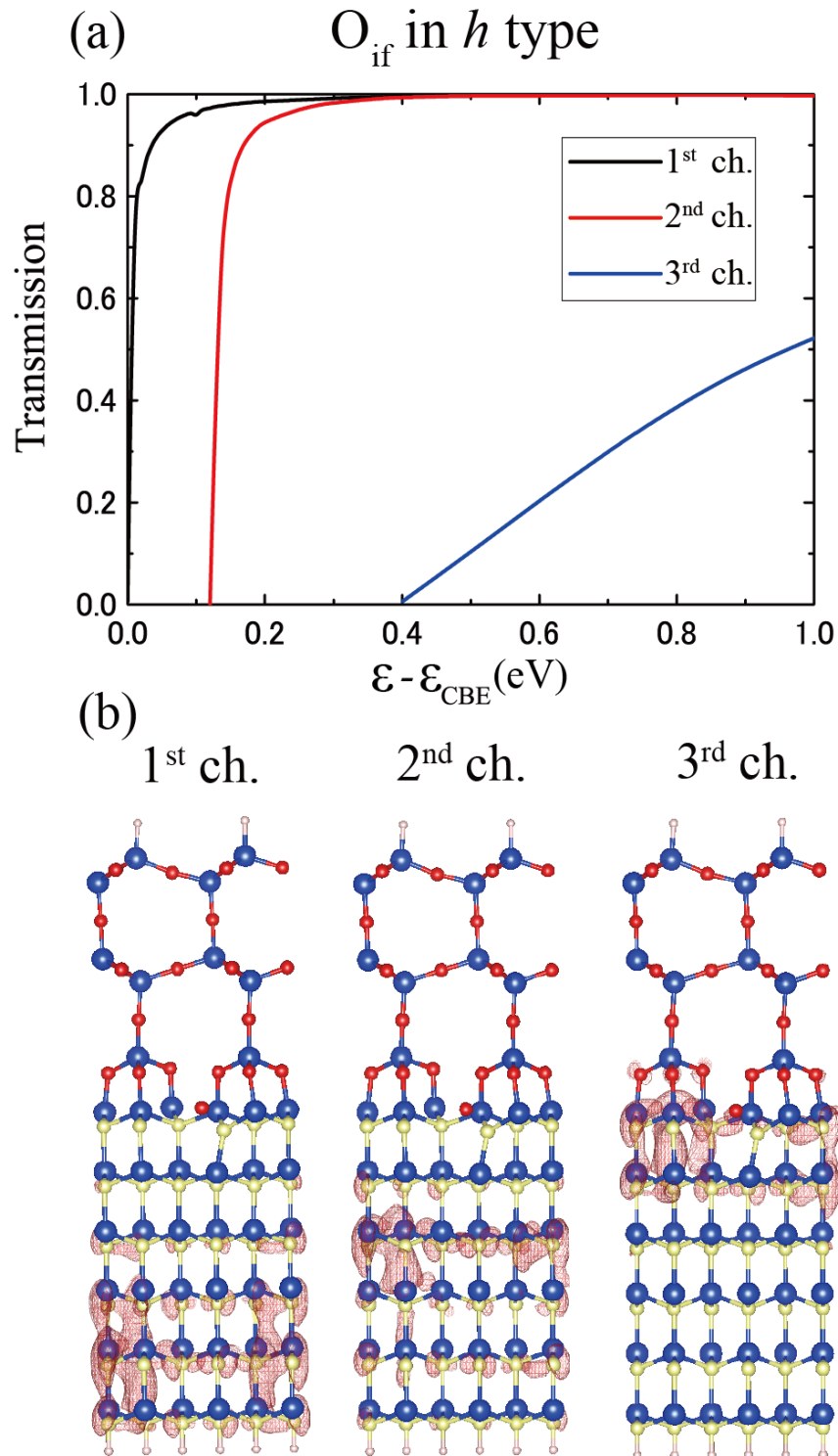
Figure 5.9: (a) Channel transmission and (b) spatial distributions of scattering wave function for eigenchannel. In (b), the case for O_{if} in h type at 1 eV from the CBE is shown. Channels are labeled in descending order of transmission probability.

Chapter 6

Contour integral method for evaluating the self-energy matrices

In Chap. 4, I have introduced the RSFD implementation to compute the self-energy matrices. Combining the partitioning and singular value decomposition technique, I derived the contracted form of the Hamiltonian matrix whose dimension is significantly smaller than the original Hamiltonian matrix. Nevertheless, it still requires inverting the matrix repeatedly or solving the quadratic eigenvalue problem for the dense matrix. In this chapter, I present the alternative approach to evaluate the generalized Bloch states, complex band structure, and self-energy matrices in order to reduce the computational cost of the first-principles transport calculation. The key concept of this approach is to construct the self-energy matrices by propagating and moderately decaying/growing waves via WFM method [48, 68], which makes it feasible to apply the novel algorithm based on the contour integral method to obtain such important waves efficiently. In addition, the proposed algorithm attains not only speed-up and memory reduction but also an excellent parallel efficiency compared with the conventional methods developed so far. The numerical tests for various electrode materials demonstrate that the self-energy matrices are evaluated more efficiently than simply applying the standard eigensolver while keeping with the moderate accuracy. The results and discussions in this chapter are based on my published papers [122, 123].

6.1 Modified WFM method

In this section, I introduce a modified WFM method [48, 124] for handling one of the most computationally expensive parts of the real-space transport calculations, namely, the computation of the self-energy matrices. If one applies the semi-analytical or WFM method, the most computationally demanding part is determining the generalized Bloch states, i.e., solving the quadratic eigenvalue problem (QEP)

$$[-\lambda_n^{-1}\hat{B}^\dagger + (\varepsilon\hat{I} - \hat{A}) - \lambda_n\hat{B}]\phi_n = 0, \tag{6.1}$$

Table 6.1: Number of right-going waves that satisfy $10^{-8} \leq |\lambda_n|^l \leq 1$ for several electrode materials as a function of the number of unit cells l. The Fermi energy is used as an input energy, and all calculations are performed by OBM method using QZ algorithm.

Material	#solutions	$l = 1$	$l = 2$	$l = 3$	$l = 4$	$l \to \infty$
Au chain	21632	102	17	6	4	1
Al(100) wire	51200	334	83	38	20	7
(6,6)CNT	41472	997	278	126	77	2
Graphene	7168	99	32	14	11	2
Silicene	24576	64	19	13	11	2

for a given input energy ε, where $\lambda_n = e^{ik_n a}$. Here, $\hat{A}$ and $\hat{B}$ are $M \times M$ matrices with M being the total number of real-space grid points in the unit cell. The solutions with $|\lambda_n| = 1$ correspond to the propagating waves, whereas remaining states with $|\lambda_n| \neq 1$ are decaying or growing evanescent waves. It is important to note that a large majority of the generalized Bloch states in a real-space grid decay completely within one unit cell of the electrode, and therefore, they contribute little to the electron transport. Table 6.1 shows the number of right-going waves that satisfy $10^{-8} \leq |\lambda_n|^l \leq 1$ for several electrode materials as a function of the number of unit cells l. The Fermi energy is used as an input energy. In all systems, the number of right-going waves satisfying the above criterion is relatively small compared to the total number of nontrivial solutions even when $l = 1$. From the Bloch ansatz, several studies [68,125] assume that the majority of right-going waves decay so fast that they contribute negligibly to the electron transport if 1–2 unit cells are added as a buffer layer. The left-going waves behave in exactly the same manner as the right-going waves because their eigenvalues are pairwise.

Consequently, the self-energy matrices are expected to be well approximated by a relatively small number of propagating and moderately decaying waves that correspond to the solutions of Eq. (6.1) with λ_n being close to the unit circle in the complex plane, that is,

$$\lambda_{min} \leq |\lambda_n| \leq \lambda_{min}^{-1}, \tag{6.2}$$

where λ_{min} is the radius of the inner circle in Fig. 6.1. If λ_{min} is set to a reasonably small value, the transport calculations remain accurate [48,68,124]. Actually, as shown in the section later, the results obtained using the approximated self-energy matrices that remove fast-decaying evanescent waves are visibly indistinguishable from those obtained using exact ones, and they reproduce the previous plane-wave transport calculations accurately. In addition, it is possible to obtain the exact self-energy matrices from the approximated ones construed only by propagating and moderate evanescent waves by employing the recursive Green's function method [126]. Thus, it is enough to find the eigenpairs satisfying Eq. (6.2), and for this purpose, I employ the Sakurai-Sugiura (SS) method which will be presented in the next section.

$$\left[-\lambda_n^{-1} B^\dagger + (\varepsilon - A) - \lambda_n B\right]\boldsymbol{\phi}_n = 0$$
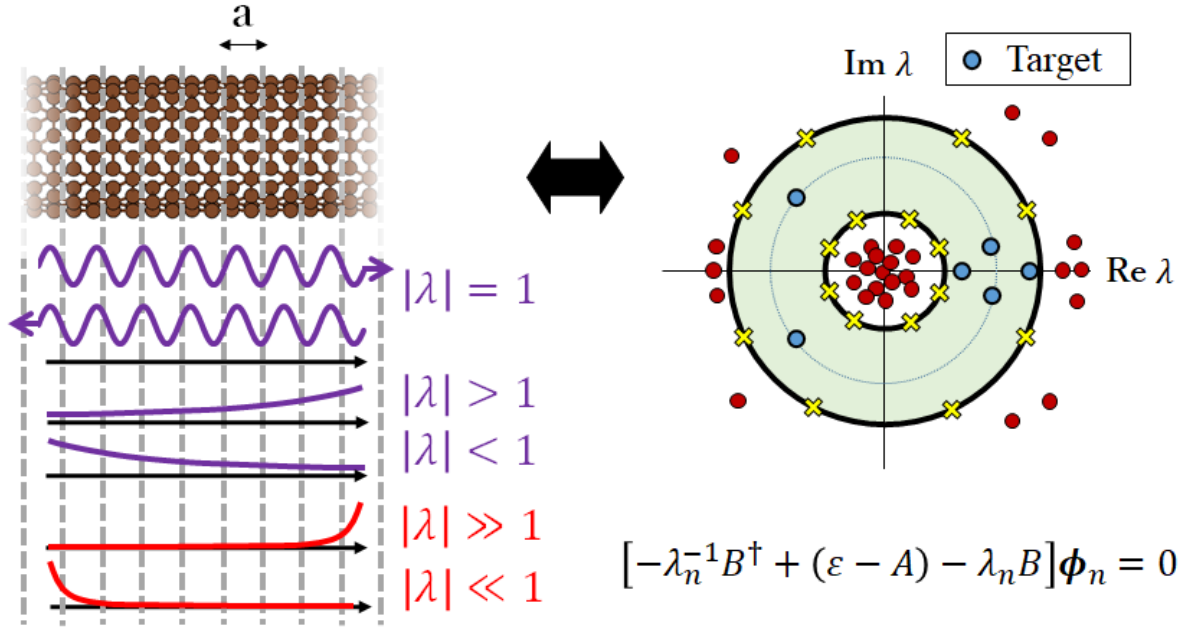
Figure 6.1: Relationship between the generalized Bloch states and solutions of the quadratic eigenvalue problem. Physically important eigenvalues within the green shaded area and the others are plotted as filled blue and open red dots, respectively.

6.2   Sakurai-Sugiura method for quadratic eigenvalue problem

Several efficient eigensolvers for finding the target eigenvalues of Eq. (6.1) located inside the ring-shaped region have been proposed so far. From the viewpoint of the eigensolver, these approaches are classified into two types: (i) shift-and-invert Krylov subspace approach [48] and (ii) contour integral approach based on Sakurai-Sugiura (SS) method [77,122], Polizzi's FEAST method [127,128], and Beyn method [125, 129]. In general, the shift-and-invert Krylov subspace approach is designed for determining eigenvalues close to a given shift, and thus, it is unsuitable for searching all target eigenvalues distributed in a wide range of the complex plane. By contrast, the contour integral approach finds eigenvalues in a given domain, and therefore, it is more appropriate for our target problem. In this subsection, I present the basic algorithm of the SS method and discuss its advantages against the FEAST and Beyn methods.

The complex moment-based eigensolver, proposed by Sakurai and Sugiura in 2003 [77], computes all eigenpairs inside the target region by using the contour integration. Although the original algorithm is developed for the generalized eigenvalue problem, the SS method can also be applied for the nonlinear eigenvalue problem without loss of the efficiency [130].

The SS method consists of two steps. The first step is generating the subspace by contour integration. Let Γ be a contour along the ring-shaped region that encloses the target eigenvalues $\lambda_1, \lambda_2, ..., \lambda_{N_\Gamma}$, where N_Γ is the number of eigenvalues inside Γ. Then, the moment matrix $\hat{S}_p$ associ-

ated with the target eigenpairs is defined as

$$\hat{S}_p = \frac{1}{2\pi i} \oint_\Gamma z^p [\varepsilon \hat{I} - \hat{H}(z)]^{-1} \hat{V} \mathrm{d}z, \quad p = 0, 1, ..., 2N_{mm} - 1 \tag{6.3}$$

where $\hat{V}$ is a $M \times N_{rh}$ nonzero arbitrary matrix. N_{rh} and N_{mm} are the number of right hand sides and the order of moment matrices, respectively. They are input parameters that are set so as to satisfy $N_{rh}N_{mm} > N_\Gamma$. I refer the reader to Ref. [131] for details on how to determine these parameters efficiently. Here,

$$\hat{H}(\lambda) = \lambda^{-1}\hat{B}^\dagger + \hat{A} + \lambda\hat{B}. \tag{6.4}$$

In the actual calculation, $\hat{S}_p$ is approximated by N_{int} points of numerical integration as

$$\hat{S}_p \approx \sum_{j=1}^{N_{int}} w_j z_j^p [\varepsilon \hat{I} - \hat{H}(z_j)]^{-1} \hat{V}, \tag{6.5}$$

where z_j and w_j are quadrature points and weights, respectively, which are determined by a trapezoidal rule.

The second step is extracting the eigenpairs from $\hat{S}_p$. For the nonlinear eigenvalue problem, this procedure can be performed by several ways including complex moment-based (SS-Hankel) method [130], Rayleigh-Ritz projection (SS-RR) method [132], and communication-avoiding Arnoldi-type (SS-CAA) method [133]. In this study, both SS-Hankel and SS-CAA methods are implemented. I first present the basic algorithm of the SS-Hankel method following the procedure given in Ref. [130]. The SS-Hankel method uses the complex moment matrix $\hat{\mu}_p = \hat{V}^\dagger \hat{S}_p$, and the target eigenvalues are extracted by solving the $N_{mm}N_{rh}(\ll M)$-dimensional generalized eigenvalue problem

$$\hat{T}^< \boldsymbol{x}_i = \tau_i \hat{T} \boldsymbol{x}_i, \tag{6.6}$$

with $N_{mm}N_{rh} \times N_{mm}N_{rh}$ Hankel matrices $\hat{T}$ and $\hat{T}^<$ defined as

$$\hat{T}^< = \begin{bmatrix} \hat{\mu}_1 & \hat{\mu}_2 & \cdots & \hat{\mu}_{N_{mm}} \\ \hat{\mu}_2 & \hat{\mu}_3 & \cdots & \hat{\mu}_{N_{mm}+1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mu}_{N_{mm}} & \hat{\mu}_{N_{mm}+1} & \cdots & \hat{\mu}_{2N_{mm}-1} \end{bmatrix}, \tag{6.7}$$

and

$$\hat{T} = \begin{bmatrix} \hat{\mu}_0 & \hat{\mu}_1 & \cdots & \hat{\mu}_{N_{mm}-1} \\ \hat{\mu}_1 & \hat{\mu}_2 & \cdots & \hat{\mu}_{N_{mm}} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mu}_{N_{mm}-1} & \hat{\mu}_{N_{mm}} & \cdots & \hat{\mu}_{2N_{mm}-2} \end{bmatrix}. \tag{6.8}$$

From the viewpoint of numerical stability and efficiency, one should compute the rank N_Γ of $\hat{T}$ by a singular value decomposition

$$\hat{T} = [\hat{U}_{T1}, \hat{U}_{T2}] \begin{bmatrix} \hat{\Sigma}_{T1} & \hat{O} \\ \hat{O} & \hat{\Sigma}_{T2} \end{bmatrix} \begin{bmatrix} \hat{W}_{T1}^\dagger \\ \hat{W}_{T2}^\dagger \end{bmatrix} \approx \hat{U}_{T1}\hat{\Sigma}_{T1}\hat{W}_{T1}^\dagger. \tag{6.9}$$

Here the term $\hat{U}_{T2}\hat{\Sigma}_{T2}\hat{W}_{T2}^\dagger$ is omitted because $\hat{\Sigma}_{T2} \approx 0$. Upon substituting Eq. (6.8) into Eq. (6.6), the QEP is reduced to N_Γ-dimensional standard eigenvalue problem, that is,

$$\hat{U}_{T1}^\dagger \hat{T}^< \hat{W}_{T1} \hat{\Sigma}_{T1}^{-1} \boldsymbol{y}_i = \tau_i \boldsymbol{y}_i, \tag{6.10}$$

where $\boldsymbol{y}_i = \hat{\Sigma}_{T1}\hat{W}_{T1}^\dagger \boldsymbol{x}_i$. The (approximated) eigenpairs are obtained as $(\lambda_n, \boldsymbol{\phi}_n) = (\tau_n, \hat{S}\hat{W}_{T1}\hat{\Sigma}_{T1}^{-1}\boldsymbol{y}_n)$, where $\hat{S} = [\hat{S}_0, \hat{S}_1, \ldots, \hat{S}_{N_{mm}-1}]$. In the algorithm, Eq. (6.10) is used instead of Eq. (6.6). If there are too many eigenvalues inside the contour, the ring-shaped region should be divided into several subdomains to reduce the cost of solving Eq. (6.10). In this case, it might be better to set a subdomain to a ring-shaped region, i.e., the target region is sliced as an onion because the number of eigenvalues located in each slice should be the same, owing to the pairwise relationship $(\lambda_n, \lambda_n^{*-1})$. The algorithm of the SS-Hankel method is illustrated in Table 6.2.

Table 6.2: SS-Hankel method for quadratic eigenvalue problem, $[-\lambda_n^{-1}\hat{B}^\dagger + (\varepsilon\hat{I} - \hat{A}) - \lambda_n\hat{B}]\boldsymbol{\phi}_n = 0$, with complex matrices $\hat{A}, \hat{B} \in \mathbb{C}^{M\times M}, \boldsymbol{\phi}_n \in \mathbb{C}^M, \lambda_n, \varepsilon \in \mathbb{C}$.

1: Input $N_{rh}, N_{mm}, N_{int} \in \mathbb{N}, \delta \in \mathbb{R}, \hat{V} \in \mathbb{C}^{M\times N_{rh}}, (z_j, \omega_j)$ for $j = 1, 2, \ldots, N_{int}$

2: Purpose: Obtain N_Γ approximate eigenpairs $(\lambda_n, \boldsymbol{\phi}_n)$

3: Compute $\hat{S}_p = \sum_{j=1}^{N_{int}} \omega_j z_j^k [\varepsilon\hat{I} - \hat{H}(z_j)]^{-1} V$ and $\hat{\mu}_p = \hat{V}^\dagger \hat{S}_p$

4: Set $\hat{S} = [\hat{S}_0, \hat{S}_1, \ldots, \hat{S}_{N_{mm}-1}]$ and block Hankel matrices $\hat{T}^<$ and $\hat{T}$

5: Perform a singular value decomposition for $\hat{T}$ using the threshold δ:
   $\hat{T} = [\hat{U}_{T1}, \hat{U}_{T2}][\hat{\Sigma}_{T1}, \hat{O}; \hat{O}, \hat{\Sigma}_{T2}][\hat{W}_{T1}, \hat{W}_{T2}]^\dagger \approx \hat{U}_{T1}\hat{\Sigma}_{T1}\hat{W}_{T1}^\dagger$

6: Solve $U_{T1}^\dagger \hat{T}^< \hat{W}_{T1} \hat{\Sigma}_{T1}^{-1} \boldsymbol{y}_n = \tau_n \boldsymbol{y}_n$, and compute $(\lambda_n, \boldsymbol{\phi}_n) = (\tau_n, \hat{S}\hat{W}_{T1}\hat{\Sigma}_{T1}^{-1}\boldsymbol{y}_n)$

The SS-CAA method is the novel complex moment-based eigensolver using the block communication-avoiding Arnoldi procedure. As proved in Ref. [133], the subspace $\hat{S}$ is equivalent to the Krylov subspace, i.e.,

$$\hat{S} = \mathcal{K}_{N_{mm}}(\hat{C}, \hat{S}_0) = [\hat{S}_0, \hat{C}\hat{S}_0, ..., \hat{C}^{N_{mm}-1}\hat{S}_0], \tag{6.11}$$

where $\hat{C}$ is not generated explicitly, instead, computed implicitly via the block Arnoldi procedure. Then, the target eigenpairs are obtained by the Arnoldi method with $\mathcal{K}_{N_{mm}}(\hat{C}, \hat{S}_0)$,

$$\hat{C}\boldsymbol{x}_i = \tau_i \boldsymbol{x}_i. \tag{6.12}$$

To obtain the expression of $\hat{C}$, one might perform the QR factorization of the Krylov subspace $\mathcal{K}_{N_{mm}+1}(\hat{C}, \hat{S}_0)$,

$$\mathcal{K}_{N_{mm}+1}(\hat{C}, \hat{S}_0) = [\hat{S}_0, \hat{S}_1, \ldots, \hat{S}_{N_{mm}}] = \hat{Q}_{N_{mm}+1}\hat{R}_{N_{mm}+1}, \tag{6.13}$$

where $\hat{Q}_{N_{mm}+1}$ is $M \times (N_{mm}+1)N_{rh}$ orthogonal matrix and $\hat{R}_{N_{mm}+1}$ is $(N_{mm}+1)N_{rh} \times (N_{mm}+1)N_{rh}$ upper triangle matrix. I also set $\hat{Q}_{N_{mm}} = \hat{Q}_{N_{mm}+1}(1 : M, 1 : N_{mm}N_{rh})$ and $\hat{R}_{N_{mm}} = \hat{R}_{N_{mm}+1}(1 : N_{mm}N_{rh}, 1 : N_{mm}N_{rh})$. The unitary transformation rewrites Eq. (6.12) to

$$\hat{C}'\boldsymbol{x}_i' = \tau_i \boldsymbol{x}_i', \tag{6.14}$$

where $\hat{C}' = \hat{Q}_{N_{mm}}^{\dagger} \hat{C} \hat{Q}_{N_{mm}}$ and $\boldsymbol{x}' = \hat{Q}_{N_{mm}}^{\dagger} \boldsymbol{x}$. Based on the communication-avoiding Arnoldi procedure, $\hat{C}'$ is given by

$$\hat{C}' = \hat{R}_{N_{mm}+1}(1 : N_{mm}N_{rh}, N_{rh} + 1 : N_{mm}N_{rh} + N_{rh})\hat{R}_{N_{mm}}^{-1}. \tag{6.15}$$

In practice, to reduce the computational cost and improve the accuracy, the rank N_{Γ} of $\hat{R}_{N_{mm}}$ is computed by singular value decomposition:

$$\hat{R}_{N_{mm}} = [\hat{U}_{R1}, \hat{U}_{R2}] \begin{bmatrix} \hat{\Sigma}_{R1} & \hat{O} \\ \hat{O} & \hat{\Sigma}_{R2} \end{bmatrix} \begin{bmatrix} \hat{W}_{R1}^{\dagger} \\ \hat{W}_{R2}^{\dagger} \end{bmatrix} \approx \hat{U}_{R1}\hat{\Sigma}_{R1}\hat{W}_{R1}^{\dagger}. \tag{6.16}$$

Substituting Eqs. (6.15) and (6.16) into Eq. (6.14), one might obtain the reduced eigenvalue problem

$$\hat{U}_{R1}^{\dagger}\hat{R}_{N_{mm}+1}(1 : N_{mm}N_{rh}, N_{rh} + 1 : N_{mm}N_{rh} + N_{rh})\hat{W}_{R1}\hat{\Sigma}_{R1}^{-1}\boldsymbol{y}_i = \tau_i\boldsymbol{y}_i, \tag{6.17}$$

where the approximated eigenpairs are obtained as $(\lambda_n, \boldsymbol{\phi}_n) = (\tau_n, \hat{Q}_{N_{mm}}\hat{U}_{R1}\boldsymbol{y}_n)$. The algorithm of the SS-CAA method is illustrated in Table 6.3.

Finally, I briefly mention about the advantage of the SS method compared with the FEAST and Beyn methods. Considering that the FEAST method is employed to solve QEP, the QEP must be transformed into the generalized eigenvalue problem with the twice-sized matrix because the FEAST method is not designed for nonlinear eigenvalue problem. Due to the twice-sized eigenvalue problem, the FEAST method is inefficient. In addition, the FEAST and Beyn methods do not use the higher-order moment matrices $\hat{S}_p$, which requires to solve linear systems with much larger number of right-hand sides than the SS method. Because the computational cost of solving linear systems is proportional to the number of the right-hand side, Beyn method is also inefficient if the Krylov subspace solver is employed.

Table 6.3: SS-CAA method for quadratic eigenvalue problem, $[-\lambda_n^{-1}\hat{B}^{\dagger} + (\varepsilon\hat{I} - \hat{A}) - \lambda_n\hat{B}]\boldsymbol{\phi}_n = 0$, with complex matrices $\hat{A}, \hat{B} \in \mathbb{C}^{M \times M}, \boldsymbol{\phi}_n \in \mathbb{C}^M, \lambda_n, \varepsilon \in \mathbb{C}$.

1: Input $N_{rh}, N_{mm}, N_{int} \in \mathbb{N}, \delta \in \mathbb{R}, \hat{V} \in \mathbb{C}^{M \times N_{rh}}, (z_j, \omega_j)$ for $j = 1, 2, \dots, N_{int}$
2: Purpose: Obtain N_{Γ} approximate eigenpairs $(\lambda_n, \boldsymbol{\phi}_n)$
3: Compute $\hat{S}_p = \sum_{j=1}^{N_{int}} \omega_j z_j^k [\varepsilon\hat{I} - \hat{H}(z_j)]^{-1}\hat{V}$
4: Set $\mathcal{K}_{N_{mm}+1}(\hat{C}, \hat{S}_0) = [\hat{S}_0, \hat{S}_1, \dots, \hat{S}_{N_{mm}}]$
5: Perform a QR factorization $\mathcal{K}_{N_{mm}+1}(\hat{C}, \hat{S}_0) = \hat{Q}_{N_{mm}+1}\hat{R}_{N_{mm}+1}$
6: Perform a singular value decomposition for $\hat{R}_{N_{mm}}$ using the threshold δ: $\hat{R}_{N_{mm}} = [\hat{U}_{R1}, \hat{U}_{R2}][\hat{\Sigma}_{R1}, \hat{O}; \hat{O}, \hat{\Sigma}_{R2}][\hat{W}_{R1}, \hat{W}_{R2}]^{\dagger} \approx \hat{U}_{R1}\hat{\Sigma}_{R1}\hat{W}_{R1}^{\dagger}$
7: Solve $U_{R1}^{\dagger}\hat{R}_{N_{mm}+1}(1 : N_{mm}N_{rh}, N_{rh} + 1 : N_{mm}N_{rh} + N_{rh})\hat{W}_{R1}\hat{\Sigma}_{R1}^{-1}\boldsymbol{y}_n = \tau_n\boldsymbol{y}_n$, and compute $(\lambda_n, \boldsymbol{\phi}_n) = (\tau_n, \hat{Q}_{N_{mm}}\hat{U}_{R1}\boldsymbol{y}_n)$

6.3   Efficient implementation of Sakurai-Sugiura method

The most time-consuming part of the SS method and other contour-integral based eigensolvers [127, 129] are the numerical integration of Eq. (6.5). I herein present several efficient implementations based

on the special structure of the target QEP.

6.3.1   Symmetry between inner and outer integrations
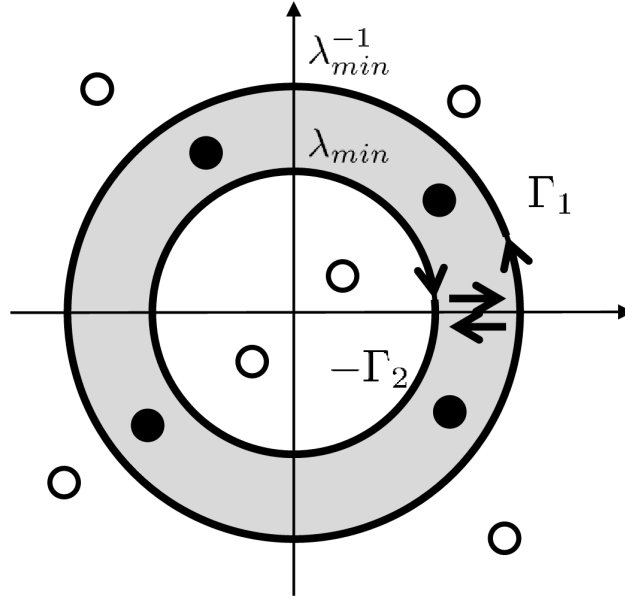
Figure 6.2: Contour path for the target ring-shaped region. The target eigenvalues and the others are shown by ● and ○, respectively.

Because the target eigenvalues are located inside the ring-shaped region surrounded by the inner and outer circles, the counter path can be set as shown in Fig. 6.2 and $\hat{S}_p$ can be split as

$$
\begin{aligned}
\hat{S}_p &= \frac{1}{2\pi i} \oint_{\Gamma_1} z^p [\varepsilon \hat{I} - \hat{H}(z)]^{-1} \hat{V} \,\mathrm{d}z - \frac{1}{2\pi i} \oint_{\Gamma_2} z^p [\varepsilon \hat{I} - \hat{H}(z)]^{-1} \hat{V} \,\mathrm{d}z \\
&\approx \sum_{j=1}^{N_{int}} w_j (z_j^{(1)})^p [\varepsilon \hat{I} - \hat{H}(z_j^{(1)})]^{-1} \hat{V} - \sum_{j=1}^{N_{int}} w_j (z_j^{(2)})^p [\varepsilon \hat{I} - \hat{H}(z_j^{(2)})]^{-1} \hat{V},
\end{aligned}
\tag{6.18}
$$

where $z_j^{(1)}$ and $z_j^{(2)}$ are quadrature points at outer and inner circles, respectively. Thus, it is required

to solve $2 \times N_{int}$ number of linear system, that is,

$$[\varepsilon \hat{I} - \hat{H}(z_j^{(1)})]\hat{Y}_j^{(1)} \quad = \quad \hat{V}, \tag{6.19}$$

$$[\varepsilon \hat{I} - \hat{H}(z_j^{(2)})]\hat{Y}_j^{(2)} \quad = \quad \hat{V}, \tag{6.20}$$

for $j = 1, ..., N_{int}$, because the target region consists of two circles. By using the symmetry of the Hamiltonian matrix, the computational cost can be reduced to half. Because $\hat{A} = \hat{A}^\dagger$, one might have

$$H(z_j^{(2)}) = H(z_j^{(1)})^\dagger. \tag{6.21}$$

Therefore, Eq. (6.20) can be regarded as the dual system of Eq. (6.19), i.e.,

$$[\varepsilon \hat{I} - H(z_j^{(1)})]^\dagger \hat{Y}_j^{(2)} = \hat{V}. \tag{6.22}$$

As noted in Appendix B, the BiCG method can solve both systems simultaneously with very little additional computational cost. Note that some other linear solvers including the direct solvers can solve dual systems efficiently.

6.3.2   Multiple energy calculations: shifted BiCG method

Basically, first-principles electron transport calculations must be conducted independently at each energy point. Therefore the total computational cost for determining the self-energy matrices is proportional to the number of energy points, if dual systems, Eqs. (6.19) and (6.22), are solved independently. However, by using the shift-invariant property of the Krylov subspace, it is possible to reduce the cost substantially. The essence of this approach is that matrix vector operation, the most time-consuming step of the Krylov subspace method, is performed only at the reference energy point and the solutions at all energy points are updated with moderate computational costs. In this case, the method is called the shifted BiCG method [81], and its details are presented in Appendix. C.

6.3.3   Modified shifted BiCG method for evaluating complex band structure

Although a significant reduction of the computational cost of the SS method will be achieved by using shifted BiCG method for solving shifted linear systems, its benefit fades out as increasing the number of energy points. This is because that the computational complexity of the (sparse) matrix-vector operation is $\mathcal{O}(cM)$, where cM is the nonzero element of $M \times M$ Hamiltonian matrix, while the scalar-vector product of the shifted BiCG method scales as $\mathcal{O}(N_s M)$, where N_s is the number of the energy points. Thus, if $N_s M \geq cM$, a cost to update the vectors, a large amount of scalar-vector products, becomes comparable to a cost of construct the Krylov subspace on a seed.

To reduce the cost of the scalar-vector product of the shifted BiCG method, I consider the efficient implementation of the SS-Hankel method [134]. From Eqs. (6.5), (6.8), and (6.9), the Hankel matrix is constructed by the complex moment matrices,

$$\hat{\mu}_p = \hat{V}^\dagger \hat{S}_p \approx \sum_{j=1}^{N_{int}} w_j z_j^p \hat{\Xi}(\varepsilon). \tag{6.23}$$

Here the reduced solution matrix is defined as

$$\hat{\Xi}(\varepsilon) = \hat{V}^{\dagger}[\varepsilon\hat{I} - \hat{H}(z_j)]^{-1}\hat{V}. \tag{6.24}$$

Note that $\hat{\Xi}(\varepsilon) \in \mathbb{C}^{N_{rh} \times N_{rh}}$ does not need to be computed explicitly at each iteration, but can be updated by the recurrences for reduced quantities. Because $[\varepsilon\hat{I} - \hat{H}(z_j)]$ is a general complex matrix, the modified shifted BiCG method is appropriate to compute the reduced solution vectors. The algorithm of the modified shifted BiCG method is written in Table 6.4. Since the only reduced vectors with dimension $N_{rh}(\ll M)$ appear in the algorithm, the computational cost of scalar-vector products on each energy shift becomes significantly smaller than those for the full-dimensional vectors.

The proposed algorithm is favorable especially for the complex band structure (CBS) calculation because the eigenvalues can be obtained only from the Hankel matrix which is constructed by reduced matrix. As for the self-energy matrices, eigenvectors, i.e., generalized Bloch states, on shift energy points are needed. This requires the solution vectors of linear equations at left and right matching planes.

6.4   Accuracy of the Sakurai-Sugiura method
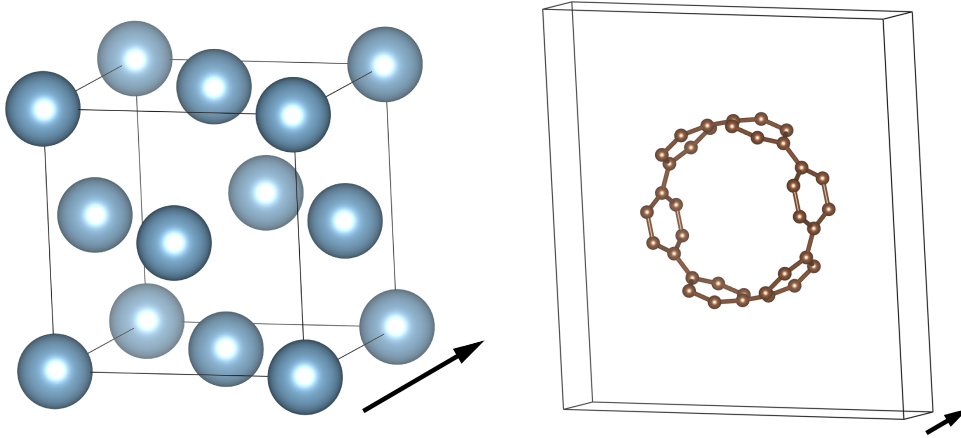
Figure 6.3: Unit cells of Al(100) and (6,6)CNT. The black arrow indicates the z-axis, i.e., transport direction.

To evaluate the accuracy of the proposed method, I calculate complex k values as a function of real energy ε. The dispersion relationship between complex k and ε is so-called CBS, which is useful to investigate the static and dynamical electronic properties of materials. Because wavevector k which satisfies $|\lambda| = 1$ corresponds to the standard band structure, plotting the CBS allows us to compare with the band structure obtained from the electronic structure calculation for periodic system. I employ fcc Aluminum (Al) and (6,6) armchair carbon nanotube (CNT) with 4 and 24 atoms per unit cell, respectively. The unit cell of two models are shown in Fig. 6.3. The number of grid points,

Table 6.4: Modified shifted BiCG algorithm for $[\hat{A} + \sigma\hat{I}]\boldsymbol{x}_n(\sigma) = \boldsymbol{b}$ and $[\hat{A}^\dagger + \sigma\hat{I}]\tilde{\boldsymbol{x}}_n(\sigma) = \boldsymbol{b}$, with a general complex matrix $\hat{A} \in \mathbb{C}^{M \times M}, \hat{V} \in \mathbb{C}^{M \times N_{rh}}, \sigma \in \mathbb{C}, \boldsymbol{x}_n, \boldsymbol{p}_n, \boldsymbol{r}_n, \tilde{\boldsymbol{x}}_n, \tilde{\boldsymbol{p}}_n, \tilde{\boldsymbol{r}}_n \in \mathbb{C}^M, \boldsymbol{\xi}_n(\sigma), \boldsymbol{\theta}_n(\sigma), \boldsymbol{\rho}_n, \tilde{\boldsymbol{\xi}}_n(\sigma), \tilde{\boldsymbol{\theta}}_n(\sigma), \tilde{\boldsymbol{\rho}}_n \in \mathbb{C}^{N_{rh}}$, and $\alpha_n, \beta_n, \alpha_n(\sigma), \beta_n(\sigma), \pi_n(\sigma) \in \mathbb{C}$.

1: Input σ_j $(j = 1, 2, ..., N_s)$
2: Set $\boldsymbol{x}_0 = \tilde{\boldsymbol{x}}_0 = \boldsymbol{p}_{-1} = \tilde{\boldsymbol{p}}_{-1} = 0$, $\boldsymbol{r}_0 = \tilde{\boldsymbol{r}}_0 = \boldsymbol{b}$, $\boldsymbol{\rho}_0 = \tilde{\boldsymbol{\rho}}_0 = \hat{V}^T\boldsymbol{b}$, $\alpha_{-1} = 1$, and $\beta_{-1} = 0$
3: Set $\boldsymbol{\xi}_0(\sigma_j) = \tilde{\boldsymbol{\xi}}_0(\sigma_j) = \boldsymbol{\theta}_{-1}(\sigma_j) = \tilde{\boldsymbol{\theta}}_{-1}(\sigma_j) = 0$
4: Set $\alpha_{-1}(\sigma_j) = \pi_{-1}(\sigma_j) = \pi_0(\sigma_j) = 1$, and $\beta_{-1}(\sigma_j) = 0$ $(j = 1, 2, ..., N_s)$
5: For $n = 0, 1, 2, ...,$ until convergence do:
6: 　　$\alpha_n = (\boldsymbol{r}_n, \boldsymbol{r}_n)/(\boldsymbol{p}_n, \hat{A}\boldsymbol{p}_n)$
7: 　　$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \alpha_n\boldsymbol{p}_n$
8: 　　$\tilde{\boldsymbol{x}}_{n+1} = \tilde{\boldsymbol{x}}_n + \bar{\alpha}_n\tilde{\boldsymbol{p}}_n$
9: 　　$\boldsymbol{r}_{n+1} = \boldsymbol{r}_n - \alpha_n\hat{A}\boldsymbol{p}_n$
10: 　　$\tilde{\boldsymbol{r}}_{n+1} = \tilde{\boldsymbol{r}}_n - \bar{\alpha}_n\hat{A}^\dagger\tilde{\boldsymbol{p}}_n$
11: 　　$\beta_n = (\boldsymbol{r}_{n+1}, \boldsymbol{r}_{n+1})/(\boldsymbol{r}_n, \boldsymbol{r}_n)$
12: 　　$\boldsymbol{p}_{n+1} = \boldsymbol{r}_n + \beta_n\boldsymbol{p}_n$
13: 　　$\tilde{\boldsymbol{p}}_{n+1} = \tilde{\boldsymbol{r}}_n + \bar{\beta}_n\tilde{\boldsymbol{p}}_n$
14: 　　$\boldsymbol{\rho}_{n+1} = \hat{V}^\dagger\boldsymbol{r}_{n+1}$
15: 　　$\tilde{\boldsymbol{\rho}}_{n+1} = \hat{V}^\dagger\tilde{\boldsymbol{r}}_{n+1}$
16: 　　For $j = 1, 2, ..., N_s$ do:
17: 　　　　$\pi_{n+1}(\sigma) = \left(1 + \frac{\beta_{n-1}\alpha_n}{\alpha_{n-1}} - \alpha_n\sigma\right)\pi_n(\sigma) - \frac{\beta_{n-1}\alpha_n}{\alpha_{n-1}}\pi_{n-1}(\sigma)$
18: 　　　　$\alpha_n(\sigma_j) = \frac{\pi_n(\sigma_j)}{\pi_{n+1}(\sigma_j)}\alpha_n$
19: 　　　　$\beta_n(\sigma_j) = \left(\frac{\pi_n(\sigma_j)}{\pi_{n+1}(\sigma_j)}\right)^2\beta_n$
20: 　　　　$\boldsymbol{\xi}_{n+1}(\sigma_j) = \boldsymbol{\xi}_n(\sigma_j) + \alpha_n(\sigma_j)\boldsymbol{\theta}_n(\sigma_j)$
21: 　　　　$\boldsymbol{\theta}_{n+1}(\sigma_j) = 1/\pi_n(\sigma_j)\boldsymbol{\rho}_n + \beta_n(\sigma_j)\boldsymbol{\theta}_n(\sigma_j)$
22: 　　　　$\tilde{\boldsymbol{\xi}}_{n+1}(\sigma_j) = \tilde{\boldsymbol{\xi}}_n(\sigma_j) + \bar{\alpha}_n(\sigma_j)\tilde{\boldsymbol{\theta}}_n(\sigma_j)$
23: 　　　　$\tilde{\boldsymbol{\theta}}_{n+1}(\sigma_j) = 1/\bar{\pi}_n(\sigma_j)\tilde{\boldsymbol{\rho}}_n + \bar{\beta}_n(\sigma_j)\tilde{\boldsymbol{\theta}}_n(\sigma_j)$
24: 　　End do:
25: End do

$M_x \times M_y \times M_z$, are $20 \times 20 \times 20$ and $72 \times 72 \times 12$, respectively. The transport direction is taken along z-axis, which is parallel to the $\langle 100 \rangle$ direction (the nanotube axis) in the case of Al (CNT). I set $N_{int} = 32, N_{mm} = 8, N_{rh} = 16, \delta = 10^{-10}$. I employ the BiCG method to solve the linear systems in the contour integral methods and set the convergence criteria by 10^{-10}. It is important to mention that I also tried other linear solvers including BiCGstab, BiCGGR, and GMRES methods for the target linear systems; however they fail to converge to the criteria within the realistic iteration numbers. The Kohn-Sham Hamiltonian matrices are obtained from the real-space pseudopotential DFT code RSPACE [85, 86]. All calculations in this and later section are performed by local density approximation [12] and norm-conserving pseudopotentials proposed by Troullier and Martins [24]. The forth-order finite-difference approximation ($N = 4$ in Ref. [75]) is employed for the Laplacian

operator. Figure 6.4 shows the CBSs for Al(100) and (6,6) CNT. The black dots are results computed by SS-Hankel method and red curves are conventionally calculated band structures. For real k region, the black dots are in good agreement with the standard band structure (red curve).

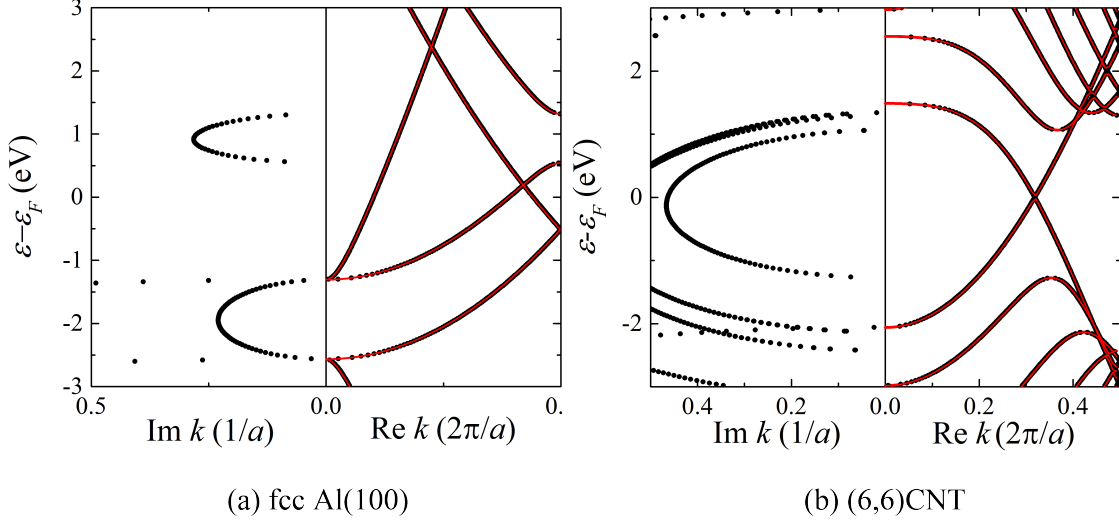(a) fcc Al(100)   (b) (6,6)CNT

Figure 6.4: Complex band structure for (a) Al(100) and (b) (6,6)CNT. The black dots indicate the numerical results obtained by the proposed method. The red curves show the conventionally calculated band structure for comparison.

Next, I evaluate the performance of the SS-Hankel (Table 6.2) and SS-CAA (Table 6.3) methods as comparison with the Beyn method [129]. In the numerical experiment, I use $N_{int} = 32$ and $\delta = 10^{-10}$ for three methods and set $N_{rh} = 16$ for the SS-Hankel and SS-CAA methods, while $N_{rh} = 128$ is used for the Beyn method so as to construct the same subspace dimension. In this study, the residual 2-norms $\|r_n\|_2 = \|[\varepsilon \hat{I} - \hat{H}(\lambda_n)]\phi_n\|_2$ is used for checking the accuracy of the computed eigenpairs. Note that the generalized Bloch states are normalized so as to satisfy $\|\phi_n\|_2 = 1$. Calculations are performed using the single node of Xeon-Phi^{TM} 7250 (code name: Knights Landing) on Oakforest-PACS.

Figure 6.5 shows the residual norms of (6,6)CNT computed by three methods. It can be seen that the Beyn method shows the best accuracy and the SS-CAA method is more accurate than the SS-Hankel method. This tendency is in accordance with the numerical tests in Ref. [133]. In all methods, the accuracies deteriorate from outer to inner circle. This numerical error mainly comes from the round-off error of the moment matrices $\hat{S}_p$ because $\hat{S}_p$ at inner and outer contours are large and small values, respectively, from Eq. (6.5). Thus, the higher-order moment matrices does not have the information of the eigenpairs near the inner circle, leading the accuracy deterioration. The maximum and minimum residual norms and elapsed time are presented in Table 6.5. Here, T_{SLE}, T_{Eig}, and T_{Total} are elapsed times of solving linear equations and constructing moment matrices, performing singular value decomposition and solving the reduced eigenvalue problem, and whole computation, respectively. From Table 6.5, the accuracy deterioration is observed clearly in SS-Hankel and SS-CAA

methods. On the other hand, the SS-Hankel method is the best regarding the elapsed time, and SS-CAA method is slightly slower than the SS-Hankel method. Because the T_{SLE} is the dominant part of the total elapsed time and linearly increases with N_{rh}, the Beyn method is almost 8 times slower than the SS-Hankel and SS-CAA methods. Based on this result, the SS-CAA method is the best balanced method in terms of the accuracy and computational cost.

In the actual calculations, parameters should be chosen such that all residues become less than a given criteria. Table 6.5 also shows the breakdowns of the three methods when modulating the number of right-hand side N_{rh}. T_{Total} (N_{rh}) of SS-Hankel, SS-CAA, and Beyn methods whose maximum residuals become less than 10^{-8} are 1438.22 sec. (24), 715.67 sec. (12), and 2862.30 sec. (48), respectively. I thus conclude that the SS-CAA method is roughly twice and four times faster than SS-Hankel and Beyn method, respectively.

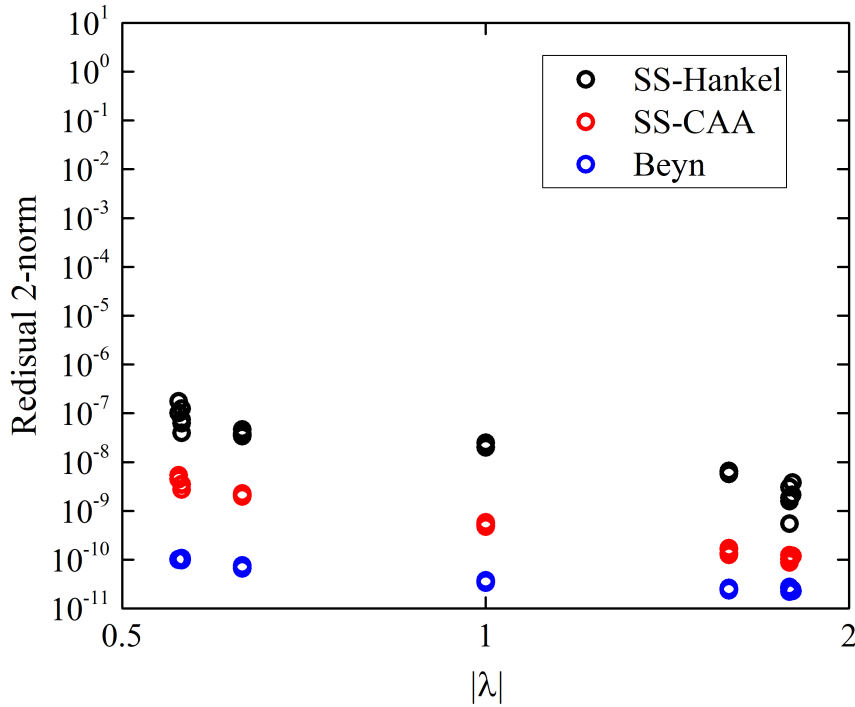Figure 6.5: Residual norms of (6,6) CNT computed by SS-Hankel, SS-CAA, and Beyn methods. I set $N_{int} = 32, \delta = 10^{-10}, \lambda_{min} = 0.5$ for all methods. $N_{rh} = 16$ and $N_{mm} = 8$ are used for SS-Hankel and SS-CAA methods while $N_{rh} = 128$ is used for the Beyn method to span the same dimension of the subspace. The number of eigenvalues excluding the spurious eigenvalues is 24.

Table 6.5: Breakdown of the SS-Hankel, SS-CAA, and Beyn methods for (6,6)CNT.

Parameters		Elapsed time[sec.]			Residual 2-norm	
$\#N_{rh}$	$\#N_{mm}$	T_{SLE}	T_{Eig}	T_{Total}	$\max\|\boldsymbol{r}_n\|_2$	$\min\|\boldsymbol{r}_n\|_2$
SS-Hankel method						
24	8	1438.22	5.14	1443.36	8.46×10^{-09}	1.93×10^{-10}
16	8	957.94	3.45	961.39	1.74×10^{-07}	5.45×10^{-10}
SS-CAA method						
16	8	960.67	9.70	970.37	5.39×10^{-09}	8.71×10^{-11}
12	8	715.67	5.89	721.56	7.86×10^{-09}	1.41×10^{-10}
Beyn method						
128	–	7631.01	5.51	7636.52	1.03×10^{-10}	2.27×10^{-11}
64	–	3810.99	2.58	3813.57	2.47×10^{-10}	4.24×10^{-11}
48	–	2862.30	2.11	2864.41	2.39×10^{-9}	1.89×10^{-10}
32	–	1906.06	1.45	1907.51	1.48×10^{-6}	3.50×10^{-7}

6.5 Parallel implementation

Even though using the efficient techniques presented the above, solving N_{int} dual linear systems with N_{rh} multiple right-hand side is still heavy task. Another important way to reduce the computational cost is to use different number of processors, that is, to use parallel computing. For this purpose, I introduce three layers of hierarchical parallelism of SS method as shown in Fig. 6.7. Note that communication between layers is negligible. The details of each layer parallelism are summarized as below.

- **Top layer parallelism: Multiple right-hand sides**
  In the top layer, N_{rh} right-hand sides of linear systems are solved in parallel. Because it is possible to solve linear systems with different right-hand side independently, this parallelism requires no communication. In addition, it is expected that the convergence of the shifted BiCG method does not strongly depend on right-hand sides, which leads the good load balancing.

- **Middle layer parallelism: Quadrature points**
  As the dual systems (6.19) and (6.22) are independent of z_j (j-th quadrature point), one can independently solve these linear systems in N_{int} parallel without communication. While middle layer parallelism also requires no communication; however, one need to take care of load balancing due to the imbalance of the convergence of the shifted BiCG method. In general, the convergency of the shifted BiCG method becomes slow when the quadrature points are close to the eigenvalues. To achieve good load balancing, it is recommended to use the following two stopping conditions for the BiCG method.

  – Relative residual 2-norm becomes less than certain stopping criteria. (This is a standard

stopping criteria.)

– The BiCG method is stopped at over half of quadrature points. (This is used to achieve good load balancing.)

Fortunately, the convergence behavior does not strongly depend on the quadrature points as shown in Fig. 6.6. This uniform convergency guarantees the accuracy of the SS method when using two stopping criteria.
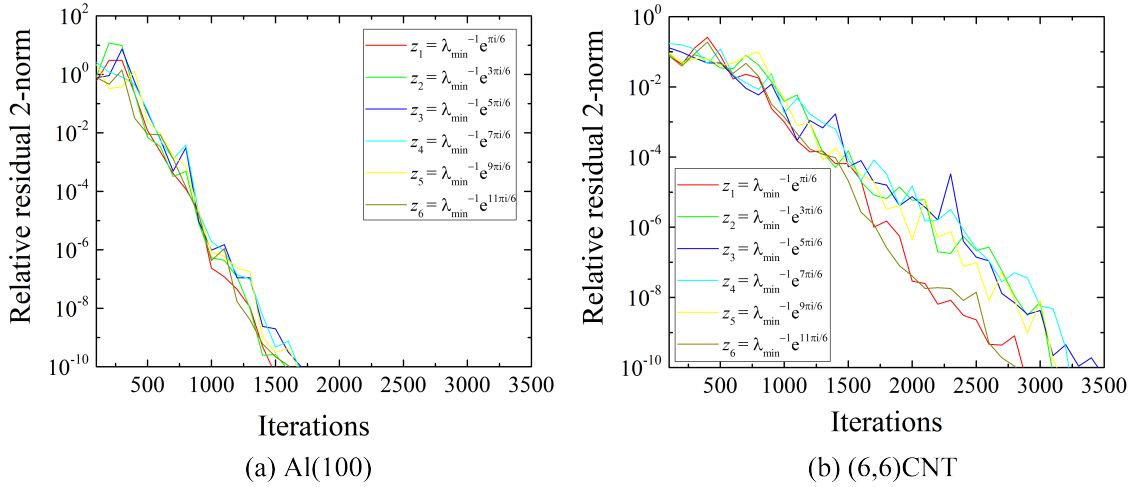

(a) Al(100) (b) (6,6)CNT

Figure 6.6: Convergence behavior of the BiCG method for (a) Al(100) and (b) (6,6)CNT at $\varepsilon = \varepsilon_F$. The figure shows the residual norms as a function of the number of iteration at each quadrature point z_j.

- **Bottom layer parallelism: Domain decomposition technique**
  The real-space grid representation of the Hamiltonian matrix is highly sparse and easily parallelized by the domain decomposition technique. Although this parallelism requires communication in matrix-vector multiplications and inner products per BiCG iteration, it is known that high parallel performance is obtained when the number of grid points is sufficiently large [135].

The total parallelism N_{total} is

$$N_{total} = N_{dm} \times N_{int} \times N_{rh},$$

where N_{dm} is the number of processors assigned for the domain decomposition. If the number of processors available is less than $N_{int} \times N_{rh}$, one should use top layer parallelism first, because upper layer is expected to show better scalability than lower layers.

Top layer: Multiple right-hand sides are solved in parallel
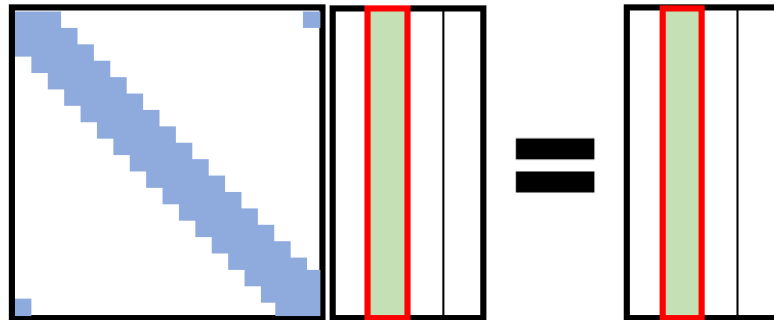
Middle layer: Quadrature points are computed in parallel
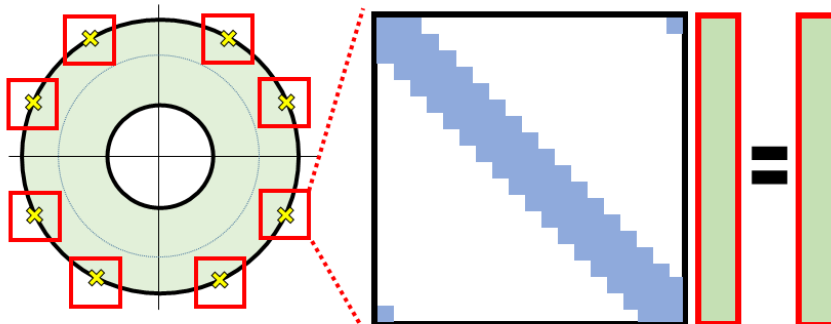
Bottom layer: Linear equation is solved in parallel

1D domain decomposition

Figure 6.7: Hierarchical parallelism of the Sakurai-Sugiura method used in this study.

6.6   Parallel performance

The computational cost of the SS method mainly depends on the part that solves the linear equations (6.20) and (6.22), as shown in Table 6.5. Consequently, to evaluate the parallel performance of the method, I parallelized only this part of the code by using OpenMP directives and the Intel Message Passing Interface (MPI) library. As mentioned in the previous section, the three layers of hierarchical parallelis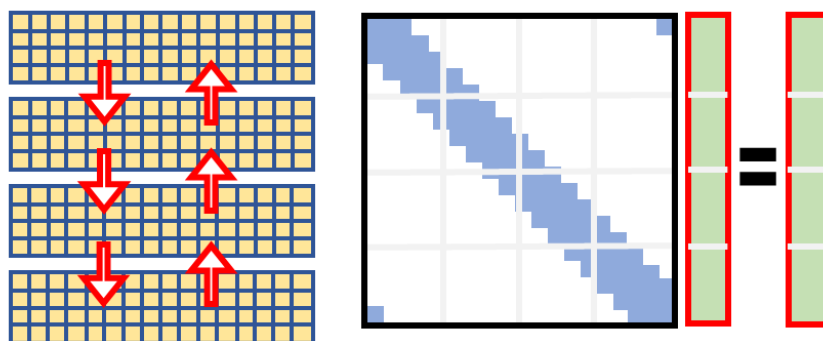m of the SS method, i.e., parallelisms for multiple right-hand sides (top layer), quadratic points (middle layer), and the domain decomposition (bottom layer) were introduced.

All calculations in this subsection were performed on Oakforest-PACS. Each computation node is an Intel Xeon Phi$^{\text{TM}}$ 7250; each node has 68 cores (1.4 GHz) and 96 GB of system memory. I here conducted numerical experiments on three different (8,0) CNTs with 32, 1024, and 10240 atoms per unit cell. The computational models of the pristine (8,0) CNT with 32 atoms, the boron- and nitrogen-doped (8,0) CNT (BN-doped (8,0) CNT) with 1024 atoms, and BN-doped (8,0) CNT with 10240 atoms are shown in Fig. 6.8. The BN-doped (8,0) CNT was made by randomly inserting boron and nitrogen into pristine (8,0) CNT.
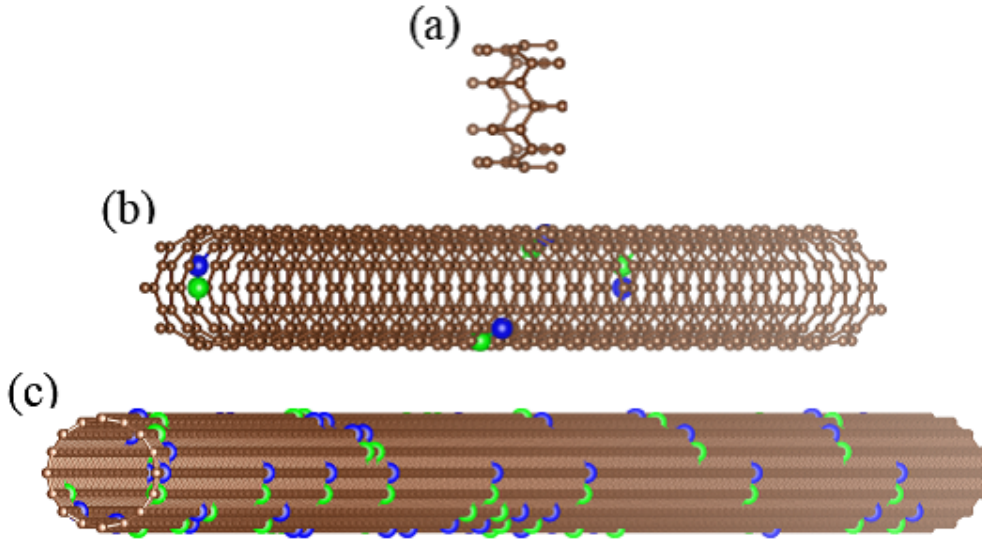
Figure 6.8: Schematic diagrams of (a) pristine (8,0) CNT, (b) BN-doped (8,0) CNT with 1024 atoms, and (c) BN-doped (8,0) CNT with 10240 atoms. Carbon, boron, and nitrogen are depicted as brown, green, and blue balls, respectively. The BN-doped (8,0) CNTs are made by randomly inserting boron and nitrogen into a pristine (8,0) CNT.

6.6.1   Scalability in small system

The system that I first tested was a (8,0) CNT with 32 atoms (the number of grid points is $72 \times 72 \times 20$). I set $N_{int} = 32, N_{mm} = 8, N_{rh} = 64, \delta = 10^{-10}$, and $\lambda_{min} = 0.5$. The convergence criteria for the BiCG method was set to 10^{-10}. One MPI process was assigned to each node, which allowed us to use 68

OpenMP threads per MPI process. Figure 6.9 shows strong scaling of the three layers of parallelism. The total time to solve the linear equations and the time to solve the remaining part are shown on the left. Note that the remaining part does not include the disk I/O to write the solutions. The right graph in Fig. 6.9 shows the speed-ups in solving the eigenvalue problem and linear equations.

Figure 6.9(a) shows the runtime and actual speed-up at the top layer of parallelism, where two MPI processes were assigned to the middle layer and the number of processes in the top layer was varied from 1 to 64. As I expected, the time of the remaining part is negligibly small compared with the time to solve the linear equations. The method achieved almost ideal scaling at the top layer because the linear equations with different right-hand sides can be solved in parallel without communication and very good load-balancing can be achieved. It should be noted that the total runtime for the small system decreased from 14392 to 234 seconds when the number of processes for the top layer increased from 1 to 64.

The middle-layer scalability is shown in Fig. 6.9(b), where two MPI processes were assigned to the top layer and the number of processes in the middle layer was varied from 1 to 32. The parallel efficiency of the middle layer is slightly lower than that of the top layer, although the computations are almost independent. The degradation of scalability at the middle layer comes from the difference in convergence behaviors of the BiCG procedure at each quadrature point and it becomes more significant as the number of processes assigned to the middle layer increases. Nevertheless, the strong scaling was almost linear and a speed-up of about 21 times was achieved when I assigned 32 MPI processes to the middle layer.

Figure 6.9(c) shows the bottom-layer scalability; here, two MPI processes were assigned to the middle layer and the number of processes in the bottom layer was varied as 1, 2, 4, 8, and 16, where the corresponding domain decompositions, $n_x \times n_y \times n_z$, were $1 \times 1 \times 1$, $1 \times 1 \times 2$, $1 \times 1 \times 4$, $2 \times 1 \times 4$, and $2 \times 2 \times 4$, respectively. Here n_x, n_y, and n_z are the number of domains in the x, y, and z directions, respectively. The bottom layer scalability based on a domain-decomposition technique is much worse than the top or middle layer scalability because of frequent communications between processes in every BiCG iteration. For the small system, the poor scaling in the bottom layer is not serious because parallelization using only the top and middle layers is enough to reduce the computation time.

I also considered how to divide the cores in the node among the OpenMP and bottom layer parallelism because it is very difficult to take full advantage of the OpenMP scalability on a many-core processor and the parallel resources are usually limited to the specific number. Table 6.6 shows the elapsed times of 1000 BiCG iterations for (8,0) CNT with 32 atoms by fixing the number of cores to 64 and varying the number of OpenMP threads and N_{dm}. In the small system, the best performance was obtained when the 16 threads were assigned to OpenMP and four MPI processes were assigned to the bottom layer.

(a) Top layer parallelism

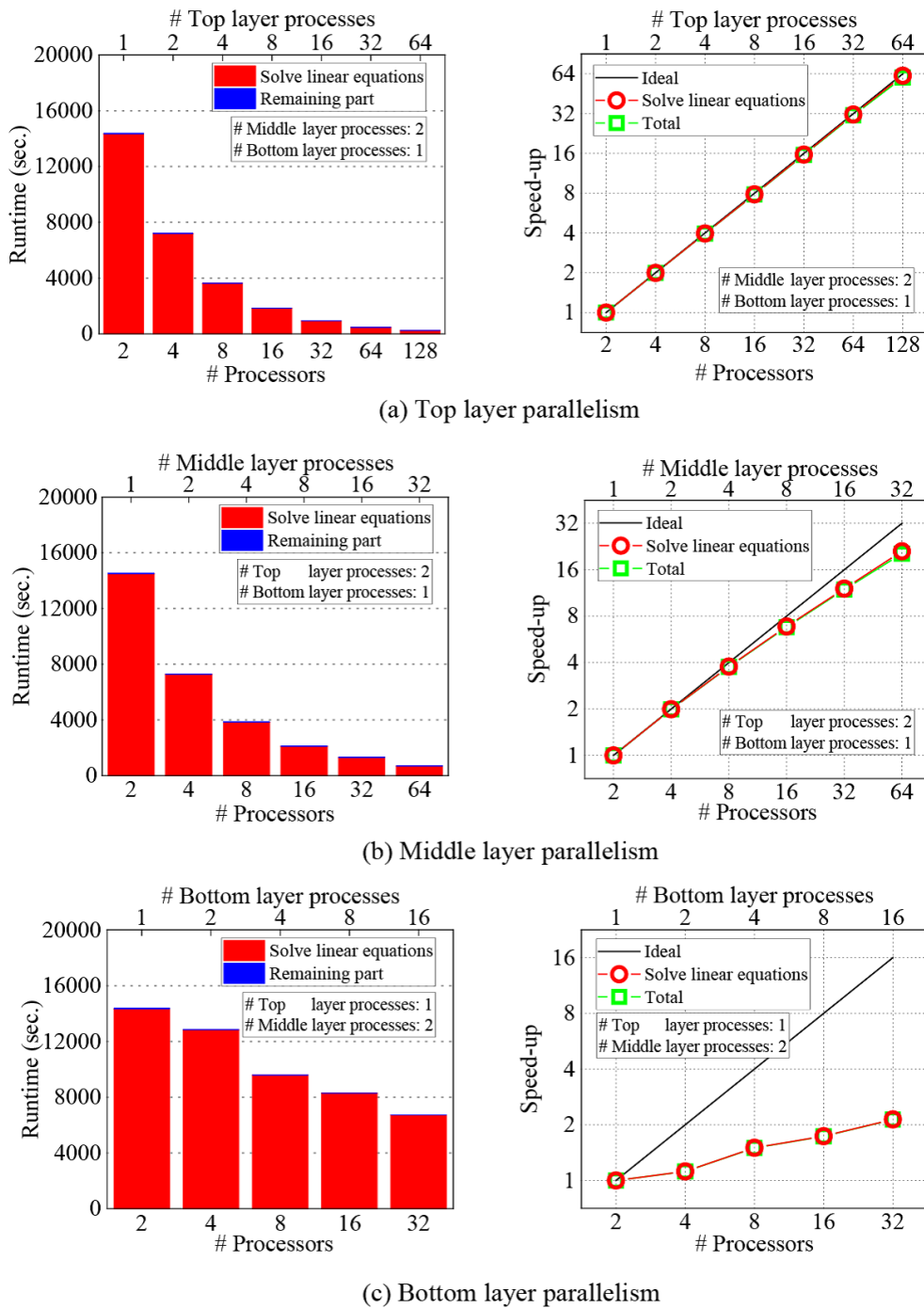(b) Middle layer parallelism

(c) Bottom layer parallelism

Figure 6.9: Scalability of three layers of parallelism for (8,0) CNT with 32 atoms. 68 OpenMP threads were assigned to each MPI process.

Table 6.6: Parallel performance inside the node. Elapsed times of 1000 iterations of the BiCG procedure for the (8,0) CNT with 32 atoms and (8,0) BN-doped (8,0) CNTs with 1024 and 10240 atoms were measured by fixing the total number of cores and splitting their allocation between the OpenMP and bottom layer parallelism.

|		Elapsed time [sec.]		
# OpenMP	# N_{dm}	(8,0)CNT	BN-doped CNT (1024 atoms)	BN-doped CNT (10240 atoms)
1	64	7.77	104.95	795.42
2	32	6.78	90.37	776.35
4	16	5.18	**84.77**	**774.75**
8	8	4.50	86.32	811.43
16	4	**3.98**	96.02	916.12
32	2	5.19	118.12	1132.11
64	1	6.16	161.24	1486.64

6.6.2 Scalability in medium-sized system

Next, I tested the (8,0) BN-CNT with 1024 atoms (the number of grid points is $72 \times 72 \times 640$). I set $N_{int} = 32, N_{mm} = 8, N_{rh} = 16, \delta = 10^{-10}$, and $\lambda_{min} = 0.5$. The convergence criteria for the BiCG method was set to 10^{-10}. Four MPI processes were assigned to each node, which allowed us to use 17 OpenMP threads per MPI process. Figure 6.10(a) shows the runtime and actual speed-up at the top layer parallelism, where 32 and 4 MPI processes were assigned to the middle and bottom layers, respectively, and the number of processes in the top layer was varied from 1 to 16. The middle-layer scalability is shown in Fig. 6.10(b), where 16 and 4 MPI processes were assigned to the top and bottom layer and the number of processes in the middle layer was varied from 1 to 32. The bottom-layer scalability is shown in Fig. 6.10(c), where 16 and 32 MPI processes were assigned to the top and middle layer, and the number of processes in the bottom layer was varied from 1 to 16. The domain decomposition was performed at the grid points along the z direction to minimize communications and achieve better load-balancing. As shown in Figure 6.10, the top- and middle-layer performances are similar to the results for the small system; i.e., the top layer has the almost ideal scaling and the middle-layer scalability is slightly lower than the top-layer one. In contrast with the case of the small system, good scalability is obtained at the bottom layer. In Table 6.6, I can see that the computational time of 1000 BiCG iterations increases almost linearly relative to the number of atoms, which indicates that the communications per iteration decreases and a domain-decomposition becomes more and more efficient as the number of atoms increases. Figure 6.10(c) shows the CBS calculation using 2048 nodes (139264 cores) of Oakforest-PACS, i.e., 25 % of total nodes. Even when using 2048 nodes, the method scales favorably, and the total time needed to solve the eigenvalue equation is reduced to about 905 seconds. Table 6.6 shows the elapsed times of 1000 BiCG iterations for the BN-doped (8,0) CNT with 1024 atoms with 64 cores, while varying the number of OpenMP threads and N_{dm}. The best
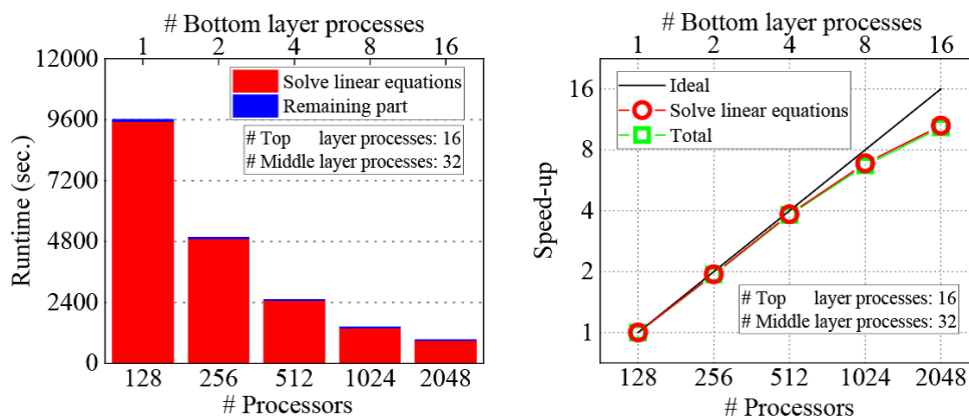
performance was obtained when the four threads were assigned to OpenMP and 16 MPI processes were assigned to the bottom layer.
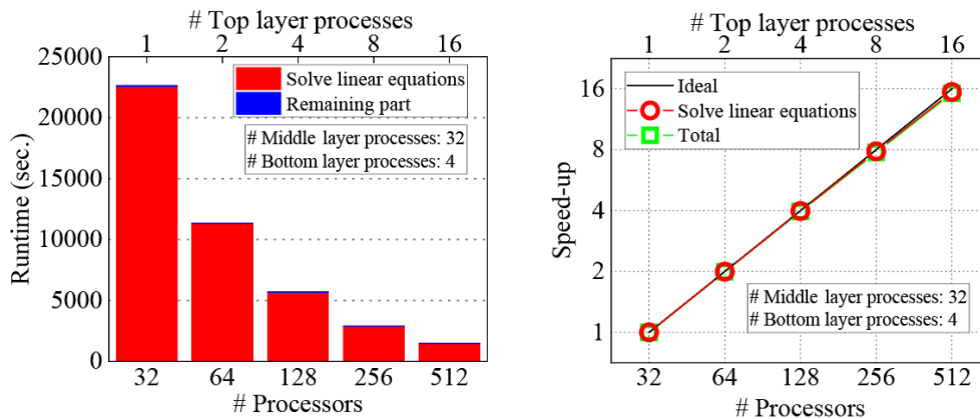
(a) Top layer parallelism
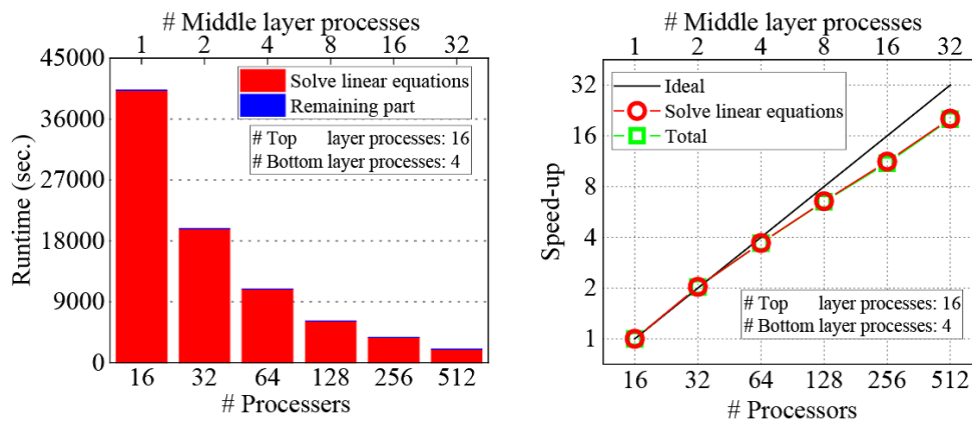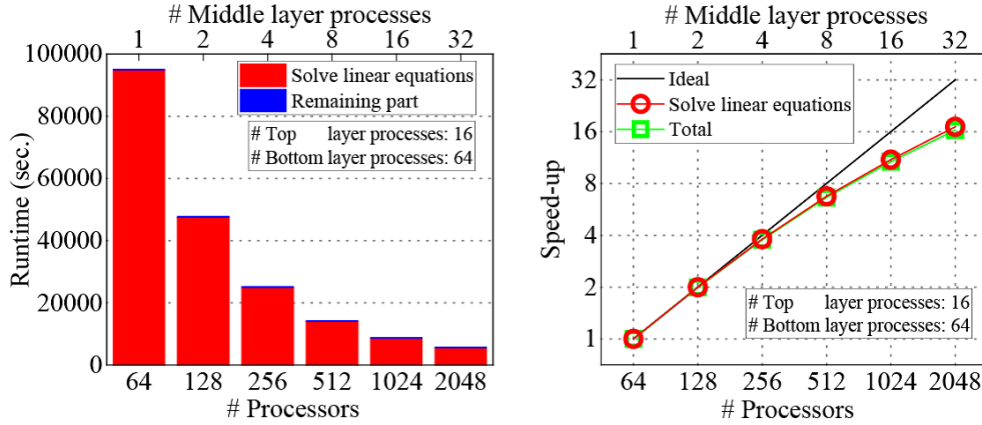
(b) Middle layer parallelism
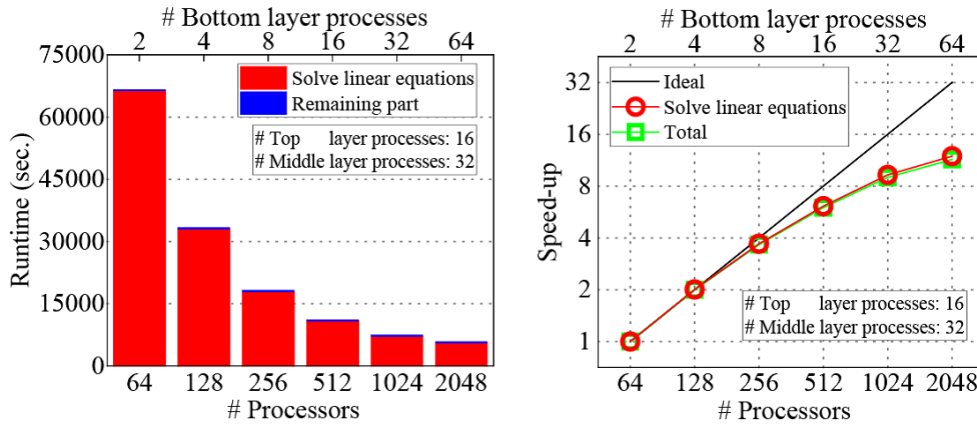
(c) Bottom layer parallelism

Figure 6.10: Scalability of three layers of parallelism for BN-doped (8,0) CNT with 1024 atoms. 17 OpenMP threads were assigned to each MPI process.

6.6.3   Scalability in large system

Finally, I investigated the parallel performance on the (8,0) BN-CNT with 10240 atoms ($72 \times 72 \times 6400$ grid points). The top layer scalability is omitted due to lack of space. I set $N_{int} = 32, N_{mm} = 8, N_{rh} = 16, \delta = 10^{-10}$, and $\lambda_{min} = 0.5$. The convergence criteria for the BiCG method was set to 10^{-10}. 16 MPI processes were assigned to each node, which allow us to use four OpenMP threads per MPI process. The domain decomposition was performed at the grid points along the z direction. Figure 6.11(a) shows the runtime of the middle layer parallelism, where 16 and 64 MPI processes were assigned to the top and bottom layer, respectively, and the number of processes for the middle layer was varied from 1 to 32. The bottom layer scalability is shown in Fig. 6.11(b), where 16 and 32 MPI processes were assigned to the top and middle layers and the number of processes in the bottom layer was varied from 2 to 64. The reduced efficiency at the large number of MPI process at bottom layer is caused by the computational cost for the global communication in the operations of nonlocal pseudopotential-vector products, which can be reduced by replacing it to local communication. Although I still need further tuning of the code, the proposed method is efficient and scalable to a large number of processors. I also note that the CBS calculations of BN-doped CNT with 10240 atoms can be executed in 2 hours using 25 % of the computational power of Oakforest-PACS. Table 6.6 shows the elapsed times of 1000 BiCG iterations for the BN-doped (8,0) CNT with 10240 atoms on 64 cores while varying the number of OpenMP threads and N_{dm}. The best performance was obtained when four threads were assigned to OpenMP and 16 MPI processes were assigned to the bottom layer.

(a) Middle layer parallelism

(b) Bottom layer parallelism

Figure 6.11: Scalability of middle and bottom layers of parallelism for BN-doped (8,0) CNT with 10240 atoms. Four OpenMP threads were assigned to each MPI process.

6.7 Variable conversion from λ space to k space

In this subsection, I focus on the numerical difficulty of the SS method presented in the previous sections. In general, the contour integral approach generates a subspace spanned by eigencomponents inside a given region and extracts target eigenpairs from this subspace. However, When $\lambda_{min} \ll 0.1$, the subspace quality degrades owing to a significant round-off error, which comes from the large difference of the absolute values of the quadrature points at inner and outer circles. Consequently, the information obtained from the contour integrations along the inner circle will not be properly contained, and accuracy degradation occurs from the outer circle toward the inner circle. To avoid this numerical difficulty arising from the explicit computation of eigenvalues in the λ plane, I convert the variable space from the λ plane to the k plane, as shown in Fig. 6.12. The variable conversion, $k = \ln \lambda / ia$, replaces the ring-shaped region with the rectangular region, as shown in Fig. 6.12(b). In

addition, the QEP for λ [Eq. (6.1)] is replaced by exponential-type eigenvalue problem (EEP) for k [Eq. (3.66)]. The contour integration along the rectangular region can be performed without suffering from the round-off error because the integration points always take moderate values.
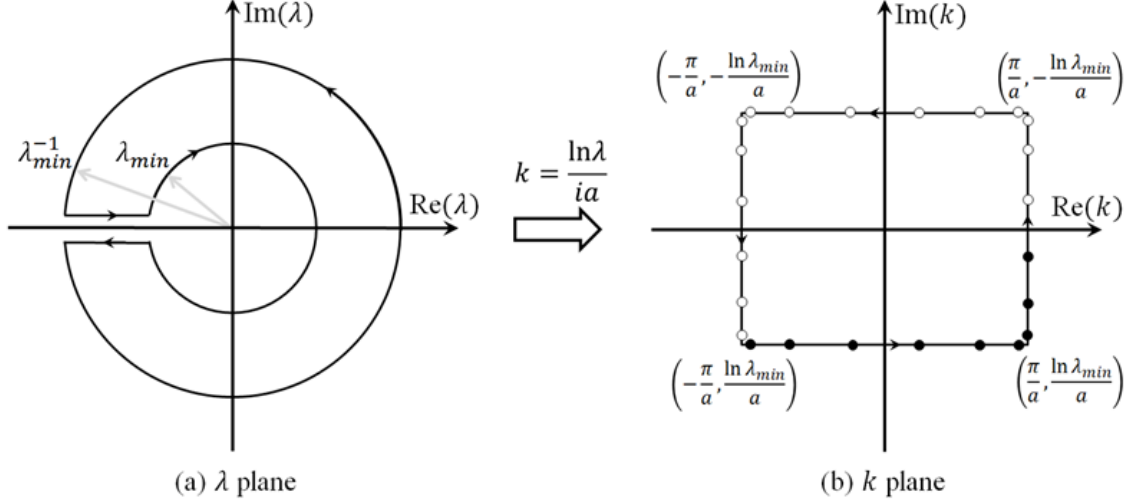

(a) λ plane   (b) k plane

Figure 6.12: Two equivalent contours in complex λ and k planes

Let Γ be a counterclockwise contour along each side of the rectangular region. Then, Eq. (6.3) can be rewritten as the sum of four definite integrals:

$$\hat{S}_p = \hat{S}_p^{(1)} + \hat{S}_p^{(2)} + \hat{S}_p^{(3)} + \hat{S}_p^{(4)}, \tag{6.25}$$

where

$$\hat{S}_p^{(1)} = \frac{1}{2\pi i} \int_{-\pi/a}^{\pi/a} \left(x + i\frac{\ln \lambda_{min}}{a} \right)^p \left[\varepsilon \hat{I} - \hat{H}\left(x + i\frac{\ln \lambda_{min}}{a} \right) \right]^{-1} \hat{V} dx, \tag{6.26}$$

$$\hat{S}_p^{(2)} = \frac{1}{2\pi i} \int_{\ln \lambda_{min}/a}^{-\ln \lambda_{min}/a} \left(\frac{\pi}{a} + iy \right)^p \left[\varepsilon \hat{I} - \hat{H}\left(\frac{\pi}{a} + iy \right) \right]^{-1} \hat{V} i dy, \tag{6.27}$$

$$\hat{S}_p^{(3)} = \frac{1}{2\pi i} \int_{\pi/a}^{-\pi/a} \left(x - i\frac{\ln \lambda_{min}}{a} \right)^p \left[\varepsilon \hat{I} - \hat{H}\left(x - i\frac{\ln \lambda_{min}}{a} \right) \right]^{-1} \hat{V} dx, \tag{6.28}$$

$$\hat{S}_p^{(4)} = \frac{1}{2\pi i} \int_{-\ln \lambda_{min}/a}^{\ln \lambda_{min}/a} \left(-\frac{\pi}{a} + iy \right)^p \left[\varepsilon \hat{I} - \hat{H}\left(-\frac{\pi}{a} + iy \right) \right]^{-1} \hat{V} i dy. \tag{6.29}$$

Here, $\hat{H}(k)$ is redefined as

$$\hat{H}(k) = e^{-ika} \hat{B}^\dagger + \hat{A} + e^{ika} \hat{B}. \tag{6.30}$$

The N_q-point Gauss-Legendre quadrature rule is applied to evaluate the definite integrals. Here, N_q is a pair of polynomial orders, that is, $N_q = (N_{q1}, N_{q2})$. Figure 6.12(b) shows $N_{q1} = 6$ quadrature points along the Re(k) axis and $N_{q2} = 6$ quadrature points along the Im(k) axis. The total number of quadrature points is $N_{int} = 2N_{q1} + 2N_{q2}$. As with the calculation on λ plane, one need to solve

linear systems with multiple right-hand sides at each quadrature point. By using the symmetry of the Hamiltonian matrix $H(k)$, the number of linear systems to be solved can be reduced to $N_{q1} + \frac{1}{2}N_{q2}$. If time-reversal symmetry holds,

$$[\varepsilon \hat{I} - \hat{H}(z_j)]^\dagger = \varepsilon \hat{I} - \hat{H}(z_j^*). \tag{6.31}$$

Equation (6.31) suggests that linear systems with $\mathrm{Im}(z_j) > 0$ are adjoints of linear systems with $\mathrm{Im}(z_j) < 0$. As noted in the previous section, the BiCG method can solve both systems simultaneously with very little additional computational cost. Furthermore, owing to the translational symmetry, $e^{\mathrm{i}ka} = e^{\mathrm{i}(ka+2\pi)}$ holds, and this leads to

$$\varepsilon \hat{I} - \hat{H}\Big(-\frac{\pi}{a} + z_j \Big) = \varepsilon \hat{I} - \hat{H}\Big(\frac{\pi}{a} + z_j \Big). \tag{6.32}$$

It is clear from Eq. (6.32) that the linear systems in $S_p^{(2)}$ and $S_p^{(4)}$ are the same; thus, one only need to solve either one of them. From the above, the numerical integration can be performed by solving the linear systems indicated by black dots in Fig. 6.12(b).

6.8 Numerical experiments on k space.

In this subsection, I demonstrate the numerical accuracy and robustness of the SS method on k space through a series of test calculations. In all cases, Γ-point sampling in the two-dimensional Brillouin zone is used. Unless noted otherwise, a grid spacing of 0.38 bohr is used.

6.8.1 Accuracy of eigenpairs inside Γ

First, I confirm the accuracy of the SS method described in the previous subsection. Besides the Hamiltonian matrix and inner radius λ_{min}, the SS method requires several other parameters: the order of Gauss-Legendre quadrature rule $N_q = (N_{q1}, N_{q2})$, the number of right-hand sides N_{rh}, and the order of moment matrices N_{mm}. It is essential to select these parameters appropriately to make the algorithm robust and efficient. Among these parameters, N_{rh} and N_{mm} are also used in the conventional SS method, and their effects on numerical errors have been studied elsewhere. Thus, it is expected that the general principles can be applied to N_{rh} and N_{mm} [136]. Care must be taken when selecting N_q because the SS method features an ellipsoid-type contour, with numerical integration performed using the trapezoidal rule. Because the numerical integration method for the SS method using a rectangular-type contour has not been proposed, the Gauss-Legendre quadrature rule is examined in this study. In this subsection, the selection of N_q is examined by monitoring the residual norms of the obtained eigenpairs. Note that I remove eigenpairs whose residuals are larger than 10^0 or located outside Γ as spurious eigenpairs. Here, I consider the fcc Al bulk with 4 atoms whose transport direction is parallel to the $\langle 100 \rangle$ direction. It is used that $N_{mm} = 8$, $N_{rh} = 16$, and $\lambda_{min} = 0.001$. The criterion of the singular value decomposition is set to 10^{-14} and that for the shifted BiCG method is set to 10^{-15}. The SS-Hankel method is employed. Figure 6.13(a) shows the
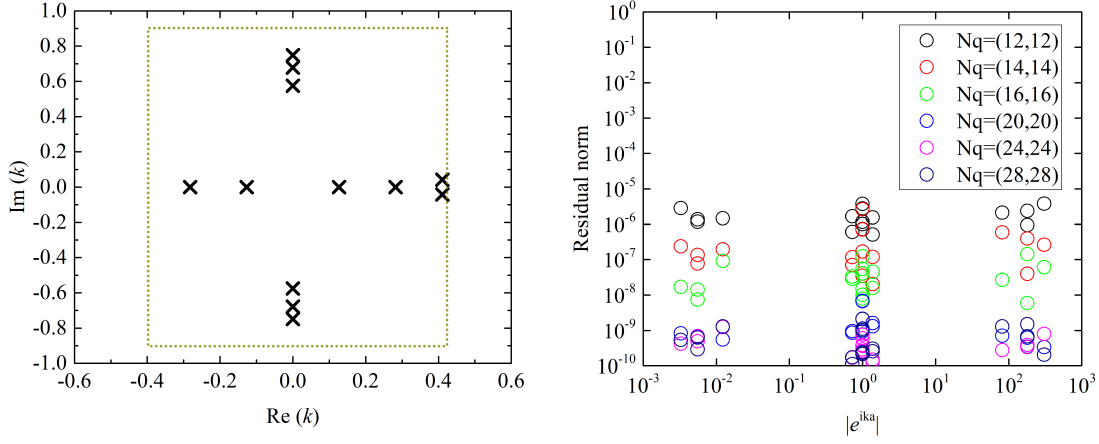
Figure 6.13: Numerical results for fcc Al bulk at Fermi energy of (a) distribution of eigenvalues within the domain enclosed by Γ and (b) residuals $||[\varepsilon\hat{I} - H(k_n)]\phi_n||_2$ when varying the order of the Gauss-Legendre quadrature rule, $N_q = (N_{q1}, N_{q2})$. The number of target eigenvalues that do not include spurious eigenpairs is 18. The plots clearly show that the positions of the eigenpairs are almost unchained and that the accuracy is straightforwardly improved by increasing N_q. Convergence is achieved at $N_q = (24, 24)$: further improvement is not achieved by increasing N_q over $(24, 24)$. Contour pass on k plane is indicated by the broken line.

distribution of the eigenvalues when $N_q = (24, 24)$. It should be noted that the obtained eigenvalues are pairwise, that is, $(k_i, k_j) \approx (k_i, k_i^*)$, and the number of eigenvalues is unchanged irrespective of the selection of N_q. In Fig. 6.13(b), the residuals $||[\varepsilon\hat{I} - H(k_n)]\phi_n||_2$ are plotted as a function of N_q. The accuracy of the obtained eigenpairs is uniformly improved by increasing N_q until $N_q = (24, 24)$, but the noticeable change in accuracy is not observed between $N_q = (24, 24)$ and $N_q = (28, 28)$. It means that numerical integration by the Gauss-Legendre rule is accurate enough when $N_q = (24, 24)$, and other parameters such as N_{rh} should be modified to achieve the further improvement, see Fig. 6.14.

6.8.2　Robustness of algorithm

Next, I demonstrate the robustness of the computations on the k plane in Fig. 6.12(b) against those on λ plane in Fig. 6.12(a). For comparison, I also apply the algorithm proposed in Ref. [122] for solving the QEP for λ by using the contour along the ring-shaped region in Fig. 6.12(a). The input parameters of the SS-Hankel method are set to $N_{mm} = 8$ and $N_{rh} = 16$ for $\lambda_{min} = 0.001$. The criterion of the singular value decomposition is set to 10^{-14} and that for the shifted BiCG method is set to 10^{-15}. The order of the Gauss-Legendre quadrature rule is $N_q = (24, 24)$ in the k plane computation. Instead, I use the trapezoidal rule to approximate the contour integrals in Fig. 6.12(a), with the number of quadrature points being 36 per circle. Figure 6.15 shows the eigenvalues and residuals calculated on
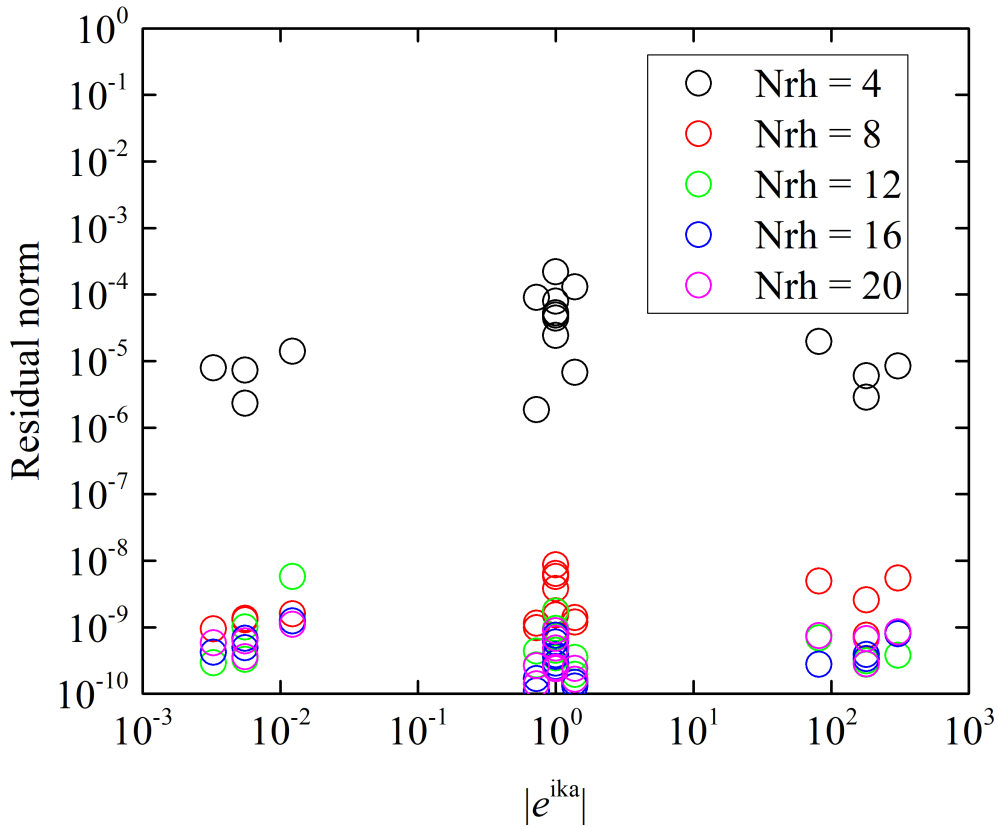
Figure 6.14: Numerical results for fcc Al bulk at Fermi energy of residuals $||[\varepsilon\hat{I} - H(k_n)]\phi_n||_2$ when varying the number of right-hand sides, N_{rh}. $N_{mm} = 8$ and $N_q = (24, 24)$ are used. The plots clearly show that the positions of the eigenpairs are almost unchained and that the accuracy is straightforwardly improved by increasing N_{rh}. Convergence is achieved at $N_{rh} = 16$: further improvement is not achieved by increasing N_{rh} over 16.

the k plane and λ plane for $\lambda_{min} = 0.001$. In both cases, all residuals computed on the k plane are below the convergence criteria (10^{-8}) irrespective of the eigenvalues. On the contrary, the accuracy of the solutions computed on the λ plane is quite poor, and the accuracy is seen to degrade from the outer circle toward the inner circle. The round-off error of the λ-plane computation can be reduced by adding some circles between the outer and the inner circles with the appropriate choice of N_{rh} per additional circle. If the BiCG method is employed as a solver, the number of iteration for the BiCG method increases linearly or more with the common of λ_{min}^{-1}, which leads to reduce the cost of total computational time. Although the appropriate choice of N_{rh} per additional circle is difficult *a priori*, N_{rh} can be tuned by monitoring the accuracy of the obtained eigenpairs. Note that this procedure is not so time-consuming because solving linear system is the dominant part of the computational time. By contrast, if the LU decomposition method is used as a solver, the additional computational cost is roughly proportional to the number of added circles because the computation of LU decompositions at quadrature points takes up a dominant part of the total execution time.

6.9 Transmission calculation

In this section, I present the transmission calculations for Au atomic chain with a CO molecule. I chose this system because transport properties have been investigated extensively using other methods [137, 138]. To validate the accuracy of the method for electron transport calculations, I study the effect of excluding rapidly decaying evanescent waves on the zero-bias transmission calculation, and I compare this result with those obtained using other methods. In all calculations, the Kohn-Sham Hamiltonian matrices in the transition region are obtained from the DFT calculation under periodic boundary conditions, and the scattering states in real-space grids are calculated using the IOBM method [70,71].

I present the transmission calculation of the Au atomic chain with a CO molecule. Prior to this work, transmission calculations for this system have been done by Calzolari *et al.* [137] and Strange *et al* [138]. Interestingly, both groups reported relatively different transmission curves, even though they employed the same methodology combined with the Green's function method and maximally localized Wannier function. As stated in Ref. [138], the disagreement might be related to the manner of construction of the tight-binding Hamiltonian, but it is still unresolved as to which one is correct. Thus, I revalidate the transmission calculation of the Au atomic chain with a CO molecule using the real-space grid method. Figure 6.16(a) shows the atomic structure of the Au atomic chain with a CO molecule. The transition region is a rectangular box of $22.68 \times 22.68 \times 49.32$ bohr^3, and electron transport occurs along the z direction. As in Refs. [137, 138], the bond lengths are set as $d_{\mathrm{Au-Au}} = 5.48$ bohr, $d_{\mathrm{Au-C}} = 3.70$ bohr, and $d_{\mathrm{C-O}} = 2.17$ bohr, and the Au atom attached to CO is shifted toward CO by 0.38 bohr. The left and right electrodes are infinite Au chains with two atoms in the unit cell. A grid spacing of 0.43 Å is used in the real-space grid calculation. Figure 6.16(b) shows the transmission spectra obtained using the self-energy matrices calculated by the proposed method with $\lambda_{min} = 0.999, 0.1, 0.01, 0.001$. In all calculations, I reproduce the main features in Fig. 2 of Ref. [138]: (i) the drop in the transmission at the Fermi energy that originates from resonant
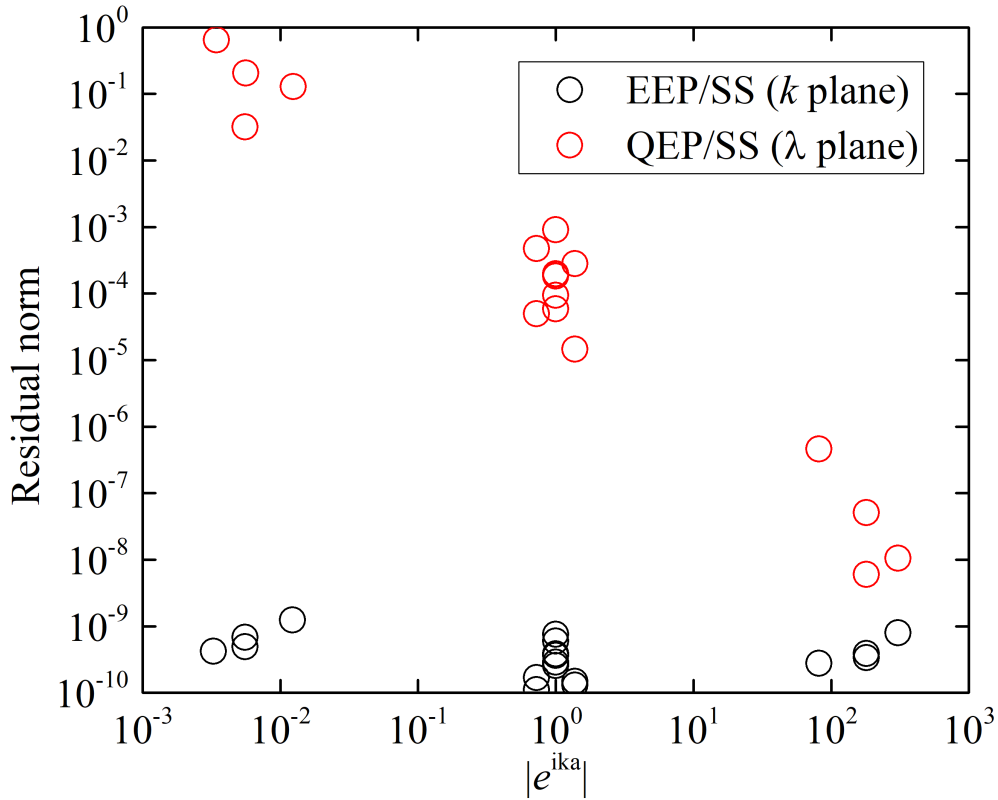
Figure 6.15: Residuals $||[\varepsilon\hat{I} - H(k_n)]\phi_n||_2$ for fcc Al bulk at $\varepsilon = \varepsilon_F + 1.0$ eV calculated on the k plane (EEP/SS) and λ plane (QEP/SS). For $\lambda_{min} = 0.001$, the number of eigenvalues that do not include spurious eigenpairs is 18. In both cases, the EEP/SS method uses $N_q = (24, 24)$ as the order of the Gauss-Legendre quadrature rule; by contrast, the QEP/SS method uses the trapezoidal rule with the number of quadrature points being 36 per circle in Fig. 6.12(a). The other parameters are kept the same.

scattering by CO adsorption, (ii) the single broad peak at $\varepsilon - \varepsilon_F \in [0, 2]$ eV, and (iii) spiky peaks at $\varepsilon - \varepsilon_F \in [-4, 0]$ eV. As the λ_{min} value decreases, the transmission spectra rapidly converge toward the correct values, and visible differences are not observed when $\lambda_{min} \leq 0.01$. In Fig. 6.16(c), the real-space grid calculation shows the qualitatively good agreement with the curve of Ref. [138]; however, I found some discrepancies between the real-space grid calculation and the curve of Ref. [137].
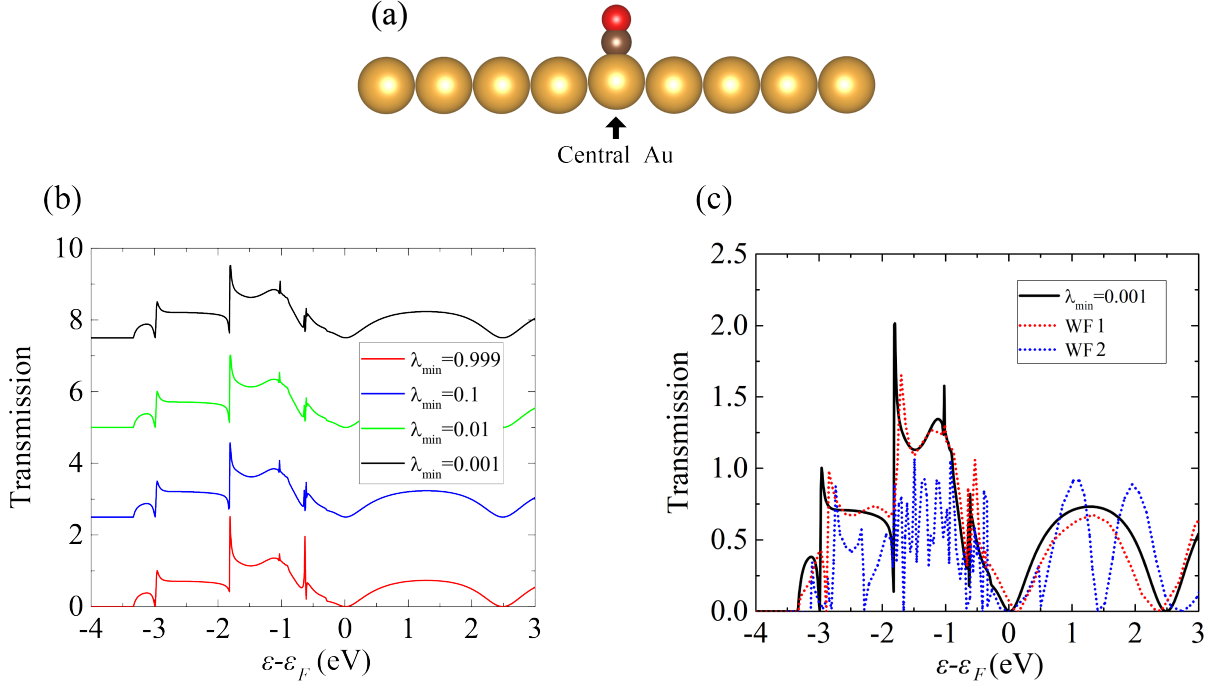
Figure 6.16: (a) Transition region of Au atomic chain with CO adsorption. Au, C, and O atoms are represented as gold, brown, and red balls, respectively. (b) Transmission spectra obtained using self-energy matrices calculated by the proposed method with four different λ_{min} values: 0.999 (red line), 0.1 (blue line), 0.01 (green line), and 0.001 (black line). For clarity, transmission spectra are shifted by the amount of 2.5 with respect to the original values in descending order of the legend. (c) The transmission spectrum obtained with the proposed method (black line). The results obtained with the maximally localized Wannier function of Ref. [138] (WF1: red dashed line) and Ref. [137] (WF2 blue dashed line) are also shown in (c).

6.10   Serial performance

In this section, I experimentally evaluate the serial performance of the SS method for the eigenvalue problem arising from the computation of self-energy matrices. To demonstrate speed-ups, the computational time is compared with that of the OBM method, which is categorized as a WFM method. Although continuous improvements [73, 76, 78, 126] have been made after the first study of the OBM

method, the computation of the first and last N_r columns of $(\varepsilon\hat{I} - \hat{A})^{-1}$ and the $2N_r$-dimensional generalized eigenvalue problem is still required. In this study, the matrix inversion is calculated using the CG method, and the generalized eigenvalue problem is solved by the SS method [73]. It should be noted that another method based on a real-space grid approach proposed by Khomyakov *et al.* [68] is not considered here because its computational procedure and cost are almost the same as those of the OBM method. In addition, popular methods [48–50, 54, 55] used in the NEGF method are also excluded from consideration because they involve the inversion of very dense matrices with the size of the real-space grids in the unit cell of the electrode.

Table 6.7 shows the breakdown of the profiling results in various test systems. All calculations are performed on a two-socket Intel Xeon E5-2667v2 with 16 cores (3.3 GHz) and 256 GB of system memory. 4 MPI processes and 4 OpenMP threads are assigned to the CPU. The parameter λ_{min} is set to 0.1, as in Ref. [48]. The input parameters of the proposed method are set as $N_{mm} = 8$, $N_q = (24, 24)$, and $N_{ene} = 100$, and the criterion of the singular value decomposition and shifted BiCG method is set to 10^{-15}. The number of right-hand sides N_{rh} is set such that all residuals are less than 10^{-8}. Equidistant energy points are chosen in the interval $\varepsilon - \varepsilon_F \in [-1, 1]$ eV, where ε_F is the Fermi energy. The CPU times of the proposed method listed in the sixth column in Table 6.7 represent the average calculation times for 100 energy points. For the SS method used in the OBM method, I employ the trapezoidal rule with the number of quadrature points being 32 per circle in Fig. 6.2, which is the default value used in z-Pares, [136] and other parameters are set as the same in the proposed method. The CPU times including both CG and SS contributions listed in the fourth column are evaluated from the computation at the Fermi energy owing to the limitation of computational resources. The CPU times of these two reference methods can be reduced by employing the shifted CG method instead of the standard CG method and using smaller number of quadrature points. To use the computer resources efficiently, I need to optimize the parameters for the shifted CG and SS methods, which depends on the test systems. Since the usage of uneven parameters become an obstacle to demonstrate the characteristic advantage of the proposed method, the standard CG method is employed and the number of quadrature points is set to be the default value in z-Pares. It was verified that the proposed method is faster than the reference methods in order by one-, two-, and three-dimensional systems. The computational cost of the proposed method scales as $O(MN_{itr}N_{int}N_{rh})$, while that of the CG/SS method does as $O(N_r^3 N_{int})$, where N_{itr} is the number of iterations for the shifted BiCG method and increases linearly or more with the common logarithm of $1/\lambda_{min}$. The number of target eigenvalues generally decreases against its matrix size as the dimension of the systems becomes smaller, indicating that the proposed method is much more efficient in the low dimensional systems because N_{rh} can be set to be small number.

6.11   Application

Silicene, which is a two-dimensional honeycomb structure of Si atoms, is a promising candidate for future nanoelectronic devices due to its unique electronic structures, as represented by a zero-gap

Table 6.7: CPU times in hours for computing the eigenvalue problems arising from the self-energy computations for various electrode materials. Here, M is the size of the Hamiltonian matrix, and $2N_r$ is the number of nontrivial solutions of Eq. (3.65). N_{rh} is the number of right-hand sides used in the SS method. The CPU times of the proposed method (this work) are averaged by the computation times at 100 different energy points between $\varepsilon_F - 1$ eV and $\varepsilon_F + 1$ eV, where ε_F is the Fermi energy. On the other hand, the CPU times of the OBM method (CG/SS) are measured only at the Fermi energy owing to the limitation of computational resources.

Material	M	$2N_r$	N_{rh}	CG/SS	This work
Au chain [a]	64896	21632	8	4.11	0.01
Al(100) wire [b]	153600	51200	16	28.96	0.08
(6,6)CNT [c]	62208	41472	32	13.15	0.12
Graphene [d]	14336	7168	8	0.13	0.00
Silicene [e]	110592	24576	16	5.66	0.09
Au(111) bulk [f]	34560	7680	8	0.66	0.01

[a] Geometry description is found in the transmission calculation of the Au atomic chain with a CO molecule. See Sec. 6.9.

[b] Geometry description is found in Ref. [42].

[c] Ideal armchair (n,n) carbon nanotube with C-C bond length of 2.68 Å.

[d] Graphene with four atoms whose transport direction is along the armchair direction and C-C bond length is 2.68 Å.

[e] Geometry description is found in the transmission calculation of the free-standing silicene. See Sec. 6.11.

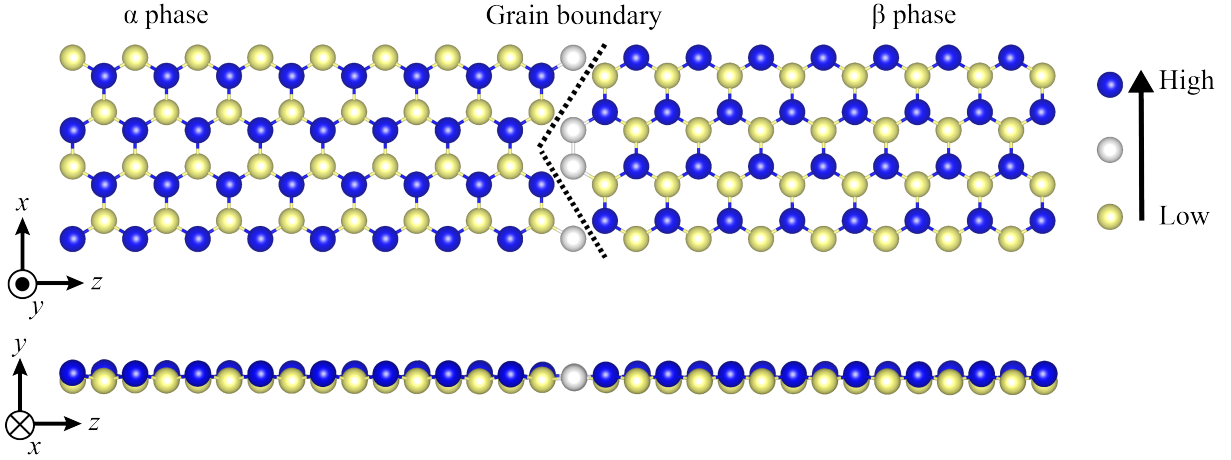[f] Geometry description is found in Ref. [138].

Figure 6.17: Optimized interface structure of silicene with α-β interface.

semiconductor with Dirac cone [139, 140]. In fact, a silicene field effect transistor (FET) using the transfer-fabrication process was recently reported. [141] However, the measured mobility values of silicene FET are considerably lower than the theoretical calculation [142] by an order of magnitude, and grain boundary scattering has been proposed as a possible cause. Despite the demand for the detailed information on the electron scattering at the grain boundaries of silicene, the electron transport behavior across the grain boundary of silicene is not well understood. This is because the fabrication of silicene is still challenging, especially on a dielectric substrate.

Unlike graphene, it is well known that silicene forms a low-buckled structure, which leads to two energetically equivalent geometrical phases, whose buckling directions are opposite to each other, as shown in Fig. 6.17. Following the notations in Ref. [143], I call these phases the α and β phases, respectively. In this study, I present the first-principles analysis of the transport properties of the free-standing silicene sheet across the interface between the α and β phases. To keep the focus on application of the methodology to the transport calculations rather than on comprehensive understanding of the scattering mechanism in silicene, attention is paid only to the grain boundary between the α and β phases along the armchair direction.

Initially, I perform the relaxation of the interface structure using a grid spacing of 0.21 Å and a $4 \times 1 \times 1$ k-point sampling on Brillouin zone. The interface model is constructed with a 256-atom supercell using a value of 2.27 Å for the Si-Si bonding length. To avoid the spurious interaction between silicene layers, a vacuum region of 10 Å is introduced in the simulation cell. The interface structure is relaxed until the residual forces become lower than 0.003 eV/Å. The relaxed geometrical structure is shown in Fig. 6.17. The reconstruction of chemical bonds at the interface does not occur, but instead the rearrangement of the out-of-plane dislocation is observed at the interface. This result is in agreement with previous theoretical work. [143] I subsequently perform the transport calculations along the z direction using the self-energy matrices evaluated with $\lambda_{min} = 0.01$. The transition region

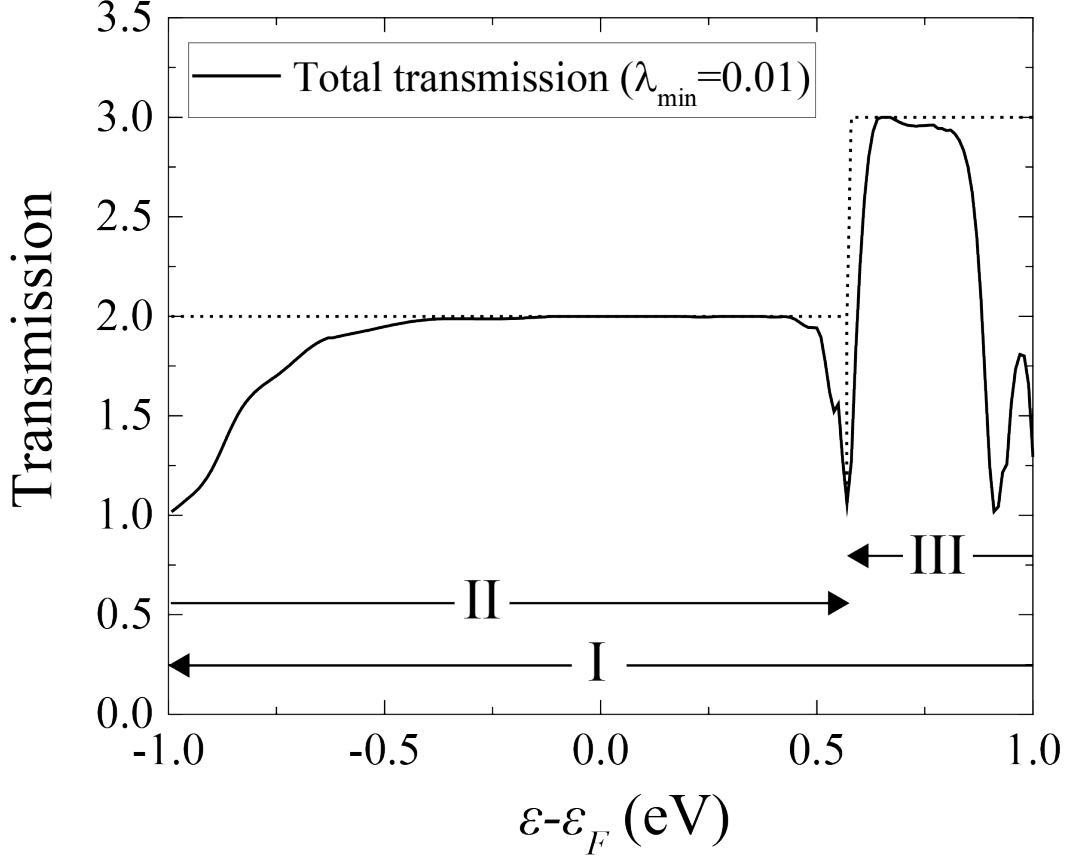contains 128 atoms and the Γ point in the transverse direction is used in the transport calculation.

Figure 6.18: Effects of the α-β interface on the transmission spectra. The solid line is the result of real-space grid calculation using the self-energy matrices obtained with $\lambda_{min} = 0.01$. Empty dots denote the transmission spectrum without the defect.

The total transmission is shown in Fig. 6.18. A feature of immediate interest is that the transmission at the Fermi energy is unchanged for a pristine silicene; that is, this type of grain boundary does not scatter incoming electrons at this energy. On the other hand, I found three transmission dips below and above the Fermi energy in Fig. 6.18. Aiming to understand the origin of such dips, I also plot the group velocity of the incident electrons and the band structure of silicene in Fig. 6.19. The important bands near the Fermi energy are labeled by I, II and III, and they contribute to the transmission in $[-1.0, 1.0]$ eV, $[-1.0, 0.57]$ eV, and $[0.57, 1.0]$ eV, respectively. Note that the III band is doubly degenerated. At $\varepsilon = \varepsilon_F - 1.0$ and 0.57 eV where the transmission dips are observed, the group velocities go to zero with opening or closing the channel in bands. The scattering where the group velocity becomes zero is understood by the one-dimensional tight-binding model with a single impurity. According to Eq. (75) in Ref. [69], it is easy to show that the transmission probability becomes zero when the group velocity becomes zero, which indicates that the perturbation of the po-

tential induced by the geometrical disorder at the silicene interface causes the scattering near the band edge. In addition, I numerically examine the scattering at the band edge using the one-dimensional Kronig-Penny model and observe the strong scatterings at the band edge (see, Appendix D.)
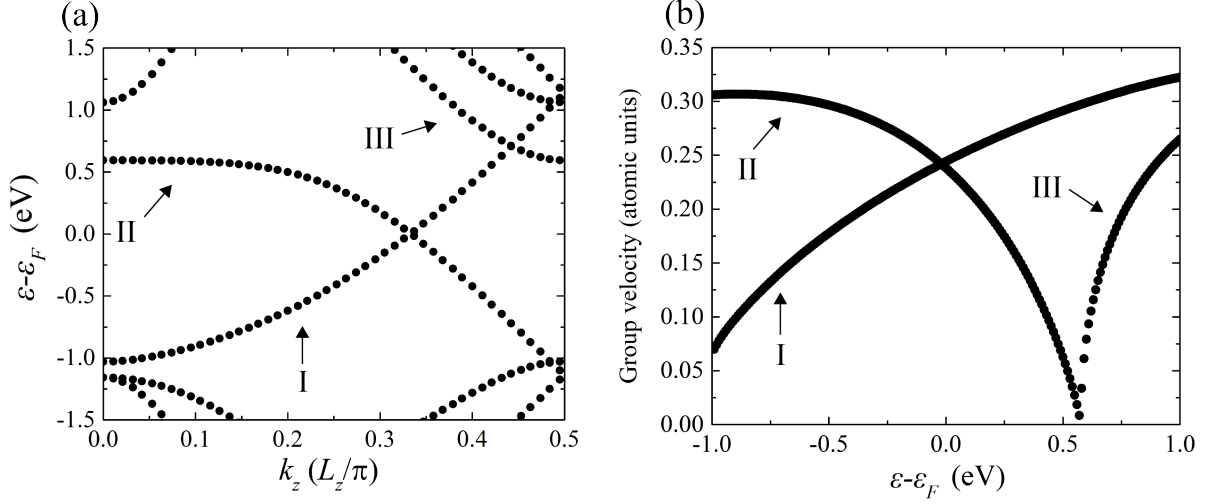

Figure 6.19: (a) Band structure and (b) group velocity of silicene.

I next consider the origin of the dip at $\varepsilon = \varepsilon_F + 0.91$ eV, where two bulk modes in III band are completely reflected at the interface. To obtain the more detailed information about the scattering, I plot the charge densities of two bulk modes for left electrode at $\varepsilon = \varepsilon_F + 0.6$ and 0.91 eV. As seen from Fig. 6.20, the charge densities of two channels at $\varepsilon = \varepsilon_F + 0.6$ eV distribute on inner and outer sides of silicene atoms, while they turn to concentrate on the only outer side at $\varepsilon = \varepsilon_F + 0.91$ eV. By expanding the result for left electrode to the interface, the new insight of the scattering is obtained. Figure 6.21 illustrates the scattering of the incident electron coming from the left electrode at $\varepsilon = \varepsilon_F + 0.91$ eV. The bulk modes distribute around the outer side of the silicene atoms in both right and left electrodes, however, the scattering states in the right electrodes will be inner side of the silicene atoms because the buckling of the silicene is reversed. Therefore, the scattering states which come from the left electrode hardly connect with the bulk modes in right electrodes, which leads the transmission reduction at $\varepsilon = \varepsilon_F + 0.91$ eV.
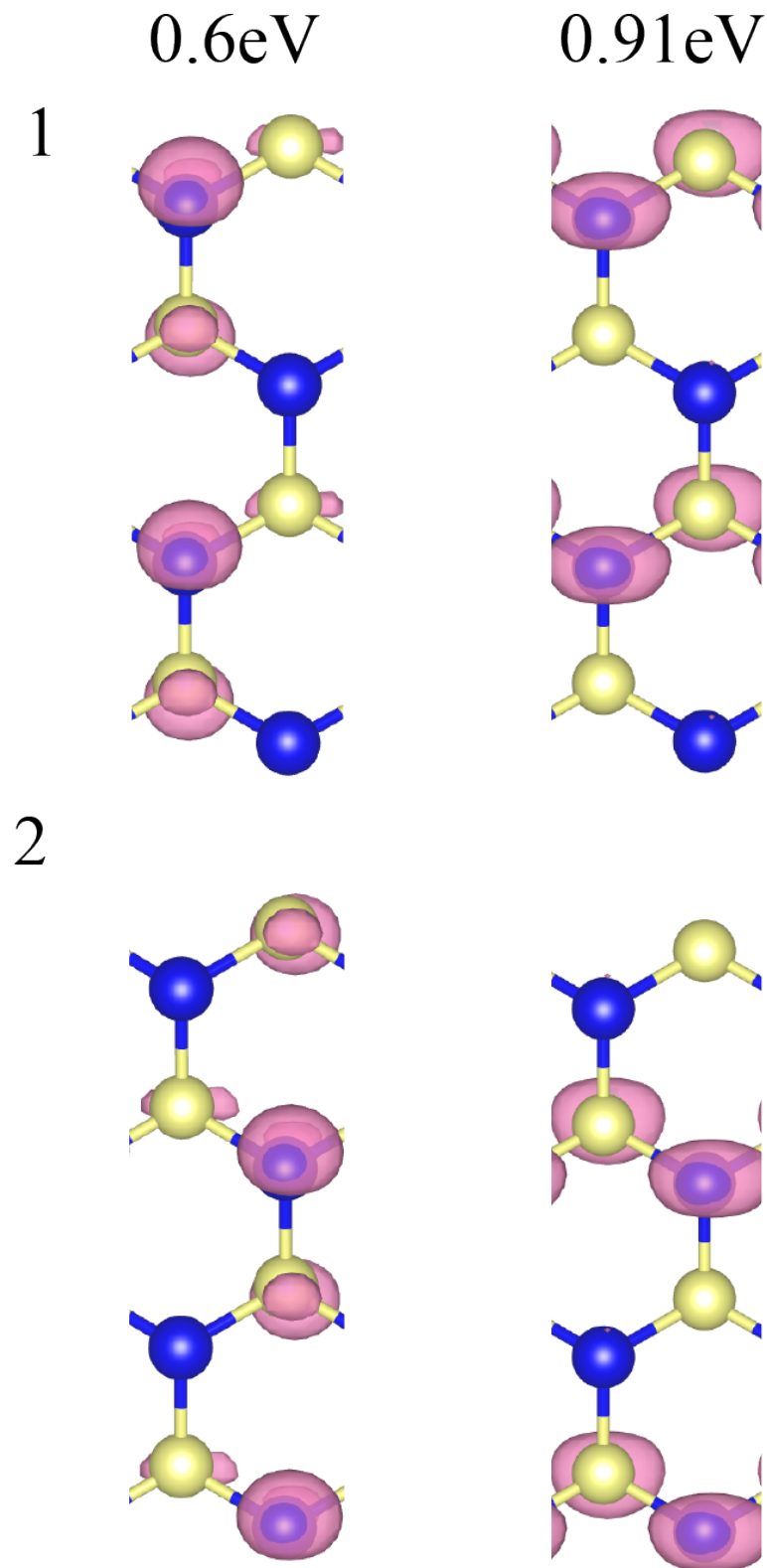
Figure 6.20: Charge densities of two bulk modes in III band for left electrode at $\varepsilon = \varepsilon_F + 0.6$ and 0.91 eV.
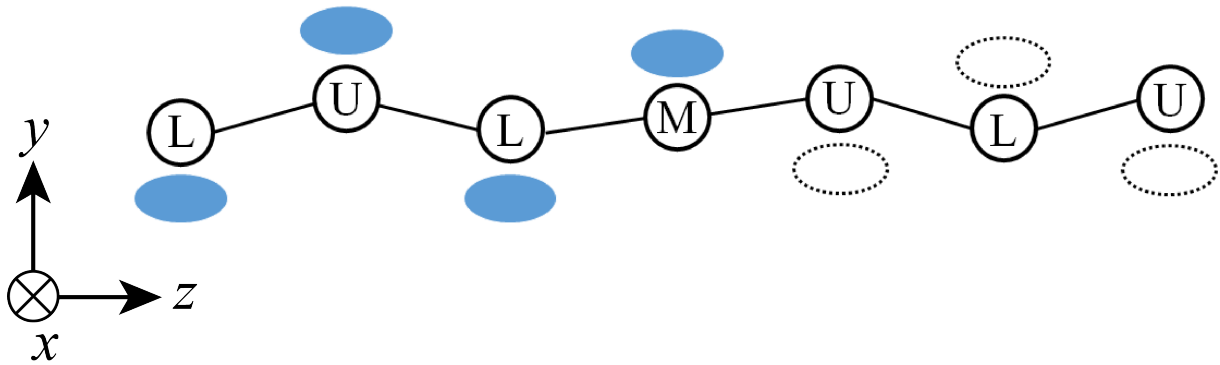
Figure 6.21: An illustration of the scattering of silicene at $\varepsilon = \varepsilon_F + 0.91$ eV. The symbols L, M, and U represent the lower buckled, non-bucked, and upper buckled atoms, respectively.

Chapter 7

Conclusion and outlook

Quantum transport calculations based on DFT have been recognized as powerful tools to investigate the transport properties of the nanoscale conductor. The most popular method for such calculations is the NEGF method which is used frequently with the localized atomic basis. The WFM method which is the alternative of the NEGF method has recently been refocused due to its efficiency. In order to apply these methods to the large-scale system with millions of atoms such as semiconductor device simulations, however, default implementations like a full eigensolver for self-energy matrices of electrodes or Gaussian elimination method for Green's function are inefficient and therefore it is important to develop new theoretical and computational methods for the most time-consuming part of the quantum transport calculations.

The purpose of this thesis is to develop and implement the numerical methods which are faster than existing methods and show its capability for quantum transport calculations. Because the implementation of the NEGF or WFM method depends strongly on the basis set, it is important to select the appropriate one according to their purpose. Here, I focused on the real-space grid method as a basis set for transport calculations. The first reason is that the real-space grid method enables us to treat the spatially delocalized electronic states which are important because these states play a central role in the transport phenomena such as tunneling current in the scanning-tunneling measurement, electron scattering process of the SiC-MOSFET, and so on. The second reason is that the Hamiltonian matrix represented by a RSFD scheme is large but very sparse, which makes it possible to perform the massively parallel computing if one can utilize the sparsity of the Hamiltonian matrix. However, real-space transport calculations are computationally demanding and only several tens atoms can be treated so far. To reduce the computational cost and memory usage of the real-space transport calculation, I thus reformulated the most time-consuming part of the NEGF method, namely, computations of the self-energy matrix and Green's function. At the same time, new algorithms have been proposed to accelerate the time-to-solution compared with the standard solver. I briefly summarize my contributions and present an outlook on future works related with them.

- **Contracted form of the Hamiltonian matrix (Chapter 4)**: Combining the partitioning

and singular value decomposition techniques, I have derived the contracted form of the Hamiltonian matrix that allows us to use the sophisticated numerical methods such as the RGF, quick decimation, and semi-analytical methods. Although the contracted formula has already developed in the literature of the semi-analytical or WFM method, the developed formula in this thesis is applicable not only semi-analytical method but also for the RGF and quick decimation methods, which makes it possible to compute the self-energy matrices at complex energy efficiently. Because the size of the contracted Hamiltonian matrix is much smaller than that of the Hamiltonian matrix of the unit cell, the contracted formula is expected to enable us to achieve a significant speed-up and memory reduction compared to the original methods without large modification. In addition, useful relations to compute the generalized Bloch states which are required for the semi-analytical method based on the contour integral eigensolver have been proposed. I believe that the contracted formula is crucial contribution toward the self-consistent calculation under the non-equilibrium condition.

- **Fast evaluation of the retarded and lesser Green's function (Chapter 5)**: The procedure for computing the retarded and lesser Green's function was developed using the sparsity of the coupling matrix of electrodes and Dyson equation. In this procedure, the unperturbed Green's function is calculated first and the retarded Green's function are obtained by the mathematically strict relationship between the retarded and unperturbed Green's function. Because physically important quantities such as transmission, density of state, and charge density can be evaluated by the only block matrix elements of the unperturbed Green's function, it is suitable to compute them by solving a set of liner equations with the use of the iterative solver. In this direction, the shifted COCG method is employed to obtain the reduced solution vector of the shifted linear system. This algorithm is much faster than the conventional COCG method especially when the many energy points are required to evaluate the Green's function or the system involves many atoms and long along the transport direction, that are usually satisfied in the large-scale transport calculations. Preliminary testing shows that the transmission calculation is accelerated by an order of magnitude in the computation of the unperturbed Green's function. I believe that the this scheme has great potential for studying the quantum transport properties of the large-scale system.

- **Contour integral method (Chapter 6) for self-energy matrix**: Based on the idea that the transmission of nanoscale system can be evaluated with practically sufficient accuracy using a comparatively small number of propagating and moderately decaying waves, a contour integral eigensolver based on the Sakurai-Sugiura method combined with the shifted BiCG method is developed to obtain the such important waves as solutions of the quadratic eigenvalue problems. Because the quadratic eigenvalue problem is solved directly by the iterative solver, one can utilize the fast technique of the conventional first-principles calculations, which speeds up the computations and reduces memory usage dramatically. Furthermore, owing to the inherently rich parallelism of the Sakurai-Sugiura method and the use of a domain-decomposition technique

for real-space grids, excellent scalability on modern massively parallel computers is achieved. Further improvements of the iterative solvers which might avoid the instability of the BiCG method, such as preconditioning, has the potential to outperform the proposed method. I leave this to future work.

Appendix A

Group velocity

In this appendix we give a derivation of the expression of the group velocity,

$$v_n = \frac{1}{\hbar} \frac{\partial \varepsilon}{\partial k_n^+} = \frac{2a}{\hbar} \text{Im}[(\lambda_n^+)^{-1} (\phi_n^+)^\dagger B^\dagger \phi_n^+]. \tag{A.1}$$

For ease of notation, we drop the index $+$ in the following. By introducing the Bloch factor $\lambda_n = e^{ik_n a}$, we can rewrite Eq. (3.66) to the quadratic eigenvalue problem,

$$[-\lambda_n^{-1} \hat{B}^\dagger + (\varepsilon \hat{I} - A) - \lambda_n \hat{B}] \phi_n = 0. \tag{A.2}$$

Multiplying $\phi_m^\dagger$ from the left hand side, we obtain

$$-\lambda_n^{-1} \phi_m^\dagger \hat{B}^\dagger \phi_n + \phi_m^\dagger (\varepsilon \hat{I} - A) \phi_n - \lambda_n \phi_m^\dagger \hat{B} \phi_n = 0. \tag{A.3}$$

From the conjugate of Eq. (A.2),

$$-\lambda_m^* \phi_m^\dagger \hat{B}^\dagger \phi_n + \phi_m^\dagger (\varepsilon \hat{I} - A) \phi_n - (\lambda_m^*)^{-1} \phi_m^\dagger \hat{B} \phi_n = 0. \tag{A.4}$$

Subtracting Eq. (A.4) from Eq. (A.3) and multiplying $\lambda_m^* \lambda_n$ gives

$$(\lambda_n \phi_m^\dagger \hat{B} \phi_n - \lambda_m^* \phi_m^\dagger \hat{B}^\dagger \phi_n)(1 - \lambda_n \lambda_m^*) = 0 \tag{A.5}$$

Here we introduce the velocity matrix $\hat{\mathcal{V}}$ defined by

$$\hat{\mathcal{V}}_{m,n} = i(\lambda_n \phi_m^\dagger \hat{B} \phi_n - \lambda_m^* \phi_m^\dagger \hat{B}^\dagger \phi_n) \tag{A.6}$$

According to Eq. (A.5), if $\lambda_n \lambda_m^* \neq 1$ then $\hat{\mathcal{V}}_{m,n} = 0$. Because right going evanescent states have $\text{Im}(k_n) > 0$, i.e., $\lambda_n > 0$, $\hat{\mathcal{V}}_{m,n} = 0$ for evanescent waves and $\hat{\mathcal{V}}_{m,n}$ might has nonzero value only when both ϕ_m and ϕ_n are right-going propagating states.

Next, taking the derivative with respective to ε of Eq. (A.3), we obtain

$$\frac{\partial \lambda_n}{\partial \varepsilon} (\lambda_n^{-1} \phi_m^\dagger \hat{B}^\dagger \phi_n - \lambda_n \phi_m^\dagger \hat{B} \phi_n) + \lambda_n \phi_m^\dagger \phi_n = i \frac{\partial \lambda_n}{\partial \varepsilon} \hat{\mathcal{V}}_{m,n} + \lambda_n \delta_{m,n} = 0. \tag{A.7}$$

Here we used $\lambda^* = \lambda^{-1}$ and $\phi_m^\dagger \phi_n = \delta_{m,n}$ for propagating waves. The above equation indicates that $\hat{\mathcal{V}}_{m,n} = 0$ for $m \neq n$. In addition, differentiating the both side of $\lambda_n = e^{ik_n a}$ yields

$$\frac{\partial \lambda_n}{\partial \varepsilon} = ia\lambda_n \frac{\partial k_n}{\partial \varepsilon}. \tag{A.8}$$

Finally substituting Eq. (A.8) into Eq. (A.3), we obtain

$$
\begin{aligned}
v_n &= \frac{a}{\hbar}\hat{\mathcal{V}}_{n,n} & \text{(A.9)}\\
&= \frac{ia}{\hbar}(\lambda_n \phi_n^\dagger \hat{B} \phi_n - \lambda_n^* \phi_n^\dagger \hat{B}^\dagger \phi_n) & \text{(A.10)}\\
&= \frac{2a}{\hbar}\mathrm{Im}(\lambda_n^{-1} \phi_n^\dagger \hat{B}^\dagger \phi_n). & \text{(A.11)}
\end{aligned}
$$

This expression is seen to be equivalent to Eq. (A.1).

Appendix B

Dual singular value decomposition for self-energies

In Sec. 4.3, I assume that $\hat{\mathcal{C}} = \hat{\mathcal{B}}^\dagger$, where

$$\hat{\mathcal{C}} = \hat{\beta}_2^\dagger (z\hat{I} - \hat{\alpha}_2)^{-1} \hat{\beta}_1^\dagger, \tag{B.1}$$

in order to write the unitary transformed matrix as

$$\hat{\mathcal{C}}' = \begin{bmatrix} \hat{\mathcal{C}}'_{11} & 0 \\ \hat{\mathcal{C}}'_{12} & 0 \end{bmatrix}. \tag{B.2}$$

The matrix structures of Eqs. (B.1) and (4.31) are essential to derive Eqs. (4.33) and (4.39). However, it is clear that $\hat{\mathcal{C}} \neq \hat{\mathcal{B}}^\dagger$ when z is the complex number, and therefore $\hat{\mathcal{C}}'$ does not have such a matrix structure. Here, we present a simple transformation scheme to overcome this difficulty. If $\hat{A}_i$ and $\hat{B}_i$ are complex matrices and z is a complex number, one might consider a dual transformation based on the singular value decomposition for $\hat{\mathcal{B}}$ and $\hat{\mathcal{C}}$,

$$\hat{\mathcal{B}} = \hat{U}_B \hat{S}_B \hat{V}_B^\dagger, \tag{B.3}$$

$$\hat{\mathcal{C}} = \hat{U}_C \hat{S}_C \hat{V}_C^\dagger \tag{B.4}$$

where $\hat{U}_B, \hat{V}_B, \hat{U}_C$, and $\hat{V}_C$ are unitary matrices and $\hat{S}_B$ and $\hat{S}_C$ are diagonal matrices with the singular values on the diagonal. From the construction, the rank of $\hat{\mathcal{C}}$ should be same as the rank of $\hat{\mathcal{B}}$. Using $\hat{U}_B$ and $\hat{V}_C$, a dual transformation

$$\hat{\mathcal{B}}' = \hat{U}_B^\dagger \hat{\mathcal{B}} \hat{V}_C, \tag{B.5}$$

$$\hat{\mathcal{C}}' = \hat{U}_B^\dagger \hat{\mathcal{C}} \hat{V}_C, \tag{B.6}$$

bring the same matrix form of Eqs. (B.1) and (4.31). Thus, all derivations in Sec. 4.3 are valid when $\hat{\mathcal{C}} \neq \hat{\mathcal{B}}^\dagger$.

If $\hat{A}_i$ and $\hat{B}_i$ are real matrices, which leads $\hat{\mathcal{C}} = \hat{\mathcal{B}}^T$, a transformation,

$$\hat{\mathcal{B}}' = \hat{U}_B^\dagger \hat{\mathcal{B}} \hat{U}_B^\dagger, \tag{B.7}$$

$$\hat{\mathcal{C}}' = \hat{U}_B^\dagger \hat{\mathcal{C}} \hat{U}_B^\dagger, \tag{B.8}$$

brings both $\hat{\mathcal{B}}'$ and $\hat{\mathcal{C}}'$ to the matrix form of Eqs. (B.1) and (4.31), respectively. In this case, the singular value decomposition for $\hat{\mathcal{C}}$ is not required.

Appendix C

Shifted BiCG method and seed switching technique

The BiCG method is one of the Krylov subspace methods for solving dual linear systems such as

$$\hat{A}\boldsymbol{x} = \boldsymbol{b}, \quad \hat{A}^\dagger \tilde{\boldsymbol{x}} = \boldsymbol{b}, \tag{C.1}$$

where the matrix $\hat{A}$ need not to be a Hermitian matrix. The algorithm updates the solution vectors x and $\tilde{\boldsymbol{x}}$ using the vectors $\boldsymbol{p}$, $\boldsymbol{r}$, $\tilde{\boldsymbol{p}}$, and $\tilde{\boldsymbol{r}}$ and the scalars α and β via the following recurrences:

$$\alpha_n = \frac{(\tilde{\boldsymbol{r}}_n, \boldsymbol{r}_n)}{(\tilde{\boldsymbol{p}}_n, \hat{A}\boldsymbol{p}_n)}, \tag{C.2}$$

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \alpha_n \boldsymbol{p}_n, \tag{C.3}$$

$$\tilde{\boldsymbol{x}}_{n+1} = \tilde{\boldsymbol{x}}_n + \bar{\alpha}_n \tilde{\boldsymbol{p}}_n, \tag{C.4}$$

$$\boldsymbol{r}_{n+1} = \boldsymbol{r}_n - \alpha_n \hat{A}\boldsymbol{p}_n, \tag{C.5}$$

$$\tilde{\boldsymbol{r}}_{n+1} = \tilde{\boldsymbol{r}}_n - \bar{\alpha}_n A^\dagger \tilde{\boldsymbol{p}}_n, \tag{C.6}$$

$$\beta_n = \frac{(\tilde{\boldsymbol{r}}_{n+1}, \boldsymbol{r}_{n+1})}{(\tilde{\boldsymbol{r}}_n, \boldsymbol{r}_n)}, \tag{C.7}$$

$$\boldsymbol{p}_{n+1} = \boldsymbol{r}_{n+1} + \beta_n \boldsymbol{p}_n, \tag{C.8}$$

$$\tilde{\boldsymbol{p}}_{n+1} = \tilde{\boldsymbol{r}}_{n+1} + \bar{\beta}_n \tilde{\boldsymbol{p}}_n, \tag{C.9}$$

where the initial conditions are set as $\boldsymbol{x}_0 = \tilde{\boldsymbol{x}}_0 = 0$ and $\boldsymbol{r}_0 = \boldsymbol{p}_0 = \tilde{\boldsymbol{r}}_0 = \tilde{\boldsymbol{p}}_0 = \boldsymbol{b}$.

Now, we focus on solving the m sets of shifted dual linear systems:

$$(\hat{A} + \sigma_i \hat{I})\boldsymbol{x}(\sigma_i) = \boldsymbol{b}, \quad (\hat{A}^\dagger + \sigma_i \hat{I})\tilde{\boldsymbol{x}}(\sigma_i) = \boldsymbol{b}, \tag{C.10}$$

for $i = 1, 2, \ldots, m$, using the reference system $\hat{A}\boldsymbol{x} = \boldsymbol{b}$ and $\hat{A}^\dagger \tilde{\boldsymbol{x}} = \boldsymbol{b}$, where σ_i is a real-valued scalar shift and $\hat{I}$ is the identity matrix. When we choose the initial conditions as $\boldsymbol{x}_0(\sigma_i) = \tilde{\boldsymbol{x}}_0(\sigma_i) = 0$, the Krylov subspace of the reference system and shifted dual linear systems are identical. As a result, the residual vectors $\boldsymbol{r}_n(\sigma_i)$ and $\tilde{\boldsymbol{r}}_n(\sigma_i)$ are collinear with $\boldsymbol{r}_n$ and $\tilde{\boldsymbol{r}}_n$, respectively, that is,

$$\boldsymbol{r}_n(\sigma_i) = \frac{1}{\pi_n(\sigma_i)}\boldsymbol{r}_n, \quad \tilde{\boldsymbol{r}}_n(\sigma_i) = \frac{1}{\bar{\pi}_n(\sigma_i)}\tilde{\boldsymbol{r}}_n, \tag{C.11}$$

where $\pi_n(\sigma_i)$ is a scalar that is updated by the following recurrence:

$$\pi_{n+1}(\sigma_i) = (1 + \frac{\beta_{n-1}\alpha_n}{\alpha_{n-1}} + \alpha_n\sigma_i)\pi_n(\sigma_i) - \frac{\beta_{n-1}\alpha_n}{\alpha_{n-1}}\pi_{n-1}(\sigma_i). \tag{C.12}$$

Here, $\pi_0(\sigma_i) = \pi_{-1}(\sigma_i) = 1$. By using the collinear relation given in Eq. (C.11), the shifted dual linear systems are updated by the following recurrences:

$$\alpha_n(\sigma_i) = \frac{\pi_{n-1}(\sigma_i)}{\pi_n(\sigma_i)}\alpha_n, \tag{C.13}$$

$$\beta_n(\sigma_i) = \left(\frac{\pi_{n-1}(\sigma_i)}{\pi_n(\sigma_i)}\right)^2\beta_n, \tag{C.14}$$

$$\boldsymbol{x}_{n+1}(\sigma_i) = \boldsymbol{x}_n(\sigma_i) + \alpha_n(\sigma_i)\boldsymbol{p}_n(\sigma_i), \tag{C.15}$$

$$\tilde{\boldsymbol{x}}_{n+1}(\sigma_i) = \tilde{\boldsymbol{x}}_n(\sigma_i) + \bar{\alpha}_n(\sigma_i)\tilde{\boldsymbol{p}}_n(\sigma_i), \tag{C.16}$$

$$\boldsymbol{p}_{n+1}(\sigma_i) = \boldsymbol{r}_{n+1}(\sigma_i) + \beta_n(\sigma_i)\boldsymbol{p}_n(\sigma_i), \tag{C.17}$$

$$\tilde{\boldsymbol{p}}_{n+1}(\sigma_i) = \tilde{\boldsymbol{r}}_{n+1}(\sigma_i) + \bar{\beta}_n(\sigma_i)\tilde{\boldsymbol{p}}_n(\sigma_i). \tag{C.18}$$

Because the recurrences in Eqs. (C.11)-(C.18) consist of only scalar-scalar and scalar-vector products, the shifted dual linear systems can be solved very quickly rather than applying the standard BiCG method to them.

The iterations continue until the residual norms of the entire system become sufficiently small. However, when the residual norm of the reference system becomes too small, the numerical precision of the residual vectors of shifted dual linear systems decreases. To avoid this problem, we use the seed switching technique that replaces the reference system with a shifted dual linear system whose residual norm is the largest in the entire system. To switch the reference system to the new one $\tilde{s} = \arg\max_{i \in I}\{||r_n(\sigma_i)||\}$, we need a scalar in Eq. (C.12) for the new reference system:

$$\pi_n(\sigma_{\tilde{s}}, \sigma_i) = \frac{\pi_n(\sigma_i)}{\pi_n(\sigma_{\tilde{s}})}. \tag{C.19}$$

The seed switching technique for an arbitrary shift $\sigma(\notin \{\sigma_1, \sigma_2, \ldots, \sigma_m\})$ is presented in Ref. [83].

Appendix D

Kronig-Penny model

I here discuss the effect of the small perturbation of the potential to the electron scattering in a one-dimensional system with square potentials. The system is divided into L, R, and C and I consider that L and R are semi-infinite electrodes with periodic square potentials and the barrier height of square potential in C is shifted. The parameters are given in Fig. D.1. Figs. D.2(a) and (b) show the energy dispersion for the periodic square potentials and transmission spectra, respectively. Naturally, it is observable that the scatterings at the band edges where the group velocity becomes zero. Note that the same tendency can be seen when varying parameters.
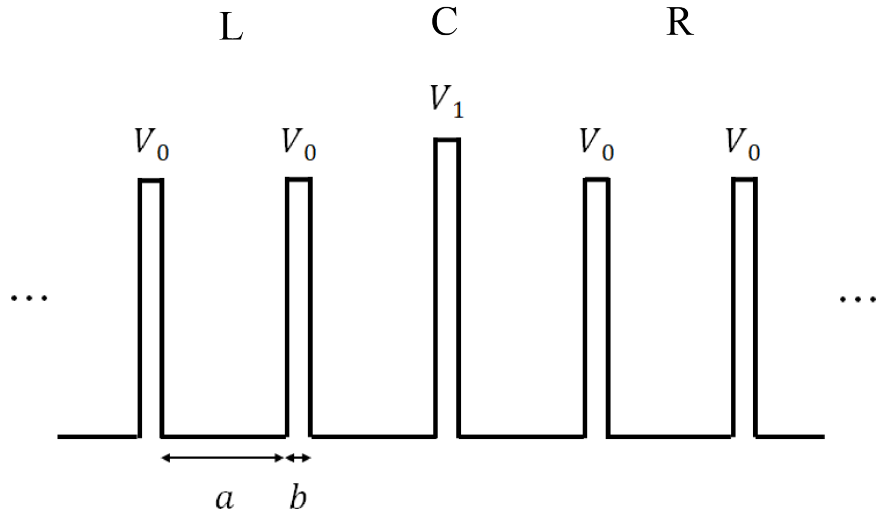
Figure D.1: An illustration of a one-dimensional system with square potential barriers. The parameters a, b, V_0, and V_1 represent the width of depths, width of barriers, barrier height in L and R regions, and barrier height in C region, respectively.
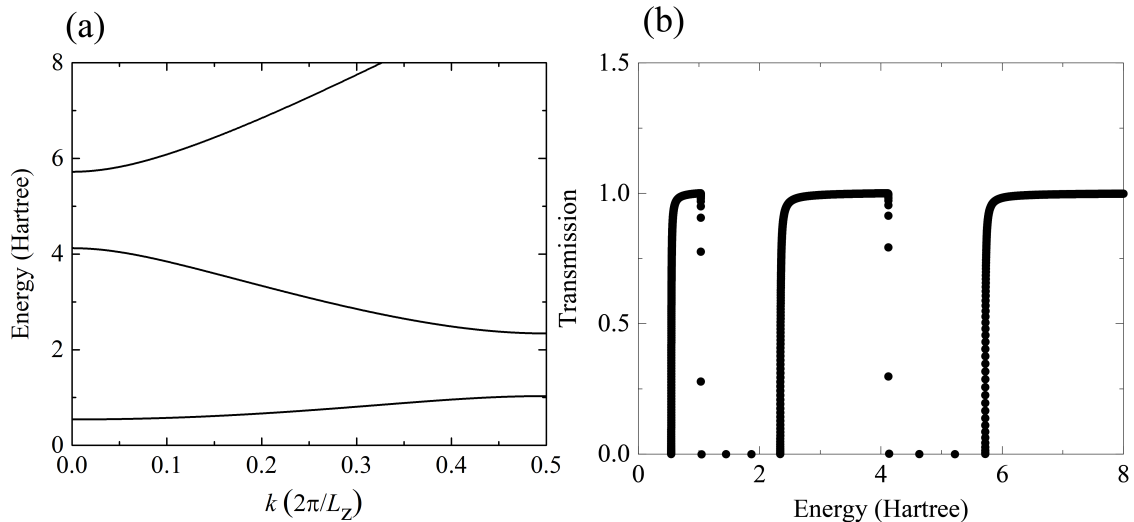

Figure D.2: (a) Energy dispersion of the Kronig-Penny model with periodic square potentials and (b) transmission spectra. The parameters in atomic units are set as $a = 2.0, b = 0.2, V_0 = 10$, and $V_1 = 11$.

Bibliography

[1] P. Hohenberg and W. Kohn. "Inhomogeneous Electron Gas". *Phys. Rev.*, **136**, B864, (1964).

[2] W. Kohn and L. J. Sham. "Self-Consistent Equations Including Exchange and Correlation Effects". *Phys. Rev.*, **140**, A1133, (1965).

[3] D. R. Hamann, M. Schlüter, and C. Chiang. "Norm-Conserving Pseudopotentials". *Phys. Rev. Lett.*, **43**, 1494, (1979).

[4] C. Lee, W. Yang, and R. G. Parr. "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density". *Phys. Rev. B*, **37**, 785, (1988).

[5] A. D. Becke. "Density-functional thermochemistry. III. The role of exact exchange". *J. Phys. Chem.*, **98**, 5648, (1993).

[6] J. P. Perdew, K. Burke, and M. Ernzerhof. "Generalized Gradient Approximation Made Simple". *Phys. Rev. Lett.*, **77**, 3865, (1996).

[7] R. Car and M. Parrinello. "Unified Approach for Molecular Dynamics and Density-Functional Theory". *Phys. Rev. Lett.*, **55**, 2471, (1985).

[8] S. Datta. *"Electronic Transport in Mesoscopic Systems"*. Cambridge Studies in Semiconductor Physics and Microelectronic Engineering. Cambridge University Press, (1995).

[9] E. Runge and E. K. U. Gross. "Density-Functional Theory for Time-Dependent Systems". *Phys. Rev. Lett.*, **52**, 997, (1984).

[10] M. Levy. "Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem". *Proc. Natl. Acad. Sci. USA*, **76**, 6062, (1979).

[11] M. Levy. "Electron densities in search of Hamiltonians". *Phys. Rev. A*, **26**, 1200, (1982).

[12] J. P. Perdew and A. Zunger. "Self-interaction correction to density-functional approximations for many-electron systems". *Phys. Rev. B*, **23**, 5048, (1981).

[13] S. H. Vosko, L. Wilk, and M. Nusair. "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis". *Can. J. Phys.*, **58**, 1200, (1980).

[14] J. P. Perdew and Y. Wang. "Accurate and simple analytic representation of the electron-gas correlation energy". *Phys. Rev. B*, **45**, 13244, (1992).

[15] D. M. Ceperley and B. J. Alder. "Ground State of the Electron Gas by a Stochastic Method". *Phys. Rev. Lett.*, **45**, 566, (1980).

[16] O. Gunnarsson, M. Jonson, and B. I. Lundqvist. "Descriptions of exchange and correlation effects in inhomogeneous electron systems". *Phys. Rev. B*, **20**, 3136, (1979).

[17] R. M. Martin. *"Electronic Structure: Basic Theory and Practical Methods"*. Cambridge University Press, (2004).

[18] E. Fermi. "Sopra lo Spostamento per Pressione delle Righe Elevate delle Serie Spettrali". *Il Nuovo Cimento (1924-1942)*, **11**, 157, (2008).

[19] H. Hellmann. "A New Approximation Method in the Problem of Many Electrons". *J. Phys. Chem.*, **3**, 61, (1935).

[20] J. C. Phillips and L. Kleinman. "New Method for Calculating Wave Functions in Crystals and Molecules". *Phys. Rev.*, **116**, 287, (1959).

[21] J. R. Chelikowsky and M. L. Cohen. "Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors". *Phys. Rev. B*, **14**, 556, (1976).

[22] D. Vanderbilt. "Optimally smooth norm-conserving pseudopotentials". *Phys. Rev. B*, **32**, 8412, (1985).

[23] A. M. Rappe, K. M. Rabe, E. Kaxiras, and J. D. Joannopoulos. "Optimized pseudopotentials". *Phys. Rev. B*, **41**, 1227, (1990).

[24] N. Troullier and J. L. Martins. "Efficient pseudopotentials for plane-wave calculations". *Phys. Rev. B*, **43**, 1993, (1991).

[25] D. R. Hamann. "Optimized norm-conserving Vanderbilt pseudopotentials". *Phys. Rev. B*, **88**, 085117, (2013).

[26] OPIUM. http://opium.sourceforge.net/index.html.

[27] L. Kleinman and D. M. Bylander. "Efficacious Form for Model Pseudopotentials". *Phys. Rev. Lett.*, **48**, 1425, (1982).

[28] X. Gonze, P. Käckell, and M. Scheffler. "Ghost states for separable, norm-conserving, Iab initioP pseudopotentials". *Phys. Rev. B*, **41**, 12264, (1990).

[29] X. Gonze, R. Stumpf, and M. Scheffler. "Analysis of separable potentials". *Phys. Rev. B*, **44**, 8503, (1991).

[30] D. Vanderbilt. "Soft self-consistent pseudopotentials in a generalized eigenvalue formalism". *Phys. Rev. B*, **41**, 7892, (1990).

[31] P. E. Blöchl. "Projector augmented-wave method". *Phys. Rev. B*, **50**, 17953, (1994).

[32] R. G. Parr and W. Yang. *"Density-Functional Theory of Atoms and Molecules (International Series of Monographs on Chemistry)"*. Oxford University Press, USA, (1994).

[33] R. Landauer. "Spatial Variation of Currents and Fields Due to Localized Scatterers in Metallic Conduction". *IBM J. Res. Dev.*, **1**, 223, (1957).

[34] M. Büttiker, Y. Imry, R. Landauer, and S. Pinhas. "Generalized many-channel conductance formula with application to small rings". *Phys. Rev. B*, **31**, 6207, (1985).

[35] M. Büttiker. "Four-Terminal Phase-Coherent Conductance". *Phys. Rev. Lett.*, **57**, 1761, (1986).

[36] Y. Meir and N. S. Wingreen. "Landauer formula for the current through an interacting electron region". *Phys. Rev. Lett.*, **68**, 2512, (1992).

[37] L. Keldysh. "Diagram Technique for Nonequilibrium Processes". *Sov. Phys. JETP*, **20**, 1018, (1965).

[38] C. Caroli, R. Combescot, P. Nozieres, and D. Saint-James. "Direct calculation of the tunneling current". *J. Phys. C*, **4**, 916, (1971).

[39] J. B. Neaton, M. S. Hybertsen, and S. G. Louie. "Renormalization of Molecular Electronic Levels at Metal-Molecule Interfaces". *Phys. Rev. Lett.*, **97**, 216405, (2006).

[40] K. S. Thygesen and A. Rubio. "Renormalization of Molecular Quasiparticle Levels at Metal-Molecule Interfaces: Trends across Binding Regimes". *Phys. Rev. Lett.*, **102**, 046802, (2009).

[41] J. Taylor, H. Guo, and J. Wang. "Ab initio modeling of quantum transport properties of molecular electronic devices". *Phys. Rev. B*, **63**, 245407, (2001).

[42] M. Brandbyge, J.-L. Mozos, P. Ordejón, J. Taylor, and K. Stokbro. "Density-functional method for nonequilibrium electron transport". *Phys. Rev. B*, **65**, 165401, (2002).

[43] T. Ozaki, K. Nishio, and H. Kino. "Efficient implementation of the nonequilibrium Green function method for electronic transport calculations". *Phys. Rev. B*, **81**, 035116, (2010).

[44] J. Chen, K. S. Thygesen, and K. W. Jacobsen. "Ab initio nonequilibrium quantum transport and forces with the real-space projector augmented wave method". *Phys. Rev. B*, **85**, 155140, (2012).

[45] H. Haug and A. Jauho. *"Quantum Kinetics in Transport and Optics of Semiconductors"*. Springer Series in Solid-State Sciences. Springer Berlin Heidelberg, (1998).

[46] R. Haydock, V. Heine, and M. J. Kelly. "Electronic structure based on the local atomic environment for tight-binding bands". *J. Phys. C*, **5**, 2845, (1972).

[47] R. Haydock, V. Heine, and M. J. Kelly. "Electronic structure based on the local atomic environment for tight-binding bands. II". *J. Phys. C*, **8**, 2591, (1975).

[48] H. H. B. Sørensen, P. C. Hansen, D. E. Petersen, S. Skelboe, and K. Stokbro. "Krylov subspace method for evaluating the self-energy matrices in electron transport calculations". *Phys. Rev. B*, **77**, 155301, (2008).

[49] M. P. L. Sancho, J. M. L. Sancho, and J. Rubio. "Quick iterative scheme for the calculation of transfer matrices: application to Mo (100)". *J. Phys. F*, **14**, 1205, (1984).

[50] M. P. L. Sancho, J. M. L. Sancho, and J. Rubio. "Highly convergent schemes for the calculation of bulk and surface Green functions". *J. Phys. F*, **15**, 851, (1985).

[51] D. H. Lee and J. D. Joannopoulos. "Simple scheme for surface-band calculations. I". *Phys. Rev. B*, **23**, 4988, (1981).

[52] A. Umerski. "Closed-form solutions to surface Green's functions". *Phys. Rev. B*, **55**, 5266, (1997).

[53] M. G. Reuter, T. Seideman, and M. A. Ratner. "Probing the surface-to-bulk transition: A closed-form constant-scaling algorithm for computing subsurface Green functions". *Phys. Rev. B*, **83**, 085412, (2011).

[54] S. Sanvito, C. J. Lambert, J. H. Jefferson, and A. M. Bratkovsky. "General Green's-function formalism for transport calculations with spd Hamiltonians and giant magnetoresistance in Co- and Ni-based magnetic multilayers". *Phys. Rev. B*, **59**, 11936, (1999).

[55] I. Rungger and S. Sanvito. "Algorithm for the construction of self-energies for electronic transport calculations based on singularity elimination and singular value decomposition". *Phys. Rev. B*, **78**, 035407, (2008).

[56] C. J. O. Verzijl, J. S. Seldenthuis, and J. M. Thijssen. "Applicability of the wide-band limit in DFT-based molecular transport calculations". *J. Phys. Chem.*, **138**, 094102, 2013.

[57] K. Varga and S. T. Pantelides. "Quantum Transport in Molecules and Nanotube Devices". *Phys. Rev. Lett.*, **98**, 076804, (2007).

[58] J. A. Driscoll and K. Varga. "Calculation of self-energy matrices using complex absorbing potentials in electron transport calculations". *Phys. Rev. B*, **78**, 245118, (2008).

[59] B. Feldman, T. Seideman, O. Hod, and L. Kronik. "Real-space method for highly parallelizable electronic transport calculations". *Phys. Rev. B*, **90**, 035445, (2014).

[60] N. D. Lang. "Resistance of atomic wires". *Phys. Rev. B*, **52**, 5335, (1995).

[61] M. Di Ventra, S. T. Pantelides, and N. D. Lang. "First-Principles Calculation of Transport Properties of a Molecular Device". *Phys. Rev. Lett.*, **84**, 979, (2000).

[62] M. Di Ventra and N. D. Lang. "Transport in nanoscale conductors from first principles". *Phys. Rev. B*, **65**, 045402, (2001).

[63] K. Hirose and M. Tsukada. "First-Principles Theory of Atom Extraction by Scanning Tunneling Microscopy". *Phys. Rev. Lett.*, **73**, 150, (1994).

[64] K. Hirose and M. Tsukada. "First-principles calculation of the electronic structure for a bielectrode junction system under strong field and current". *Phys. Rev. B*, **51**, 5278, (1995).

[65] T. Ando. "Quantum point contacts in magnetic fields". *Phys. Rev. B*, **44**, 8017, (1991).

[66] H. Joon Choi and J. Ihm. "Ab initio pseudopotential method for the calculation of conductance in quantum wires". *Phys. Rev. B*, **59**, 2267, (1999).

[67] Y. Fujimoto and K. Hirose. "First-principles treatments of electron transport properties for nanoscale junctions". *Phys. Rev. B*, **67**, 195315, (2003).

[68] P. A. Khomyakov and G. Brocks. "Real-space finite-difference method for conductance calculations". *Phys. Rev. B*, **70**, 195402, (2004).

[69] P. A. Khomyakov, G. Brocks, V. Karpan, M. Zwierzycki, and P. J. Kelly. "Conductance calculations for quantum wires and interfaces: Mode matching and Green's functions". *Phys. Rev. B*, **72**, 035450, (2005).

[70] L. Kong, M. L. Tiago, and J. R. Chelikowsky. "Real-space pseudopotential method for electron transport properties of nanoscale junctions". *Phys. Rev. B*, **73**, 195118, (2006).

[71] L. Kong, J. R. Chelikowsky, J. B. Neaton, and S. G. Louie. "Real-space pseudopotential calculations of spin-dependent electron transport in magnetic molecular junctions". *Phys. Rev. B*, **76**, 235422, (2007).

[72] T. B. Boykin. "Generalized eigenproblem method for surface and interface states: The complex bands of GaAs and AlAs". *Phys. Rev. B*, **54**, 8107, (1996).

[73] S. Tsukamoto, K. Hirose, and S. Blügel. "Real-space finite-difference calculation method of generalized Bloch wave functions and complex band structures with reduced computational cost". *Phys. Rev. E*, **90**, 013306, (2014).

[74] J. R. Chelikowsky, N. Troullier, and Y. Saad. "Finite-difference-pseudopotential method: Electronic structure calculations without a basis". *Phys. Rev. Lett.*, **72**, 1240, (1994).

[75] J. R. Chelikowsky, N. Troullier, K. Wu, and Y. Saad. "Higher-order finite-difference pseudopotential method: An application to diatomic molecules". *Phys. Rev. B*, **50**, 11355, (1994).

[76] T. Ono, Y. Egami, and K. Hirose. "First-principles transport calculation method based on real-space finite-difference nonequilibrium Green's function scheme". *Phys. Rev. B*, **86**, 195406, (2012).

[77] T. Sakurai and H. Sugiura. "A projection method for generalized eigenvalue problems using numerical integration". *J. Comput. Appl. Math*, **159**, 119, (2003). 6th Japan-China Joint Seminar on Numerical Mathematics; In Search for the Frontier of Computational and Applied Mathematics toward the 21st Century.

[78] S. Iwase, T. Hoshi, and T. Ono. "Numerical solver for first-principles transport calculation based on real-space finite-difference method". *Phys. Rev. E*, **91**, 063305, (2015).

[79] S. Iwase, C. J. Kirkham, and T. Ono. "Intrinsic origin of electron scattering at the $4H$-SiC(0001)/SiO_2 interface". *Phys. Rev. B*, **95**, 041302, (2017).

[80] R. Takayama, T. Hoshi, T. Sogabe, S.-L. Zhang, and T. Fujiwara. "Linear algebraic calculation of the Green's function for large-scale electronic structure theory". *Phys. Rev. B*, **73**, 165108, (2006).

[81] A. Frommer. "BiCGStab(l) for Families of Shifted Linear Systems". *Computing*, **70**, 87, (2003).

[82] R. Freund. "On conjugate gradient type methods and polynomial preconditioners for a class of complex non-hermitian matrices". *Numerische Mathematik*, **57**, 285, (1990).

[83] S. Yamamoto, T. Sogabe, T. Hoshi, S.-L. Zhang, and T. Fujiwara. "Shifted Conjugate-Orthogonal Conjugate-Gradient Method and Its Application to Double Orbital Extended Hubbard Model". *J. Phys. Soc. Jpn.*, **77**, 114713, (2008).

[84] H. Teng, T. Fujiwara, T. Hoshi, T. Sogabe, S.-L. Zhang, and S. Yamamoto. "Efficient and accurate linear algebraic methods for large-scale electronic structure calculations with nonorthogonal atomic orbitals". *Phys. Rev. B*, **83**, 165103, (2011).

[85] K. Hirose, T. Ono, Y. Fujimoto, and S. Tsukamoto. *"First-Principles Calculations in Real-Space Formalism, Electronic Configurations and Transport Properties of Nanostructures"*. Imperial College Press, London, (2005).

[86] T. Ono and K. Hirose. "Timesaving Double-Grid Method for Real-Space Electronic-Structure Calculations". *Phys. Rev. Lett.*, **82**, 5016, (1999).

[87] Z. Liu, M. Koshino, K. Suenaga, A. c. v. Mrzel, H. Kataura, and S. Iijima. "Transmission Electron Microscopy Imaging of Individual Functional Groups of Fullerene Derivatives". *Phys. Rev. Lett.*, **96**, 088304, (2006).

[88] S. Okada, S. Saito, and A. Oshiyama. "Energetics and Electronic Structures of Encapsulated C_{60} in a Carbon Nanotube". *Phys. Rev. Lett.*, **86**, 3835, (2001).

[89] O. Dubay and G. Kresse. "Density functional calculations for C_{60} peapods". *Phys. Rev. B*, **70**, 165424, (2004).

[90] H. Kondo, H. Kino, and T. Ohno. "Transport properties of carbon nanotubes encapsulating C_{60} and related materials". *Phys. Rev. B*, **71**, 115413, (2005).

[91] R. Haddon, L. Brus, and K. Raghavachari. "Electronic structure and bonding in icosahedral C60". *Chem. Phys. Lett.*, **125**, 459, (1986).

[92] E. Manousakis. "Electronic structure of C_{60} within the tight-binding approximation". *Phys. Rev. B*, **44**, 10991, (1991).

[93] R. Tellgmann, N. Krawez, S.-H. Lin, I. V. Hertel, and E. E. B. Campbell. "Endohedral fullerene production". *Nature*, **382**, 407, (1996).

[94] H. Huang, M. Ata, and Y. Yoshimoto. "Cu@C60 formation in rf-plasma and ring-current induced magnetism of C60". *Chem. Commun.*, 1206, (2004).

[95] J. R. Heath, S. C. O'Brien, Q. Zhang, Y. Liu, R. F. Curl, F. K. Tittel, and R. E. Smalley. "Lanthanum complexes of spheroidal carbon shells". *J. Am. Chem. Soc.*, **107**, 7779, 1985.

[96] M. Saunders, H. A. Jimenez-Vazquez, R. J. Cross, S. Mroczkowski, M. L. Gross, D. E. Giblin, and R. J. Poreda. "Incorporation of helium, neon, argon, krypton, and xenon into fullerenes using high pressure". *J. Am. Chem. Soc.*, **116**, 2193, (1994).

[97] K. Komatsu, M. Murata, and Y. Murata. "Encapsulation of Molecular Hydrogen in Fullerene C60 by Organic Synthesis". *Science*, **307**, 238, (2005).

[98] V. Tilak. "Inversion layer electron transport in 4H-SiC metal-Oxide-Semiconductor field-effect transistors". *Phys. Status Solidi A*, **206**, 2391, (2009).

[99] T. Kimoto. "Material science and device physics in SiC technology for high-voltage power devices". *Jpn. J. Appl. Phys.*, **54**, 040103, (2015).

[100] G. Liu, B. R. Tuttle, and S. Dhar. "Silicon carbide: A unique platform for metal-oxide-semiconductor physics". *Appl. Phys. Rev.*, **2**, 021307, (2015).

[101] S. Dhar, S. Haney, L. Cheng, S.-R. Ryu, A. K. Agarwal, L. C. Yu, and K. P. Cheung. "Inversion layer carrier concentration and mobility in 4H-SiC metal-oxide-semiconductor field-effect transistors". *J. Appl. Phys.*, **108**, 054509, (2010).

[102] D. Okamoto, H. Yano, T. Hatayama, and T. Fuyuki. "Removal of near-interface traps at SiO2/4H-SiC (0001) interfaces by phosphorus incorporation". *Appl. Phys. Lett.*, **96**, 203508, (2010).

[103] R. H. Kikuchi and K. Kita. "Fabrication of SiO2/4H-SiC (0001) interface with nearly ideal capacitance-voltage characteristics by thermal oxidation". *Appl. Phys. Lett.*, **105**, 032106, (2014).

[104] D. J. Lichtenwalner, L. Cheng, S. Dhar, A. Agarwal, and J. W. Palmour. "High mobility 4H-SiC (0001) transistors using alkali and alkaline earth interface layers". *Appl. Phys. Lett.*, **105**, 182107, (2014).

[105] S. Potbhare, N. Goldsman, G. Pennington, A. Lelis, and J. M. McGarrity. "Numerical and experimental characterization of 4H-silicon carbide lateral metal-oxide-semiconductor field-effect transistor". *J. Appl. Phys.*, **100**, 044515, (2006).

[106] Y.-i. Matsushita, S. Furuya, and A. Oshiyama. "Floating Electron States in Covalent Semiconductors". *Phys. Rev. Lett.*, **108**, 246404, (2012).

[107] Y.-i. Matsushita and A. Oshiyama. "Interstitial Channels that Control Band Gaps and Effective Masses in Tetrahedrally Bonded Semiconductors". *Phys. Rev. Lett.*, **112**, 136403, (2014).

[108] C. J. Kirkham and T. Ono. "First-Principles Study on Interlayer States at the 4H-SiC/SiO2 Interface and the Effect of Oxygen-Related Defects". *J. Phys. Soc. Jpn.*, **85**, 024701, (2016).

[109] H. Hara, Y. Morikawa, Y. Sano, and K. Yamauchi. "Termination dependence of surface stacking at $4H$-SiC(0001)-1×1: Density functional theory calculations". *Phys. Rev. B*, **79**, 153306, (2009).

[110] K. Sawada, J.-I. Iwata, and A. Oshiyama. "Magic angle and height quantization in nanofacets on SiC(0001) surfaces". *Appl. Phys. Lett.*, **104**, 051605, (2014).

[111] K. Arima, H. Hara, J. Murata, T. Ishida, R. Okamoto, K. Yagi, Y. Sano, H. Mimura, and K. Yamauchi. "Atomic-scale flattening of SiC surfaces by electroless chemical etching in HF solution with Pt catalyst". *Appl. Phys. Lett.*, **90**, 202106, (2007).

[112] P. Liu, G. Li, G. Duscher, Y. K. Sharma, A. C. Ahyi, T. Isaacs-Smith, J. R. Williams, and S. Dhar. "Roughness of the SiC/SiO2 vicinal interface and atomic structure of the transition layers". *J. Vac. Sci. Technol. A*, **32**, 060603, (2014).

[113] I. Iskandarova, K. Khromov, A. Knizhnik, and B. Potapkin. "The role of neutral point defects in carrier mobility degradation in bulk 4H-SiC and at 4H-SiC/SiO2 interface: First-principles investigation using Green's functions". *J. Appl. Phys.*, **117**, 175703, (2015).

[114] S. Saito. PhD thesis, Osaka University, 2013.

[115] T. Akiyama, A. Ito, K. Nakamura, T. Ito, H. Kageshima, M. Uematsu, and K. Shiraishi. "First-principles investigations for oxidation reaction processes at 4H-SiC/SiO2 interface and its orientation dependence". *Surf. Sci.*, **641**, 174, (2015).

[116] T. Ono and S. Saito. "First-principles study on the effect of SiO2 layers during oxidation of 4H-SiC". *Appl. Phys. Lett.*, **106**, 081601, (2015).

[117] T. Ono, C. J. Kirkham, S. Saito, and Y. Oshima. "Theoretical and experimental investigation of the atomic and electronic structures at the $4H$-SiC(0001)/SiO$_2$ interface". *Phys. Rev. B*, **96**, 115311, (2017).

[118] J. M. Knaup, P. Deák, T. Frauenheim, A. Gali, Z. Hajnal, and W. J. Choyke. "Theoretical study of the mechanism of dry oxidation of $4H$-SiC". *Phys. Rev. B*, **71**, 235321, (2005).

[119] A. Gavrikov, A. Knizhnik, A. Safonov, A. Scherbinin, A. Bagatur'yants, B. Potapkin, A. Chatterjee, and K. Matocha. "First-principles-based investigation of kinetic mechanism of SiC(0001) dry oxidation including defect generation and passivation". *J. Appl. Phys.*, **104**, 093508, (2008).

[120] V. V. Afanas'ev and A. Stesmans. "Interfacial Defects in SiO$_2$ Revealed by Photon Stimulated Tunneling of Electrons". *Phys. Rev. Lett.*, **78**, 2437, (1997).

[121] N. Kobayashi and M. Tsukada. "Numerical Method for Local Density of States and Current Density Decomposed into Eigenchannels in Multichannel System". *Jpn. J. Appl. Phys.*, **38**, 3805, (1999).

[122] S. Iwase, Y. Futamura, A. Imakura, T. Sakurai, and T. Ono. "Efficient and Scalable Calculation of Complex Band Structure Using Sakurai-Sugiura Method". In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '17, 40:1, New York, NY, USA, 2017. ACM.

[123] S. Iwase, Y. Futamura, A. Imakura, T. Sakurai, S. Tsukamoto, and T. Ono. "Contour integral method for obtaining the self-energy matrices of electrodes in electron transport calculations". *Phys. Rev. B*, **97**, 195449, (2018).

[124] H. H. B. Sørensen, P. C. Hansen, D. E. Petersen, S. Skelboe, and K. Stokbro. "Efficient wavefunction matching approach for quantum transport calculations". *Phys. Rev. B*, **79**, 205322, (2009).

[125] S. Brück, M. Calderara, M. H. Bani-Hashemian, J. VandeVondele, and M. Luisier. "Efficient algorithms for large-scale quantum transport calculations". *J. Chem. Phys.*, **147**, 074116, (2017).

[126] T. Ono and S. Tsukamoto. "Real-space method for first-principles electron transport calculations: Self-energy terms of electrodes for large systems". *Phys. Rev. B*, **93**, 045421, (2016).

[127] E. Polizzi. "Density-matrix-based algorithm for solving eigenvalue problems". *Phys. Rev. B*, **79**, 115112, (2009).

[128] S. E. Laux. "Solving complex band structure problems with the FEAST eigenvalue algorithm". *Phys. Rev. B*, **86**, 075103, (2012).

[129] W.-J. Beyn. "An integral method for solving nonlinear eigenvalue problems". *Linear Algebra Its Appl.*, **436**, 3839, (2012). Special Issue dedicated to Heinrich Voss's 65th birthday.

[130] J. Asakura, T. Sakurai, H. Tadano, T. Ikegami, and K. Kimura. "A numerical method for nonlinear eigenvalue problems using contour integrals". *JSIAM Lett.*, **1**, 52, (2009).

[131] T. Sakurai, Y. Futamura, and H. Tadano. "Efficient Parameter Estimation and Implementation of a Contour Integral-Based Eigensolver". *J. Algorithm. Comput. Technol.*, **7**, 249, (2013).

[132] S. Yokota and T. Sakurai. "A projection method for nonlinear eigenvalue problems using contour integrals". *JSIAM Lett.*, **5**, 41, (2013).

[133] A. Imakura and T. Sakurai. "Block SS-CAA: A complex moment-based parallel nonlinear eigensolver using the block communication-avoiding Arnoldi procedure". *Parallel Computing*, **74**, 34, (2018). Parallel Matrix Algorithms and Applications (PMAA'16).

[134] Y. Futamura, A. Imakura, and T. Sakutai. (Private communication).

[135] Y. Hasegawa, J.-I. Iwata, M. Tsuji, D. Takahashi, A. Oshiyama, K. Minami, T. Boku, F. Shoji, A. Uno, M. Kurokawa, H. Inoue, I. Miyoshi, and M. Yokokawa. "First-principles Calculations of Electron States of a Silicon Nanowire with 100,000 Atoms on the K Computer". In *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '11, 1:1, New York, NY, USA, (2011). ACM.

[136] z-Pares: Parallel Eigenvalue Solver. http://zpares.cs.tsukuba.ac.jp/.

[137] A. Calzolari, C. Cavazzoni, and M. Buongiorno Nardelli. "Electronic and Transport Properties of Artificial Gold Chains". *Phys. Rev. Lett.*, **93**, 096404, (2004).

[138] M. Strange, I. S. Kristensen, K. S. Thygesen, and K. W. Jacobsen. "Benchmark density functional theory calculations for nanoscale conductance". *J. Phys. Chem.*, **128**, 114714, (2008).

[139] K. Takeda and K. Shiraishi. "Theoretical possibility of stage corrugation in Si and Ge analogs of graphite". *Phys. Rev. B*, **50**, 14916, (1994).

[140] S. Lebègue and O. Eriksson. "Electronic structure of two-dimensional crystals from ab initio theory". *Phys. Rev. B*, **79**, 115409, (2009).

[141] L. Tao, E. Cinquanta, D. Chiappe, C. Grazianetti, M. Fanciulli, M. Dubey, A. Molle, and D. Akinwande. "Silicene field-effect transistors operating at room temperature". *Nat. Nanotechnol.*, **10**, 227, (2015).

[142] X. Li, J. T. Mullen, Z. Jin, K. M. Borysenko, M. Buongiorno Nardelli, and K. W. Kim. "Intrinsic electrical transport properties of monolayer silicene and MoS_2 from first principles". *Phys. Rev. B*, **87**, 115418, (2013).

[143] M. P. Lima, A. Fazzio, and A. J. R. d. Silva. "Interfaces between buckling phases in silicene: Ab initio density functional theory calculations". *Phys. Rev. B*, **88**, 235413, (2013).

Acknowledgement