University of Regensburg   Institute of Theoretical Physics

Analysis and efficient computational scheme for bound and continuum excitons from the Bethe-Salpeter equation

Master's thesis

by

Marc Aichner

January 22, 2024

Reviewer:	Prof. Dr. Ferdinand Evers (University of Regensburg, Germany)
Second Reviewer:	Dr. Lucia Reining (École Polytechnique Paris, France)
Supervisors:	Dr. Matteo Gatti, Dr. Lucia Reining, Prof. Dr. Ferdinand Evers

Contents

Table 1: Notation

<div style="text-align:center">

Acronyms

</div>

eMB	electron many-body
DFT	density functional theory
TDDFT	Time-dependent density functional theory
KS	Kohn-Sham
IP	Independent-particle
RPA	Random phase approximation
HF	Hartree-Fock
TDHF	Time-dependent Hartree-Fock
BSE	Bethe-Salpeter equation
EELS	Electron energy-loss spectroscopy
A	Adiabatic, $O(1,2) = O(\boldsymbol{r}_1, \boldsymbol{r}_2)\delta(t_1 - t_2)$
AGW	adiabatic/static GW-approximation
TDA	Tamm-Dancoff approximation
BZ	Brillouin zone
EoM	Equation of motion
QP	Quasi-particle

General symbol notation

$\underline{\underline{O}}$	general matrix
$\underline{o}$	general vector
$\hat{O}_I$	operator in interaction picture
$\hat{O}_H$	operator in Heisenberg picture
$\hat{O}$	operator in Schroedinger picture
$\hat{\psi}^\dagger, \hat{\psi}$	field operators for fermions

Coordinates and indices

$\boldsymbol{r}$	position in real space
$\boldsymbol{R}$	realspace lattic vector
$\boldsymbol{q}, \boldsymbol{k}$	reciprocal vector of the first Brillouin zone
$\boldsymbol{G}$	reciprocal lattice vector

t	time
ω	frequency
Ω	$= \omega + i\Gamma$ complex frequency
σ	spin
m_σ	spin quantum number
j	$= (\boldsymbol{r}_j, t_j, \sigma_j)$ combined space, time, spin coordinates
$\bar{\boldsymbol{r}}$	$= \{\boldsymbol{r}_i\}$ N-electron coordinate
$\bar{\boldsymbol{R}}$	$= \{\boldsymbol{R}_I\}$ N-ion coordinate
$\boldsymbol{r}_i$	position of electron i
$\boldsymbol{R}_I$	position of ion I
$\tilde{n}_i$	general band quantum number
c	conduction band
v	valence band
n_i	$= (\tilde{n}_i, \boldsymbol{k}_i)$ merge of band quantum number and vector of the first Brillouin zone
μ	$= (v, c, \boldsymbol{k})$ Combination of valence, conduction band index and vector of the first Brillouin zone
λ	eigenvalue index

Mathematical operations

diag	diagonal-part of a matrix, $\mathrm{diag}(\underline{M}) = \mathrm{diag}(M_{00}, M_{11}, \ldots, M_{nn})$
Diag	representation of a vector as diagonal matrix $\mathrm{Diag}(\underline{v}) = \mathrm{diag}(v_{00}, v_{11}, \ldots, v_{nn})$
Tr	trace
Re, Im	real and imaginary part
F_T^t	Fourier transformation in variable time
$\langle \hat{O} \rangle$	expectation value
t^+	$= \lim_{\nu \to 0} t + \nu$
$\bar{j} = \int d_j$	$= \sum_{\sigma_j} \int d\boldsymbol{r}_j \int t_j$ varibalen integrated over
T	time-ordering operator

Physical quantities

T	temperature
β	inverse temperature, $1/k_B T$
μ_T	chemical potential
V_0	volume unit cell
N_k	number of k-points
Ω_v	volume crytal, $N_k V_0$
Γ, γ	broadening

Correlation functions and related quantities

G_0	independent-particle Green's function
G	one-body Green's function
G_2	two-particle Green's function
C	correlation function
L	two-particle correlation function
χ	general linear response function, susceptibility
ϵ	general dielectric function
n	density

Potentials and interaction variables

v_c	Coulomb potential
W	screened Coulomb potential
v_c^0	long-range part of the Coulomb potential, $v_c(\boldsymbol{q}+\boldsymbol{G}=0)$
$\bar{v}_c$	short-range part of the Coulomb potential, $v_c - v^0$
v_{eff}	effective potential
v_{ext}	external potential
v_{xc}	exchange correlation potential
v_H	Hartree potential
v_{ind}	induced potential
v_{tot}	total classic potential, $v_{per} + v_{ind}$
v_{per}	perturbation potential
Σ	self-energy $v_H + \Sigma_{xc}$
Σ_{xc}	exchange correlation part of self-energy
Ξ	interaction kernel $i\frac{\delta\Sigma}{\delta G}$

Defined quantities in this work

$\tilde{\rho}$	dipol transition matrix elements
E_λ	excitonic eigenenergy with value $\lambda \in \mathbb{N}$
$\underline{A}_\lambda$	excitonic eigenvector in reciprocal space belonging to E_λ, envelope function
A_λ^μ	μ-th component of the excitonic eigenvector with $\mu \in \mathbb{N}$ and one-to-one mapping to $\mu = (v, c, \boldsymbol{k})$
$\underline{\underline{H}}$	Hamilton matrix
$O_{\mu\mu'}$	μ,μ' component of the matrix $\underline{\underline{O}}$

$\underline{\underline{H}}^0$ independent-particle transition Hamiltonian $H^0_{\mu\mu'} = E^0_\mu \delta_{\mu\mu'}$

E^0_μ independent-particle transition energies, $E^0_\mu :=$ $\left(E^0_{c\boldsymbol{k}} - E^0_{v\boldsymbol{k}}\right)\delta_{v,v'}\delta_{c,c'}\delta_{\boldsymbol{k},\boldsymbol{k}'}$, $\mu \in \mathbb{N}$ and one-to-one mapping to $\mu = (v, c, \boldsymbol{k})$

$M^{IP}(E^0_\mu)$ set of independent-particle transition energies

$\underline{\underline{h}}$ interaction Hamiltonian, AGW-BSE: $\underline{\underline{h}} = \underline{\underline{\Xi}}^{AGW}$

$\underline{\underline{\tilde{h}}}$ off-diagonal part of the interaction Hamiltonian, $\underline{\underline{\tilde{h}}} = \underline{\underline{h}} - \text{diag}(\underline{\underline{h}})$

$\underline{\underline{\Delta\tilde{E}}}^{-1}_\lambda$ energy difference diagonal matrix, $\left(\Delta\tilde{E}_{\mu\mu'}(E_\lambda)\right)^{-1} := \frac{(1-\delta_{\mu\lambda})}{E_\lambda - (E^0_\mu + h_{\mu\mu})}$

$\underline{\underline{\Delta E(\gamma)}}^{-1}_\lambda$ energy difference diagonal matrix, $(\Delta E_{\mu\mu'}(E_\lambda + i\gamma)^{-1} := \frac{1}{(E_\lambda + i\gamma)-(E^0_\mu + h_{\mu\mu})}$

$\underline{\underline{1}}$ identity matrix

$\underline{\tilde{h}}_\lambda$ vector constructed from the λ column of $\underline{\underline{\tilde{h}}}$, $(\tilde{h}_{\mu\lambda} = (1 - \delta_{mu\lambda})h_{\mu\lambda}$

$\underline{H}_\lambda$ vector constructed from the λ column of $\underline{\underline{H}}$, $(H_{\mu\lambda} = H_{\mu\lambda})$

$\underline{1}_\lambda$ one-vector $1_{\mu\lambda} = \delta_{\mu\lambda}$

$\underline{\tilde{H}}_\lambda$ vector defined as $\underline{\tilde{H}}_\lambda := (\underline{\underline{H}} - E_\lambda\underline{\underline{1}})\underline{1}_\lambda$

e_λ matrix structure factor

Definition of terminology

optical bandgap energies in an absorption spectrum which are smaller than $\min(M^{IP}(E^0_\mu))$

bound exciton exciton with an energy lower than $\min(M^{IP}(E^0_\mu))$

continuum exciton exciton with an energy larger than $\min(M^{IP}(E^0_\mu))$

Chapter 1

Introduction

> "The great advances in science are often, perhaps always, based on the fact that a question that was not asked before is now being asked, and successfully."
>
> Carl Friedrich von Weizsäcker

The exciting field of theoretical spectroscopy explores how matter interacts with electromagnetic radiation, shedding light on various phenomena, including the intriguing effects of excitons. In an independent-particle picture, light absorption can be explained in terms of interband transitions, with electrons excited from valence to conduction bands. When interactions are included in the theoretical description, the attraction between the electrons in the conduction band and the holes left in the valence band gives rise to excitonic effects, with modification of peak positions and intensities in the spectra. Excitons play a crucial role in diverse applications [1] like non-linear optics [2] and electrically driven light emission [3]. Furthermore, new experimental observations, such as exciton-exciton transitions [4–7], interlayer excitons [8–10], or the dynamics of excitons [11–14], lead to increasing interest and relevance of the scientific community in order to understand, explain and predict these phenomena [15–18].

The fundamental equation in theoretical spectroscopy from an ab inito point of view is the Bethe-Salpeter equation [19], which describes the propagation of correlated electron-hole pairs by a two-particle correlation function, which depends on 4-points in the variables of space, time, and spin. A prevalent approach in the community [20–22] is to assume that the electron-hole interaction is instantaneous [23]. Based on this assumption, the complicated Bethe-Salpeter equation for the 4-point function can be transformed into an effective two-particle Hamiltonian whose eigenvalues and eigenvectors represent the excitons' eigenenergies and envelope functions. While computationally solvable, the diagonalization of the Hamiltonian involves high numerical effort since the typical dimension of the Hamiltonian for simple materials such as silicon and lithium fluoride is of the order $N \sim 10^4 - 10^5$, where N is the number of basis functions, i.e., of independent-particle transitions. However, if one is interested in an extended range of the excitation spectrum or complex materials such as layered materials like transition-metal dichalcogenides, which are constantly coming into the focus of science, a numerical solution employing classical diagonalization algorithms persists

as a challenge, with an error tolerance of 100 meV or more required for eigenenergies [20, 24, 25]. Due to the formidable problem of diagonalizing large matrices, which arises in many areas of applied mathematics, a large number of algorithms for the calculation of eigenvalues and eigenvectors have been developed over the last decades [24, 26–34], as well as methods to circumvent specific isolated problems of theoretical spectroscopy for more numerically favorable solutions, such as (a) the double grid method [25, 35] - designed to expedite k-grid convergence - or (b) subspace reduction of the Hamiltonian [36] or (c) mapping procedures for including spin-orbit coupling comparatively cheap [37] or (d) an algorithm for generating low-cost approximations of the first few eigenvectors and eigenvalues [24]. However, the most efficient algorithms are still not satisfactory. They often do not transparently display physics and therefore do not ease analysis. For this reason, in this work, we want to explore alternative approaches that enable the extraction of crucial information computationally efficiently and additionally allow us to perform analyses in order to develop a deeper understanding. In doing so, we focus on developing analytical and perturbation-theoretical methodologies for understanding the role of excitonic effects in absorption spectra, eigenvectors, and eigenvalues.

Initiating this exploration will involve a comprehensive study of the impact of excitonic effects on the absorption spectrum using a Neumann series (Ch. 5.1). This method provides insights into the crucial components influencing the emergence of an absorption spectrum with excitonic effects. Furthermore, it will offer a potential avenue for numerically advantageous calculations beyond standard theoretical approximations. Considerations include calculations for the dielectric function at finite frequencies.

Armed with this foundational knowledge, we will delve into fundamental inquiries, specifically addressing how the strength of excitonic effects influences eigenvalues and eigenvectors. The analysis will involve developing analytical equations beyond the conventional eigenvalue equation (Ch. 5.2), paving the way for analytical studies of the interplay between the strength of excitonic effects and the corresponding eigenvectors and eigenvalues using straightforward models (Ch. 5.3).

Subsequently, we will leverage this understanding to formulate the continuum approximation of excitonic eigenvalues (Ch. 5.4). This approach enables precise calculations of excitonic eigenenergies in the continuum, with negligible numerical costs, applicable to materials with weak and strong excitonic effects, even when the reciprocal space is discretized, as this is usually done in practice. Additionally, we will demonstrate the calculation of eigenvectors for materials with weak excitonic effects in practical applications, exhibiting a numerical scaling of approximately $O(N^2)$ through a modified perturbation-theoretic approach (Ch. 5.6).

Finally, we will show the versatility of this approach by accurately determining the lowest eigenvalues of strongly bound excitons, along with associated eigenvectors, at a numerical scaling of approximately $O(N^2)$ and an accuracy of millielectronvolt.

Chapter 2

Introduction to the field of optical spectroscopy

> "Learn to see everything. Become aware of the unclear.
> Though the path is 1000 miles long, walk it step by step.
> Always remember that!"
>
> Miyamoto Musashi

Because a summit climb never begins at the summit but always at the foot of a mountain. So we, too, will not begin by answering the main question immediately but by walking the path, step by step. This implies that we must break down our initial question into smaller sub-questions.

Although our initial question is very mathematical, we want to use the answer to this question in physics. Therefore, the most crucial question is: *In which observables are we interested in?*

The framework we work on is theoretical spectroscopy, a part of condensed matter theory in which an incoming particle beam interacts with a material. To fill this framework with physics and thus with understanding, we start our description from the macroscopic world of theoretical electrodynamics. A macroscopic description is possible since, typically materials properties are studied by having use of spectroscopies, where a beam of particles, such as a macroscopic object, is sent onto a material sample. It is a different story if such a simplification is also possible for the material. However, for the moment, we assume it is. Therefore, our primary concern is to study how a material responds to an external field in an entirely classical description. On a fundamental level, this means we want to study how a bunch of electrons and ions behave if we perturbed them with an external field, like light. An elementary description, but a nice picture, is to assume that electrons and ions form so-called dipoles p in an external field E_{per}. Roughly speaking, this could be put into the picture of tug-of-war between the attractive interaction between electrons and ions and their interaction with the external field. Since our material is a macroscopic object, we will find many tugs-of-war, so to say, a bunch of dipoles, which we call polarization density $P := P(E_{per})$. Since dipoles are nothing more than transmitters of electric fields, they generate a

field $E_{ind}(E_{per}) = -4\pi P$ that counteracts the external one.

If we assume a weak perturbative field, so a weak total electric field $E_{tot} = E_{per} + E_{ind}(E_{per})$, we can express the polarization density as a Taylor expansion:

$$P = P_0 + \frac{\partial P}{\partial E_{tot}} E_{tot} + \frac{\partial^2 P}{\partial^2 E_{tot}} E_{tot}^2 + \ldots =: P_0 + \chi_1 E_{tot} + \chi_2 E_{tot}^2 + \ldots.$$

To breathe some physics into this equation, let us assume that $\chi_1 E_{tot} \gg \chi_{n \neq 1} E_{tot}^{n \neq 1}$ holds, such that we can simplify the equation to

$$P \approx P_0 + \frac{\partial P}{\partial E_{tot}} E_{tot} =: P_0 + \chi_1 E_{tot}.$$

We can observe a term describing a ground polarization density P_0 without the system being disturbed from the outside, and we have a first-order term describing the change in polarization density with the external field. So, the second term describes how the material will respond when the external field changes, leading to the general name of this function $\chi_1 = \frac{\partial P}{\partial E_{tot}}$ response function. This indisputable quantity is at the heart of many spectroscopic applications and will accompany us throughout this work.

Many exciting effects can be observed in a wide variety of orders. However, this work will deal exclusively with materials and effects that require only a first-order description. This assumption also includes that the zero-order term disappears, which is the case for materials that do not experience spontaneous intrinsic electric polarization (ferroelectricity). Thus, we obtain the simple equation

$$P \approx \frac{\partial P}{\partial E_{tot}} E_{tot} =: \chi_1 E_{tot} =: 4\pi \chi_e E_{tot},$$

where χ_e is called the dielectric susceptibility.

We have claimed above that we would describe everything with electrodynamics, but we have only considered static equations so far. In order to motivate the necessity of dynamics, we will resort to an analogy: Let us imagine that there is a huge lake, and we are sitting on the edge of the lake. Now, at time t', someone sitting on a boat throws a rock into the lake. After a certain period, $t - t'$ we see the waves coming towards us. It is the same when we shoot a beam at the center of the material and measure somewhere else; we have to wait for the system's response $\chi_e(t, t')$, $t > t'$.

Since we now have a dynamic description of the response function, which describes the system's response to an outer perturbative field, we should ask ourselves: *What is measured in an experiment?* As one can imagine, it is extremely complicated to directly observe how the polarisation density changes and thus measure χ_e because it would mean we are watching the tug-of-war between interacting ions and electrons and the electric field. However, looking at how the electric field changes during the tug-of-war is much easier

$$E_{per} = E_{tot} - E_{ind} = E_{tot} + 4\pi P = E_{tot} + 4\pi \chi_e E_{tot} = (1 + 4\pi \chi_e) E_{tot}.$$

It is of practical use to define a new quantity, the dielectric constant, denoted as ϵ and defined as

$$\epsilon := 1 + 4\pi \chi_e. \tag{2.1}$$

This quantity describes how the entire system, i.e. the system without material $E_{per} = E_{tot}$ and with material $E_{per} = \epsilon E_{tot}$, will behave. Furthermore, since, on the one hand, the field E_{per} is

known, and on the other hand, the field E_{tot} is measured. Thus, the dielectric constant ϵ or, more generally, the dielectric function $\epsilon(t, t')$ can be measured.

Now that we have mapped out the theoretical field we are working with on the macroscopic level and understand the crucial quantities for theory and experiment, it is time to bring the quantities to life with some optical experiments. This will help us understand what observations we can make and discuss empirically what effects we need to consider as we dive deeper into theory.

2.1 Optical spectroscopy

The fascinating field of spectroscopy can be categorized into two groups based on whether the sample undergoes a change in its number of electrons. (a) Direct and inverse photoemission experiments belong to the category where the number of electrons in the sample changes, while in (b) absorption and electron energy-loss spectroscopy (EELS) it does not [38].

In this work, our focus will be on the latter category, which we will discuss in the following.

Absorption spectroscopy

In absorption spectroscopy, a macroscopic beam of photons is sent onto a material for example using a laser or a lamp. At the moment when such a photon interacts with a sample, e.g. a solid, it can either be absorbed, reflected, or transmitted. In the case of absorption, the photon's energy is utilized to create a neutral excitation within the system.

In an independent-particle description, absorbing a photon would mean exciting an electron from an occupied state to an empty state while conserving its crystal momentum (optical photons possess negligible momentum).

If we consider interactions between the particles, the consequences of excitation are manifold; a fantastic phenomenon that can be observed is the formation of so-called excitons. Excitons arise when an excited electron in the conduction band engages in an attractive interaction with its corresponding hole in the valence band. Of course, the question now arises: *As to how such effects can affect an absorption measurement?* To answer this question, we should consider what quantity we will observe. Since the measurand is a macroscopic quantity, we can again consult classical electrodynamics. From the perspective of classical electrodynamics, one could show that the absorption coefficient is given by [38, 39]:

$$\alpha = \frac{2\omega n_2}{c} = \frac{\epsilon_2}{n_1 c},$$

with c is the speed of light and ϵ_i, $i \in 1, 2$ represents the real and imaginary parts of the dielectric constant, while n_i, $i \in 1, 2$ corresponds to the refractive index n_1 and the extinction coefficient n_2. The absorption coefficient shows that assuming $n_1 = const.$, absorption is directly proportional to the imaginary part of the dielectric function. Since, in equilibrium, ϵ depends on the time difference $t - t'$, as we will discuss later, it could be Fourier transformed into the frequency space $\alpha(\omega) \sim \epsilon_2(\omega) = \text{Im}(\epsilon(\omega))$. With this quantity in our toolbox, we can now ask ourselves how an absorption spectrum builds up and how electron-hole pairs can influence this spectrum. To do this, let us carry out a little thought experiment:

We have a laser with a tuneable frequency (ω) and a material consisting of two valence and two conduction levels with energies $\omega_{c_2} > \omega_{c_1} > \omega_{v_2} > \omega_{v_1}$. In this simple case, we would have four transitions where photons with frequencies e.g. $\omega_{c_2} - \omega_{v_1} > \omega_{c_2} - \omega_{v_2}, > \omega_{c_1} - \omega_{v_1} > \omega_{c_1} - \omega_{v_2}$ could be absorbed. Since we neglect inter-particle interactions, we would observe four very sharp delta peaks, similar to the absorption spectrum of an atom. According to Fermi's golden rule, the intensity of the peaks would depend on the strength of the overlap between the valence band and the final conduction band into which the electron is excited. Now, we turn on the interaction between electrons and holes. Since this is an attractive interaction, the energy differences between the initial and final states will become smaller, so we would observe the spectrum shifting towards smaller frequencies. Furthermore, the overlap between the states can also change since we have changed the environment of the electrons and holes with the excitation. For this reason, stronger or weaker peaks can be observed. As we can imagine, depending on the strength of the interaction between electrons and holes (e-h interaction), there can be minor to massive changes in the absorption spectrum.

To see how the whole thing behaves in real materials, we will extend our thought experiment to two real materials with more than four bands: silicone and lithium fluoride.
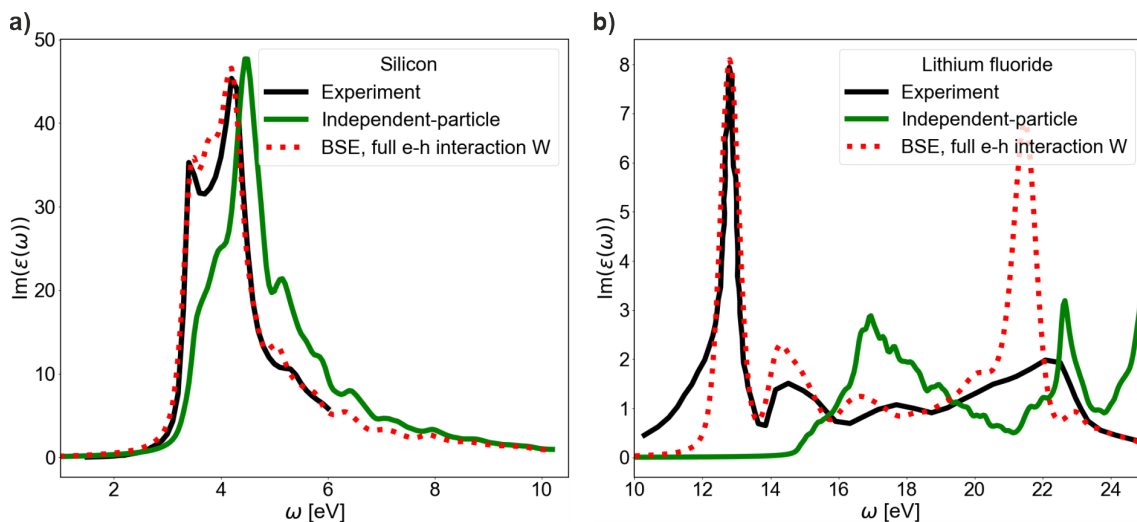
Figure 2.1: Calculated (red) and measured (black) absorption spectra, represented by the frequency-dependent imaginary part of the dielectric function $\text{Im}(\epsilon(\omega))$, for bulk **a)** silicon and **b)** lithium fluoride. The comparison shows that an independent particle spectrum (Eq. (3.35)) (green) cannot describe all the effects seen in a measurement. Instead, it is necessary to use a higher level theory, taking into account the e-h interaction (Eq. (3.40)) (red), to consider excitonic effects. The calculations were made according to the values given in Appendix A and calculated on an 8k grid (2048 k-points) using a Kohn-Sham band structure. A Gaussian-broadening of 0.0184/0.0085 [Ha] was used for silicon/lithium fluoride to compensate for the finite gridsize and to account for experimental resolution. Experimental data were taken from [40] for silicon and from [41] for lithium fluoride.

In figure 2.1, we display the imaginary part of the dielectric function for bulk Si a) and LiF b). Two distinct theoretical curves are evident: one labeled as e-h-interaction (black) and the other

as independent-particle (IP) (green). The IP curve represents the spectrum of non-interacting particles and simple transitions between valence and conduction bands. This means, in particular, that in the region where the spectrum assumes an intensity of zero, no visible excitations can be found. Furthermore, the first visible excitation between valence and conduction band is called the optical bandgap. It is important to note that the optical bandgap only corresponds to the fundamental bandgap if this corresponds to a direct transition (direct bandgap) and this transition is also allowed.

However, as we can see the experimental results (black solid curve) are far from such a simple description.

In the case of Si (Fig. 2.1 a)), calculated IP and measured spectra agree pretty well at higher frequencies but differ rapidly in the region between 3 to 4 eV.

In contrast, in LiF (Fig. 2.1 b)) we immediately observe a strong change of the IP spectrum towards the experimental one; in particular, we see a very pronounced peak at an energy of 12.8 eV, which is about 1.5 eV smaller than the original optical band edge.

As we have already considered in our thought experiment, this strong difference between the two materials is related to the strength of the electron-hole interaction. It will be of practical use to classify these observations in the following. Excitations which are within the optical bandgap, and are not captured by an independent particle spectrum, will be called bound excitons and excitations within the continuum of excitations, the independent particle spectrum, will be called continuum excitons.

After getting a first impression about absorption spectroscopy and especially about the need to consider excitonic effects in theoretical calculations so that theory and experiment coincide, we will now turn to the second optical experiment the EELS.

Electron Energy-loss spectroscopy

Unlike absorption spectroscopy measurements, in which a beam of photons is shot at a material, in EELS, a beam of high energy electrons is shot at the material. The advantage of such a measurement is that it allows observing various effects through which the electrons transfer their energy and momentum. Possible observations include phonons, plasmons, intra-interband excitations, and inner shell ionisations. However, we will limit ourselves here to the low energy range up to about 50 eV, where information about the band structure and dielectric properties of the sample, particularly plasmons, becomes visible. At this point, the first question is: *What are plasmons?* Plasmons are collective oscillations of excited electrons. These collective excitations probably correspond most closely to our lake-wave analogy from the introductory section.

Now that we have briefly discussed what to expect on a phenomenological basis, we should ask ourselves: *How can we put such an experiment into an equation?*

To answer this question, it is practical to clarify the phenomenological process and to package the individual quantities in symbols. The process can be explained as follows: A high-energy electron, treated as a non-relativistic particle, undergoes a scattering interaction with the sample from the moment it approaches. The probability that such an electron, with velocity $\boldsymbol{v}$, transfers energy E_T

and momentum $\boldsymbol{q}$ to an excitation of the system per unit time can be expressed by the energy-loss rate per unit time [39]:

$$\frac{dE_T}{dt} = \frac{1}{(2\pi)^3} \int d\omega d\boldsymbol{q}, \frac{\omega}{\boldsymbol{q}^2} \mathrm{Im}\left[-\frac{1}{\epsilon_M(\boldsymbol{q},\omega)}\right]\delta(\omega + \boldsymbol{q}\boldsymbol{v}).$$

From this equation, we get that the energy-loss rate per unit of time is represented as the product of a kinematic factor $\frac{\omega}{\boldsymbol{q}^2}$ and a loss function $L(\boldsymbol{q},\omega)$:

$$L(\boldsymbol{q},\omega) = -\mathrm{Im}\left[\frac{1}{\epsilon_M(\boldsymbol{q},\omega)}\right] = \frac{\epsilon_2(\boldsymbol{q},\omega)}{\epsilon_1^2(\boldsymbol{q},\omega) + \epsilon_2^2(\boldsymbol{q},\omega)}. \tag{2.2}$$

The loss function equation tells us which conditions ϵ_1 and ϵ_2 must fulfil for strong visible structures to form within an EELS. Equation (2.2) indicates that the most pronounced spectral features in EELS can be observed under two conditions:

1. ϵ_2 exhibits a peak corresponding to some interband transition,

2. ϵ_2 is small and $\epsilon_1 = 0$.

The first case belongs to the same category as what we have already discussed in the case of absorption spectroscopy ($\alpha(\omega) \sim \epsilon_2(\omega)$). The difference to absorption spectroscopy, however, is the denominator $\epsilon_1^2(\boldsymbol{q},\omega) + \epsilon_2^2(\boldsymbol{q},\omega)$ of the loss function. This can be seen by assuming that we are in the frequency range of the excitons, in which the imaginary part of the dielectric function would be very large ($\epsilon_2 \gg 1$), and since the denominator increases quadratically positively, we can estimate the effect that we perceive in the region of the excitons to be $L(\boldsymbol{q},\omega) < \frac{1}{\epsilon_2(\boldsymbol{q},\omega)} < 1$. This shows us that excitonic effects will generally play a less important role in EELS than in absorption spectra.

In the second case where $\epsilon_2(\boldsymbol{q},\omega)$ is small, i.e. we are far away from the optical bandgap. We can expect a substantial increase in the loss function when the real part of the dielectric function becomes small or zero $\epsilon_1(\boldsymbol{q},\omega) = 0$. Since the disappearance of the real part is not a trivial phenomenon, an interesting question arises: *What does the zero crossing of the real part of the dielectric function mean?*

To keep this discussion as simple as possible, we consider only the case $\boldsymbol{q} \to \boldsymbol{0}$ and make the simple assumption that we consider a free electron gas, which is described by non-relativistic electrons with mass $m = 1$ and charge $e = 1$ in Hartree units, the electrons experience a collision time of τ in the material and the external electric field, which perturbates the electrons, can be described by $\boldsymbol{E} = \boldsymbol{E}_0 \exp(-i\omega t)$. In such a case, we can use the damped harmonic oscillator without restoring force as a simple model

$$\frac{d\boldsymbol{v}}{dt} + \frac{1}{\tau}\boldsymbol{v} = \boldsymbol{E}_0 \exp(-i\omega t).$$

After a short calculation [42] together with the definition for the dielectric function (Eq. (2.1)) and the definition of the frequency $\omega_p = \sqrt{4\pi n}$, where n is the electron density, we get the expression for the dielectric function

$$\epsilon(\omega) = 1 - \frac{\omega_p^2\tau^2}{1 + \omega^2\tau^2} + i\frac{\omega_p^2\tau^2}{\omega(1 + \omega^2\tau^2)}. \tag{2.3}$$

If we consider only the real part of the dielectric function and further assume that the scattering time is very long $1 \ll \omega_p^2 \tau^2$ (metal limit), we obtain the equation

$$\epsilon_1(\omega) = 1 - \frac{\omega_p}{\omega}.$$

From this highly simplified equation, we can now read what it means when the real part of the dielectric function is zero, namely that the system oscillates at a frequency ω_p. Since this frequency depends on the density of the electrons, it describes a collective oscillation of electrons, the plasmons. For this reason, the frequency ω_p is called the plasma frequency.

However, since the model that was used only allows a very simplified view of the world, we should know how such effects as plasmons occur in a real spectrum. To do this, we show in figure 2.2 the EELS for Si.
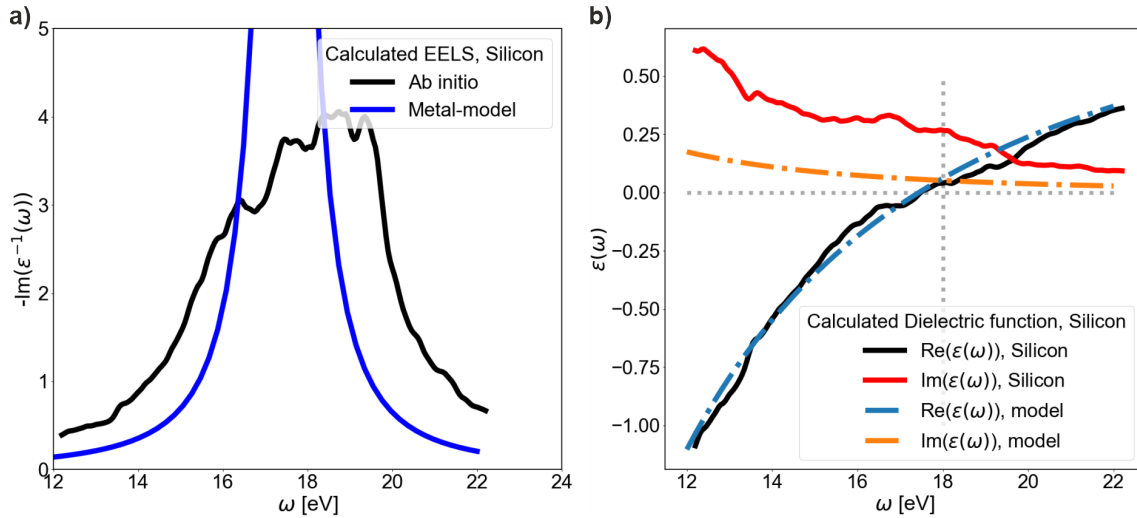
Figure 2.2: Electron energy-loss spectrum for silicon, represented by the frequency-dependent loss function $L(\omega) = -\mathrm{Im}(\epsilon^{-1}(\omega))$ (Eq. (2.2)). Shown in **a)** the calculated loss function for silicon (black) and the model loss function (blue)(Eq. (2.3), $\tau = 1$, $\omega_p = 17.46$ eV, taken from the calculated ϵ_1) and **b)** the real (black/blue) and imaginary (red/orange) part of the frequency dependent calculated/model dielectric function. When the real part of the dielectric function tends to zero (plasma frequency), the loss function increases, reaches its maximum value between the zero point of the real part and the cross point of the imaginary and real part and decreases rapidly after the crossing point. The model would predict that the highest peak should be at the zero crossing of the real part. Calculations were made neglecting the electron-hole interaction on an 8k grid (2048 k-points) and using a Kohn-Sham band structure with the parameters given in Appendix A and $\boldsymbol{q} = (0.0, 0.0, 0.0)$.

As we have discussed above, the loss function increases as the imaginary part becomes small, and the real part of the dielectric function tends towards zero. We observe this behavior in both the ab inito calculation (Fig. 2.2 a), black) and the model (Fig. 2.2 a), blue) for Si.

It is interesting to note that the real part of the dielectric function for the model reflects the ab inito calculation very well. This empirical finding validates the description of plasmons with the simple metal model we mentioned above, so that we can read the plasma frequency directly from the graph ($\epsilon_1(\omega_p) = 0$): $\omega_p = 17.46$ eV.

The imaginary part, on the other hand, is poorly described, the model underestimates the magnitude by a factor of two. This poor description of the imaginary part leads to the fact that the model EELS is also poorly described. We see this immediately when we take a look at the loss function $L(\omega) \approx \frac{1}{\epsilon_2(\omega)}$ in the plasma frequency range: If the imaginary part is described overly strongly or weakly, this will inevitably lead to an under- or overestimation of the EELS. Furthermore, we see that the ab inito calculation has much more structure than the simple metal model can reflect.

This should make it clear that models are an important support for analysis and physical understanding. However, they only contain part of the truth. This means that the results can only be trusted to a limited extent and caution is advised without a comparison with experiments or ab inito calculations.

Finally, we would like to make a brief remark about the calculations of the EELS compared to the absorption spectrum, even though we have not discussed how to do the calculation so far. For a fully converged absorption spectrum close to the optical band edge, 18 conduction bands are sufficient; for the EELS spectrum, 52 conduction bands are necessary (Appendix A). The reason for the large number of conduction bands for the EELS is, firstly, the energy range we are interested in and, secondly, the real part of the dielectric function, which requires a much larger number of bands to converge.

2.2 Intermediate summary

This section has given us an insight into the framework, optical spectroscopy, in which we will work. We have looked at both absorption and electron energy-loss spectroscopy.

The absorption spectroscopy has shown us that classical and simple quantum mechanical descriptions via Fermi's golden rule in the IP picture is insufficient to account for the crucial effects. It was necessary to go beyond the simple picture of non-interacting particles and consider the interaction between electrons and holes to describe the spectra accurately.

Furthermore, we have seen that excitonic effects can affect the spectrum differently. In particular, slight shifts and weaker peaks near the optical band edge appear when the excitonic effects are weak. On the other hand, when the excitonic effects are strong, as we have seen in LiF, a substantial change regarding the non-interacting spectrum can occur, as well as excitonic peaks within the optical bandgap, so-called strongly bound excitons.

In the case of the EELS, we have discussed that excitonic effects close to the optical bandgap have less influence on the spectrum, instead a strong, broad peak appeared within the band continuum of Si. This peak was caused by collective excitations of the electron density, so-called plasmons.

Both effects have in common that the dielectric function must be known for their theoretical description, which we have so far come to know as a macroscopic object. However, in order to be able to describe the effects that we observed, we had to use the language of quantum mechanics and not classical electron dynamics. For this reason, it must be clear that we must immerse ourselves in the microscopic world to describe the effects at the theoretical level accurately. Therefore, we should ask ourselves the question:

How can we derive the microscopic dielectric susceptibility from quantum mechanics?

Chapter 3

The microscopic formalism of the dielectric susceptibility

> "To know what one knows and to know what one does, that is knowledge."
>
> Konfuzius

The previous section showed us that we had to use the language of quantum mechanics with interactions to describe our observations in absorption and electron energy-loss spectroscopy, the excitonic and plasmonic effects.

In order to accurately embed the underlying effects of absorption and electron energy-loss spectroscopy into the theory, our goal in the following will be to describe the material and thus the susceptibility χ and the dielectric function ϵ microscopically.

This aim raises the question: *With which equation do we start on our way?* Since we want to describe the material quantum mechanically under a time-dependent perturbation, we must use the full many-particle Hamiltonian in the time-dependent Schroedinger equation for an exact description:

$$i\partial_t \Psi(\bar{r}, \bar{R}, t) = \bar{H}(\bar{r}, \bar{R}, t)\Psi(\bar{r}, \bar{R}, t), \ \bar{r} = \{r_i\}, \ \bar{R} = \{R_I\}.$$

Without going into the exact structure of the Hamiltonian, we can see that it depends on time t and the positions of the electrons ($\bar{r}$) and ions ($\bar{R}$). However, since a cubic centimeter of silicon, for example, already has 10^{22} ions and $1.4 \cdot 10^{23}$ electrons, it is impossible to set up the Hamiltonian or to solve for it. For this reason, it already advisable to start with the first approximation.

The idea of the first approximation will be based on the question: *Does one need all the information of the full many-particle Hamiltonian to describe either the collective interaction between electrons or the interaction between electrons and holes?*

The answer is no, since according to the brilliant idea of Born and Oppenheimer [43], two separate timescales exist inside the system, one for the faster moving electrons and one for the slower nuclei. The reason for this is that the proton mass exceeds the electron mass by 1836, which makes the nuclei of the atoms significantly slower than their valence electrons. This assumption allows the physics of electrons and nuclei to be almost decoupled from each other, the nuclei are only

considered as a static construct inside the Hamiltonian of the electrons

$$i\partial_t \Psi_{\bar{\boldsymbol{R}}}(\bar{\boldsymbol{r}}, t) = H^{eMB}(\bar{\boldsymbol{r}}, \bar{\boldsymbol{R}}, t)\Psi_{\bar{\boldsymbol{R}}}(\bar{\boldsymbol{r}}, t).$$

To get an impression of the Hamiltonian, we write it out in a component-wise representation

$$H^{eMB}(\bar{\boldsymbol{r}}, \bar{\boldsymbol{R}}, t) = \sum_{i=1}^{N}\left(-\frac{1}{2}\partial_{\boldsymbol{r}_i}^2 + v_{ext}(\boldsymbol{r}_i, \bar{\boldsymbol{R}}, t)\right) + \frac{1}{2}\sum_{i\neq j}^{N} v_c(|\boldsymbol{r}_i - \boldsymbol{r}_j|), \qquad (3.1)$$

where v_c is the Coulomb-interaction between the electrons and v_{ext} contains the external time-dependent perturbation, as well as the interaction between the fixed nuclei and the moving electrons.

Now that we have described the Hamiltonian of the system, we should ask ourselves: *How do we get from the Schroedinger equation to dielectric susceptibility?*

In order to find a mathematical answer, we should be clear about what we are observing. For this purpose, we will limit ourselves to plasmonic effects, as the picture is straightforward. Plasmons are a collective wave of interacting electrons, similar to a water wave in a lake. This means we do not observe individual electrons directly but a collection of them, i.e. a density $n = \langle\hat{n}\rangle$. If we now disturb such a density of electrons at the location $\boldsymbol{r}'$ at the time t' with a weak laser pulse, then we can observe at a different point in time and space $(\boldsymbol{r}, t)$ how this disturbance spreads. This means that the system responds to the disturbance, and since we look at densities, we observe a density-density response, also called dielectric susceptibility.

3.1 Density-density response function

For the sake of simplicity, we will neglect the position variables in the following and follow the approach outlined in [44] by G. Czycholl.

The way now to the density-density response function is through the measure of the electron density $n(t) = \langle\hat{n}\rangle(t)$. Since our system includes a temporal disturbance, we evaluate this quantity in the interaction picture. For this reason, it is central to define our system through the Hamiltonian (Eq. (3.1))

$$\hat{H}^{eMB}(t) = \hat{H} + \hat{H}_1(t) =: \hat{H} + \hat{n}v_{per}(t),$$

in the following with a weak perturbation $\hat{H}_1(t) = \hat{n}v_{per}(t)$, which couples to the density, and a unperturbed part $\hat{H}$.

However, since our interest lies in a time-dependent observable, the density

$$\langle\hat{n}\rangle(t) = Tr(\hat{n}\,n(t)) = Tr(\hat{n}_I(t)n_I(t)), \qquad (3.2)$$

we need an expression for the time-dependent density operator $\hat{n}_I(t)$ in order to solve the equation. For this purpose we can use the Neumann equation, which represents an equation of motion (EoM) for the density operator:

$$i\partial_t \hat{n}(t) = [\hat{H}^{eMB}(t), \hat{n}(t)].$$

In the interaction picture the EoM for the density operator becomes:

$$i\partial_t \hat{n}_I(t) = [\hat{H}_1(t), \hat{n}_I(t)] = [\hat{n}_I(t), \hat{n}_I(t)]v_{per}(t).$$

The obtained equation for the density operator represents a linear differential equation, and its solution is given by:

$$\hat{n}_I(t) = \hat{n}_0 - i \int_{-\infty}^{t} dt' \, [\hat{n}_I(t'), \hat{n}_I(t)] v_{per}(t').$$

The problem with the EoM for the density operator, even though it is exact, is that to get $\hat{n}_I(t)$, we must already know it. Such equations are usually solved self-consistently. Since we only refer to weak fields ($\hat{n}_0 \gg i \int_{-\infty}^{t} dt' \, [\hat{n}_I(t'), \hat{n}_0] v_{per}(t')$), we can turn our focus on the linear part of the equation and therefore break off the iterative process after the first order:

$$\hat{n}_I(t) \approx \hat{n}_0 - i \int_{-\infty}^{t} dt' \, [\hat{n}_I(t'), \hat{n}_0] v_{per}(t').$$

Substituting the linear equation into the expression for the observable (Eq. (3.2)) and defining the retarded density-density response function

$$\chi(t, t') := -i \langle [\hat{n}_I(t), \hat{n}_I(t')] \rangle_{n_0} \theta(t - t'), \tag{3.3}$$

we arrive at the following expression for the density as observable:

$$\langle \hat{n} \rangle_{n_0}(t) = \langle \hat{n} \rangle_{n_0} - i \int_{-\infty}^{t} dt' \, Tr(\hat{n}_I(t)[\hat{n}_I(t'), \hat{n}_0]) v_{per}(t') = \langle \hat{n} \rangle_{n_0} + \int_{-\infty}^{\infty} dt' \, \chi(t, t') v_{per}(t'). \tag{3.4}$$

The term "retarded" for χ signifies that the time propagation is forward, which manifests itself in the θ-function, meaning that there is a perturbation at time t' followed by a response of the system at a later time t ($t > t'$).

Having reached this point, we have now found a definition for the density-density response function (Eq. (3.3)) that describes exactly what we previously assumed heuristically, a coupling between two densities at different points. In the following, we will elaborate this equation a little further in order to understand it better and to obtain useful equations.

The first useful equation can be derived from equation (3.4), if one is interested in the change of density with the perturbing field:

$$\delta \langle \hat{n} \rangle_{n_0}(t) = \int_{-\infty}^{\infty} dt' \, \chi(t, t') \delta v_{per}(t') \Rightarrow \frac{\delta \langle \hat{n} \rangle_{n_0}(t)}{\delta v_{per}(t')} = \frac{\delta n(t)}{\delta v_{per}(t')} =: \chi(t, t'). \tag{3.5}$$

This equation shows that the density-density response function describes nothing other than the change in density with the perturbation.

The next step is understanding that the response function depends only on the time difference $\chi(t, t') = \chi(t - t')$. This can be shown by changing to the interaction picture or using the wave example in the water. In this analogy, if we fix the measurement conditions, it is irrelevant at what exact time the disturbance (t') in the water takes place; only the difference between the time of the disturbance, i.e. the emission of the wave and the time of the measurement (t), can play a role. For instance, it does not matter whether we experiment early or late in the evening if the conditions (external disturbances) are the same.

Since experiments are carried out with lasers, lamps, etc. of a certain frequency, i.e. photons of a certain energy, it makes sense to subject the response function to a Fourier transformation in the frequency space to carry this out. For this purpose we bring the completeness relation

$1 = \sum_\lambda |\lambda\rangle\langle\lambda|$ into equation (3.3) and obtain after Fourier transformation [38]:

$$\chi(\omega) = \lim_{\Gamma\to 0} \sum_\lambda \left[\frac{o_\lambda o_\lambda^*}{(\omega + i\Gamma) - (E_\lambda - E_0)} - \frac{o_\lambda^* o_\lambda}{(\omega + i\Gamma) + (E_\lambda - E_0)} \right],$$

$$o_\lambda = \langle N|\hat{n}|\lambda\rangle.$$

(3.6)

In the expression, $|N\rangle$ represents the N-particle ground state of the system ($\hat{H}$), with the corresponding ground state energy E_0. On the other hand, $|\lambda\rangle$ originates from the completeness relation and represents the excited state corresponding to the energy E_λ of the Hamiltonian $\hat{H}$. Finally, to guarantee convergence of the Fourier transformation [45], we shifted the beam-frequence ω by a small broadening denoted as Γ.

Having the structure of χ in frequency space leads to the question: *What is the meaning of the two terms representing the response function?* Since we are now in frequency space, the poles (denominators) correspond to the excitation of particles from an energy E_0 to an energy E_λ, without changing the number of particles (neutral excitations). In this picture the first term corresponds to a resonant transition ($-(E_\lambda - E_0)$), while the second term represents an anti-resonant transition ($+(E_\lambda - E_0)$).

Having derived the dielectric susceptibility at a microscopic level using linear response theory, the space-dependency of the response function is still missing. However, introducing the dependence is not a complicated intervention; one can work with the local quantities ($\hat{n}(\boldsymbol{r}, t), v_{per}(\boldsymbol{r}, t)$) from the beginning, which leads to the position- and frequency-dependent density-density response function:

$$\chi(\boldsymbol{r}, \boldsymbol{r}', \omega) = F_T^t \left(\frac{\delta n(\boldsymbol{r}, t)}{\delta v_{per}(\boldsymbol{r}', t')} \right) = \lim_{\Gamma\to 0} \sum_\lambda \left[\frac{o_\lambda(\boldsymbol{r})o_\lambda^*(\boldsymbol{r}')}{(\omega + i\Gamma) - (E_\lambda - E_0)} - \frac{o_\lambda^*(\boldsymbol{r})o_\lambda(\boldsymbol{r}')}{(\omega + i\Gamma) + (E_\lambda - E_0)} \right].$$

(3.7)

It is important to note that, unlike in the time domain, there is no reason why the measurement in the space domain should depend only on distance and not directly on $\boldsymbol{r}$ and $\boldsymbol{r}'$. Here, too, the wave example can give us an intuition. Let us imagine a perfect lake which looks the same in all directions and is infinitely extended and deep. We would call this lake homogeneous. In this case, it would not matter where we carry out our measurement $\boldsymbol{r}$ or where the disturbance happens $\boldsymbol{r}'$. Only the distance $|\boldsymbol{r} - \boldsymbol{r}'|$ between the origin of the wave $\boldsymbol{r}'$ and the measurement $\boldsymbol{r}$ would be decisive because we would find the same conditions everywhere. However, a natural lake is not perfect; it has differences in depth and perhaps even other boundary conditions due to protruding stones, so it may be that, depending on the place of origin of the wave and the measurement, all these irregularities play a role. It is the same in an real material. For example, the distribution of electrons has irregularities depending on the local interaction due to the underlying structure of the crystal lattice, making a difference at which position we disturb the system and measure. For this reason, we cannot generally expect a simple distance dependency.

Now that we know the dielectric susceptibility in the form of a formula on the microscopic level (Eq. (3.7)), we should ask ourselves: *How can we use this equation in practice?* This question arises from two observations: Firstly, the density-density response function depends on the oscillator strengths o_λ and secondly, on the eigenenergies of the many-body electron Hamiltonian (Eq. (3.1)) without external perturbation. The problem, however, is that we know neither. The solution to this problem is deferred to the next but one section (Ch. 3.3), in which we will learn a way around this

problem. However, before we deal with this, we should first look at how we get from the microscopic dielectric susceptibility to the macroscopic function, which is measured in experiments.

3.2   Macroscopic dielectric function

Since the decisive parameter for our optical experiments is the dielectric function, it is first necessary to establish a link between microscopic susceptibility and microscopic dielectric function. The next step will be establishing a link between microscopic and macroscopic dielectric function. To make the task as simple as possible, we introduce a notation that combines space ($\boldsymbol{r}$), time (t), and spin (σ) variables in one variable $j \in \mathbb{N}$, defined as $j = (\boldsymbol{r}_j, t_j, \sigma_j)$. The notation implies that the sum over spin and integration over space and time can be represented as $\sum_{\sigma_j} \int d\boldsymbol{r}_j \int dt_j =: \int d_j$. The reason for the notation is to focus more on physics and mathematics than on indices and symbols.

To derive an equation for the microscopic dielectric function, we use classical electrodynamics in which the total potential v_{tot} of a perturbed system is given by the sum of the external perturbative potential v_{per} and the induced potential v_{ind} in the material [46]:

$$\delta v_{tot}(1) = \delta v_{per}(1) + \delta v_{ind}(1). \tag{3.8}$$

Since the induced density creates an induced potential. Its classical part is the induced Hartree potential $v_H(1) := \int d_2 \, v_c(1,2)n(2)$, which is the quantity we need here on the classical level of our description $v_{ind}(1) \to v_H(1)$. With this relation, we can further rewrite equation (3.8) as follows:

$$\delta v_{tot}(1) = \delta v_{per}(1) + \int d_2 \, v_c(1,2)\delta n(2) = \delta v_{per}(1) + \delta v_H(1). \tag{3.9}$$

In order to bring the dielectric susceptibility into the equation, we can utilize the relationship $\delta n = \chi \delta v_{per}$ introduced in equation (3.7) to rewrite equation (3.9) as:

$$\delta v_{tot}(1) = \delta v_{per}(1) + \int d_2 d_3 \, v_c(1,2)\chi(2,3)\delta v_{per}(3) = \int d_3 \left(\delta(1,3) + \int d_2 \, v_c(1,2)\chi(2,3) \right) \delta v_{per}(3). \tag{3.10}$$

Finally, we define the inverse microscopic dielectric function

$$\epsilon^{-1}(1,2) := \delta(1,2) + \int d_2 \, v_c(1,2)\chi(2,3). \tag{3.11}$$

The equation we obtain for the dielectric function provides us with the link between the dielectric susceptibility and the dielectric function at the microscopic level. As already in the macroscopic case (Eq. (2.1)), the dielectric function encompasses two cases: the response of the environment without material inside (represented by the delta function), which leaves the outer field unchanged, and the response of the material to an external perturbation, which changes the outer field. Now that we know the link between the microscopic dielectric function and the dielectric susceptibility, the question is: *How do we get to the macroscopic dielectric function?*

To answer this question, it is worth answering another question: *What is the connection between the microscopic total potential and the macroscopic one?* This question is much easier to answer, and since the total potential depends on the dielectric function (Eq. (3.10)), the answer leads directly to the answer to our actual question.

In order to answer the question about the link between microscopic and macroscopic potential, let us briefly conduct a short thought experiment: Let us imagine that we are an electron sitting on the surface of a material, and we see a potential landscape of hills and valleys. Since we are on the material's surface, we perceive all the irregularities of the potential landscape. Now, we move far away from the material until it is just a tiny, unresolved area on the horizon. We still feel the material's potential in this case, but it is no longer finely resolved but macroscopic. Nevertheless, since all the hills and valleys contribute to this potential, it must be an averaging over the landscape, i.e. an integration, which we feel.

After this thought experiment, we will package the words in equations. For this, we consider the specific case of a perfect crystal with a periodic potential and transform the essential quantities (Eq. (3.10) and Eq. (3.11)) in the Fourier space. In order to accomplish this, we remember that all quantities in the dielectric function depend only on the time difference. Therefore, the Fourier transformation in frequency and reciprocal space domain for the dielectric function leads to

$$\epsilon^{-1}_{\boldsymbol{G},\boldsymbol{G}'}(\boldsymbol{q},\omega) = \delta_{\boldsymbol{G},\boldsymbol{G}'} + v_c(\boldsymbol{q}+\boldsymbol{G})\chi_{\boldsymbol{G},\boldsymbol{G}'}(\boldsymbol{q},\omega) \tag{3.12}$$

and similar to the total potential in reciprocal space

$$\delta v_{tot}(\boldsymbol{q}+\boldsymbol{G},\omega) = \sum_{\boldsymbol{G}'} \epsilon^{-1}_{\boldsymbol{G},\boldsymbol{G}'}(\boldsymbol{q},\omega)\delta v_{per}(\boldsymbol{q}+\boldsymbol{G}',\omega). \tag{3.13}$$

In both equations, $\boldsymbol{q}$ represents the reciprocal vector of the first Brillouin zone and $\boldsymbol{G}$ is the lattice vector in reciprocal space.

Now that we have these quantities, we can look at how we can do spatial averaging. To this end, we follow [38]. The idea is to find out what spatial averaging means in reciprocal space. For this purpose, we start with the usual assumption that lattice-related microscopic quantities are lattice periodic, and therefore, we can perform a Fourier expansion of a potential as:

$$V(\boldsymbol{r},\omega) = \sum_{\boldsymbol{q},\boldsymbol{G}} V(\boldsymbol{q}+\boldsymbol{G},\omega)\exp(i(\boldsymbol{q}+\boldsymbol{G})\boldsymbol{r}) = \sum_{\boldsymbol{q}}\exp(i\boldsymbol{q}\boldsymbol{r})\sum_{\boldsymbol{G}} V(\boldsymbol{q}+\boldsymbol{G},\omega)\exp(i\boldsymbol{G}\boldsymbol{R}) =$$

$$=: \sum_{\boldsymbol{q}}\exp(i\boldsymbol{q}\boldsymbol{r})V(\boldsymbol{q},\boldsymbol{r},\omega),$$

where $V(\boldsymbol{q},\boldsymbol{r},\omega)$ is periodic concerning the underlying Bravais lattice $V(\boldsymbol{q},\boldsymbol{r},\omega) = V(\boldsymbol{q},\boldsymbol{r}+\boldsymbol{R},\omega)$. Next, we want to perform a macroscopic average

$$V_M(\boldsymbol{q},\omega) = \frac{1}{V_0}\int d\boldsymbol{r}\, V(\boldsymbol{q},\boldsymbol{r},\omega),$$

where V_0 is the volume of the unit cell. If we work out this equation further by evaluating explicitly the Fourier transformation

$$V_M(\boldsymbol{q},\omega) = \frac{1}{V_0}\int d\boldsymbol{r}\, V(\boldsymbol{q},\boldsymbol{r},\omega) = \frac{1}{V_0}\sum_{\boldsymbol{G}} V(\boldsymbol{q}+\boldsymbol{G},\omega)\int d\boldsymbol{r}\,\exp(i\boldsymbol{G}\boldsymbol{r}) = \sum_{\boldsymbol{G}} V(\boldsymbol{q}+\boldsymbol{G},\omega)\delta_{\boldsymbol{G},\boldsymbol{0}} = V(\boldsymbol{q},\omega),$$

we obtain that the link between microscopic and macroscopic crystal periodic quantities is given by setting in reciprocal space $\boldsymbol{G} = \boldsymbol{0}$. This knowledge leads us to the equation for the macroscopic total potential (Eq. (3.13)) given as:

$$\delta v_{tot}^M(\boldsymbol{q},\omega) = \sum_{\boldsymbol{G}'} \epsilon^{-1}_{\boldsymbol{0}\boldsymbol{G}'}(\boldsymbol{q},\omega)\delta v_{per}(\boldsymbol{q}+\boldsymbol{G}',\omega).$$

Since the external field δv_{per} is, in the case of spectroscopy, a macroscopic field $\delta v_{per}(\boldsymbol{q} + \boldsymbol{G}', \omega) = \delta v_{per}(\boldsymbol{q}, \omega)$, the only component, which is different from zero, is the component $\boldsymbol{G} = \boldsymbol{0}$. For this reason, the final equation for the macroscopic total potential is:

$$\delta v_{tot}^M(\boldsymbol{q}, \omega) = \epsilon_M^{-1}(\boldsymbol{q}, \omega)\delta v_{per}^M(\boldsymbol{q}, \omega) = \epsilon_{\boldsymbol{0},\boldsymbol{0}}^{-1}(\boldsymbol{q}, \omega)\delta v_{per}^M(\boldsymbol{q}, \omega).$$

The obtained equation allows us to read directly what the connection between microscopic and macroscopic dielectric function is

$$\epsilon_M(\boldsymbol{q}, \omega) = \frac{1}{\epsilon_{\boldsymbol{0},\boldsymbol{0}}^{-1}(\boldsymbol{q}, \omega)} \neq \epsilon_{\boldsymbol{0},\boldsymbol{0}}(\boldsymbol{q}, \omega). \tag{3.14}$$

As we can see, the link between the macroscopic and microscopic dielectric functions is not a simple average like it is for the potential. Instead, it involves an inversion. This opens up an interesting question: *What is the physical reason behind this behavior of the dielectric function?* To answer this question, we need an equation in which we can separate the $\boldsymbol{G} = \boldsymbol{G}' = \boldsymbol{0}$ component from the rest of the dielectric function. For this reason, we will directly invert equation (3.13) to obtain:

$$\delta v_{per}(\boldsymbol{q}, \omega) = \sum_{\boldsymbol{G}'} \epsilon_{\boldsymbol{0},\boldsymbol{G}'}(\boldsymbol{q}, \omega)\delta v_{tot}(\boldsymbol{q} + \boldsymbol{G}', \omega).$$

If we now split off the proportion $\boldsymbol{G} = \boldsymbol{G}' = \boldsymbol{0}$, then we get

$$\delta v_{per}(\boldsymbol{q}, \omega) = \epsilon_{\boldsymbol{0},\boldsymbol{0}}(\boldsymbol{q}, \omega)\delta v_{tot}(\boldsymbol{q}, \omega) + \sum_{\boldsymbol{G}' \neq \boldsymbol{0}} \epsilon_{\boldsymbol{0},\boldsymbol{G}'}(\boldsymbol{q}, \omega)\delta v_{tot}(\boldsymbol{q} + \boldsymbol{G}', \omega).$$

The resulting equation shows us that assuming $\epsilon_M(\boldsymbol{q}, \omega) = \epsilon_{\boldsymbol{0},\boldsymbol{0}}(\boldsymbol{q}, \omega)$ implies neglecting the off-diagonal elements of the microscopic total potential. However, this physical means that since the perturbation is, by its definition, a macroscopic quantity, one would neglect information from the induced potential and, therefore, from the material, referred to as crystal local-field effects [47,48]. Therefore, the only case in which $\epsilon_M(\boldsymbol{q}, \omega) = \epsilon_{\boldsymbol{0},\boldsymbol{0}}(\boldsymbol{q}, \omega)$ holds exactly is when $\epsilon_{\boldsymbol{G},\boldsymbol{G}'}(\boldsymbol{q}, \omega) = \epsilon_{\boldsymbol{G},\boldsymbol{G}}(\boldsymbol{q}, \omega)$ is diagonal or in real-space $\epsilon(\boldsymbol{r}, \boldsymbol{r}', \omega) = \epsilon(|\boldsymbol{r} - \boldsymbol{r}'|, \omega)$, which is equivalent to $\chi(\boldsymbol{r}, \boldsymbol{r}', \omega) = \chi(|\boldsymbol{r} - \boldsymbol{r}'|, \omega)$. This condition, as we already know, requires a perfectly homogeneous system. Materials behaving similarly to the homogeneous electron gas, such as metals or some small-gap semiconductors, would approximately satisfy this condition.

With the obtained link between dielectric function and susceptibility on a microscopic level (Eq. (3.12)) and the link between microscopic and macroscopic dielectric (Eq. (3.14)), in principle, we now have everything together to do theoretical spectroscopy. We know the crucial quantity, namely the dielectric susceptibility (Eq. (3.7)) in the microscopic world, and we know how to obtain the macroscopic dielectric function (Eq. (3.14)), which we need in order to reproduce experiments. Unfortunately, we have a not-so-small problem. In order to calculate the dielectric susceptibility, we need to know the ground state energies and the excited energies of the total electron many-body problem (Eq. (3.1)), as well as the corresponding wave functions. However, this requirement means we have to solve the whole electron many-body problem, which is impossible.

To get around this problem, it is necessary to take a different approach. Since the problem is that we cannot calculate the necessary quantities for χ, we need a way that gives us the information of the Hamiltonian without having to diagonalize it. Such a path is the introduction of Green's functions,

which can be described mathematically as the inverse of an operator $L(1)G(1,2) = \delta(1,2)$, for example, a Hamiltonian $L = H$, and thus contain the total amount of information from the Hamiltonian.

3.3 Green's function formalism

The primary objective of this section is to show one possible approach to address the challenge posed by the lack of knowledge regarding the eigenstates and energies of the electron many-body Hamiltonian (Eq. (3.1)) while investigating the density-density response function (Eq. (3.7)).

A formally exact method to rewrite this demanding task is the method of Green's functions, which contains the complete information of the system and contains therefore information such as ground state energy and the excitation energies.

To enter the formalism of Green's functions, we initiate our exploration from the broader context of the correlation function. The advantage of using correlation functions, is that the description is purely mathematical and thus universally valid. For us, this means that we have an important tool to derive many quantities, such as the density-density correlation function or the Green's functions. With this tool, we will venture into the question: *How is it possible to obtain the density-density response function using Green's function formalism?*

3.3.1 Correlation functions

Before we get into mathematical formulations, it is worth getting a feel for the word correlation. We look at the density-density response function (Eq. (3.3)) to do this. We have previously physically interpreted this function as follows: We perturbed the system at a point in space and time $(\boldsymbol{r}', t')$ and measured at a later time t at a different location $\boldsymbol{r}$. Since the information was transferred from $(\boldsymbol{r}', t')$ to $(\boldsymbol{r}, t)$, these two places in space and time must be in mutual exchange, so they are correlated. However, if we were to place a second material somewhere next to the disturbed one and not disturb it but measure it, then we would not find any correlation between the disturbed and undisturbed systems. The two materials would, therefore, be uncorrelated.

In order to put this simple example into a mathematical general concept, we will follow [46]. From a mathematical point of view, we define quantities to be correlated when the expectation value of their product does not equal the product of their expectation values (uncorrelated), expressed as:

$$\langle \hat{A}\hat{B} \rangle \neq \langle \hat{A} \rangle \langle \hat{B} \rangle.$$

Elaborated further, the correlation function in the context of many-body physics can be written as:

$$C_{AB}(1,2) := \langle \hat{A}_H(1)\hat{B}_H(2) \rangle = \sum_\lambda w_\lambda \langle \lambda | \hat{A}_H(1)\hat{B}_H(2) | \lambda \rangle,$$

where $\hat{A}_H(1)$, $\hat{B}_H(2)$ are operators in the Heisenberg picture, $|\lambda\rangle$ represents a many-body state, and $w_\lambda = \frac{\exp(-\beta(E_\lambda - \mu_T N_\lambda))}{\sum_\lambda \exp(-\beta(E_\lambda - \mu_T N_\lambda))}$ denotes the weight of state $|\lambda\rangle$ in the grand-canonical ensemble at the equilibrium with the inverse temperture $\beta = 1/(k_B \mathrm{T})$ and the chemical potential μ_T. As we have seen in the case of the density-density response function (Eq. (3.3)), we will later focus on

retarded (forward traveling in time, $t_1 > t_2$) correlation functions. For this purpose, we define $\tilde{C}_{AB}(1,2) = \pm C_{BA}(1,2)$ for bosons/fermions, which allows us to obtain:

1. The **Retarded Correlation function**:

$$C_{AB}^R(1,2) := \theta(t_1 - t_2)\langle[\hat{A}(1), \hat{B}(2)]\rangle = \theta(t_1 - t_2)(C_{AB}(1,2) - \tilde{C}(1,2)). \qquad (3.15)$$

2. The **Time-ordered Correlation function**:

$$C_{AB}^T(1,2) := -\langle T[\hat{A}(1), \hat{B}(2)]\rangle = \theta(t_1 - t_2)C_{AB}(1,2) + \theta(t_2 - t_1)\tilde{C}(1,2)$$
$$= C_{AB}^R(1,2) + \tilde{C}_{AB}(1,2). \qquad (3.16)$$

Since we will work mainly in the frequency domain, we want to express the correlation function in frequency space. To do so, we assume a static Hamiltonian, we use the completeness relation $1 = \sum_{\lambda'}|\lambda'\rangle\langle\lambda'|$ and the transformation of pictures $\hat{O}_H(t) = \exp(i\hat{H}(t - t_0))\hat{O}\exp(-i\hat{H}(t - t_0))$, which allows us to express the correlation function as:

$$C_{AB}(\boldsymbol{r}_1, \boldsymbol{r}_2, t_1 - t_2) = \sum_{\lambda\lambda'} w_\lambda \exp(i(E_\lambda - E_{\lambda'})(t_1 - t_2))A_{\lambda\lambda'}(\boldsymbol{r}_1)B_{\lambda'\lambda}(\boldsymbol{r}_2).$$

From this point we can perform a Fourier transformation for the correlation function

$$C_{AB}(\boldsymbol{r}_1, \boldsymbol{r}_2, \omega) = \int_{-\infty}^{\infty} d\tau \sum_{\lambda\lambda'} w_\lambda A_{\lambda\lambda'}(\boldsymbol{r}_1)B_{\lambda'\lambda}(\boldsymbol{r}_2)\exp(i(E_\lambda - E_{\lambda'})\tau)\exp(i\omega\tau)$$
$$= 2\pi \sum_{\lambda\lambda'} w_\lambda A_{\lambda\lambda'}(\boldsymbol{r}_1)B_{\lambda'\lambda}(\boldsymbol{r}_2)\delta(E_\lambda - E_{\lambda'} + \omega).$$

To account for the theta function found within the time-ordered or retarded correlation function, we use the expression $\int d\tau\,\theta(\pm\tau)\exp(i\omega\tau - \Gamma\tau) = \pm\frac{i}{\omega\pm i\Gamma}$ [45] in order to perform the Fourier transformation for the retarded correlation function

$$C_{AB}^R(\boldsymbol{r}_1, \boldsymbol{r}_2, \omega) = \int d\tau\,\theta(\tau)(C_{AB}(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau) - \tilde{C}(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau))\exp(i\omega\tau) =: 2\pi \int d\tau\,\theta(\tau)A_{AB}(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)\exp(i\omega\tau)$$
$$= \int d\tau \int d\omega'\,\theta(\tau)A_{AB}(\boldsymbol{r}_1, \boldsymbol{r}_2, \omega')\exp(-i(\omega' - \omega)\tau) = \lim_{\Gamma\to 0} i \int d\omega'\,\frac{A_{AB}(\boldsymbol{r}_1, \boldsymbol{r}_2, \omega')}{\omega - \omega' + i\Gamma},$$

and the time-ordered correlation function

$$C_{AB}^T(\boldsymbol{r}_1, \boldsymbol{r}_2, \omega) = \lim_{\Gamma\to 0} i \int d\omega'\,\frac{A_{AB}(\boldsymbol{r}_1, \boldsymbol{r}_2, \omega')}{\omega - \omega' + i\Gamma sgn(\omega' - \mu_T)}, \qquad (3.17)$$

with the definition of the function A_{AB} in the denominator as:

$$2\pi A_{AB}(\boldsymbol{r}_1, \boldsymbol{r}_2, \omega) = C_{AB}(\boldsymbol{r}_1, \boldsymbol{r}_2, \omega)[1 \mp \exp(-\beta(\omega - \mu_T))].$$

For the reason that we are primarily interested in experiments in which we perceive effects such as excitons, i.e. where there are as few disturbances as possible, such as phonons or the like, which would smear out interesting characteristics [49–51], we usually work in the zero temperature limit $T \to 0$ and with a fixed number of particles. In this limit the function A_{AB} simplifies to:

$$A_{AB}(\boldsymbol{r}_1, \boldsymbol{r}_2, \omega) = \begin{cases} \sum_{\lambda'} \delta(\omega - \epsilon_{\lambda'})A_{0\lambda'}(\boldsymbol{r}_1)B_{\lambda 0}(\boldsymbol{r}_2) & , \text{if } \omega > \mu_T \\ \mp\sum_{\lambda'} \delta(\omega - \epsilon_{\lambda'})B_{0\lambda}(\boldsymbol{r}_2)A_{\lambda'0}(\boldsymbol{r}_1) & , \text{if } \omega < \mu_T \end{cases}$$

with $\epsilon_{\lambda'} = \pm(E_{\lambda'} - E_0)$, where $+$ is for $\omega > \mu_T$ (resonant), and $-$ is for $\omega < \mu_T$ (anti-resonant). With this tool in our toolbox, we are now able to directly write down any many-body correlation function in its spectral representation by specifying only the operators $\hat{A}$ and $\hat{B}$. As we will see in the following, there are many correlation functions that are important for us.

3.3.2   Definition of the one-body Green's function

The first correlation function we will derive is the one-body Green's function, but before we do so, we will first familiarise ourselves with the concept of Green's functions.

The Green's function formalism offers a mathematically rigorous framework to tackle the complexities of many-body physics. Furthermore, it provides a valuable physical interpretation grounded in the dynamics of moving and interacting particles. The additional notable advantage of Green's function formalism is its ability to employ the entire quantum field theory machinery, including Feynman diagrams [52], Schwinger's functional derivatives [53], and Dyson equations [45], within the solid-state physics domain.

Within this framework, we adopt the general definition of the time-ordered one-body Green's function at T = 0 K, which, in second quantization, can be expressed as [45]:

$$G(1,2) := -i\langle N|T[\hat{\psi}(1)\hat{\psi}^\dagger(2)]|N\rangle = -i\langle|T[\hat{\psi}(1)\hat{\psi}^\dagger(2)]|\rangle,$$

where $|N\rangle$ represents the ground state of the interacting system corresponding to the full many-body Hamiltonian and $\hat{\psi}^\dagger$, $\hat{\psi}$ are fermionic creation and annihilation operators. The time-ordered product, in the definition of the Green's function, can be elaborated as:

$$G(1,2) = -i[\theta(t_1 - t_2)\langle N|\hat{\psi}(1)\hat{\psi}^\dagger(2)|N\rangle - \theta(t_2 - t_1)\langle N|\hat{\psi}^\dagger(2)\hat{\psi}(1)|N\rangle]. \tag{3.18}$$

To imbue physical significance into this quantity, we associate $t_1 > t_2$ with the propagation of an electron from $\boldsymbol{r}_2$ to $\boldsymbol{r}_1$, while $t_2 > t_1$ describes the propagation of the anti-particle of the electron (backwards propagation in time), commonly referred to as a hole, from $\boldsymbol{r}_1$ to $\boldsymbol{r}_2$.

In order to understand why Green's functions are a practical tool, we can look at the remarkable attribute of the one-body Green's function, which lies in its capacity to recast the expectation value of a one-body operator $\hat{A} = \lim_{\boldsymbol{r}' \to \boldsymbol{r}} \int d\boldsymbol{r}\, A(\boldsymbol{r})\hat{\psi}^\dagger(\boldsymbol{r}')\hat{\psi}(\boldsymbol{r})$ from a time-independent Hamiltonian as [38]:

$$\langle N|\hat{A}|N\rangle = -i \lim_{\boldsymbol{r}' \to \boldsymbol{r}} \lim_{t' \to t^+} \int d\boldsymbol{r}\, A(\boldsymbol{r})G(\boldsymbol{r}, \boldsymbol{r}', t, t').$$

An illustrative example of this concept is the one-particle density, which can be expressed together with the Green's function as:

$$n(\boldsymbol{r}) = \langle N|\hat{n}(\boldsymbol{r})|N\rangle = -i \lim_{\boldsymbol{r}' \to \boldsymbol{r}} \lim_{t' \to t^+} \int d\boldsymbol{r}\, \delta(\boldsymbol{r}' - \boldsymbol{r})G(\boldsymbol{r}, \boldsymbol{r}', t, t') = -iG(\boldsymbol{r}, \boldsymbol{r}, t, t^+),$$

where $t^+ = \lim_{\nu \to 0} t + \nu$.

At this point, the result that we can write the density function using Green's functions may seem less spectacular. However, it is an essential indicator of our path because the one-body Green's function becomes a density by taking the spatial and temporal limit. Since we aim to obtain the density-density response function using Green's function formalism, the connection between Green's functions and density is already significant.

Having established the formidable utility of the Green's function, we can ask ourselves: *How can we draw a connection between the one-body Green's function and the formalism of correlation functions?* As we have seen above, the Green's function is an expectation value of two time-ordered operators, namely $\hat{B} = \hat{\psi}^\dagger$, $\hat{A} = \hat{\psi}$. According to what we have learned about correlation functions

in the last section, we can, therefore, understand the one-body Green's function as a time-ordered correlation function $-iC^T_{\hat{\psi}\hat{\psi}^\dagger}(1,2) = G(1,2)$ of these two operators.

This perspective allows us to directly formulate the one-body Green's function in spectral representation (Eq. (3.17)) at zero-temperature [38]:

$$G(\boldsymbol{r}_1, \boldsymbol{r}_2, \omega) = \lim_{\Gamma \to 0} \sum_\lambda \frac{\Psi_\lambda(\boldsymbol{r}_1)\Psi^*_\lambda(\boldsymbol{r}_2)}{\omega - E_\lambda + i\Gamma \mathrm{sgn}(E_\lambda - \mu_T)}, \tag{3.19}$$

where $E_\lambda > \mu_T$ corresponds to $\Psi_\lambda(\boldsymbol{r}_1) = \langle N|\hat{\psi}(\boldsymbol{r}_1)|\lambda\rangle$ and $E_\lambda = E(N+1, \lambda) - E_0$, while $E_\lambda < \mu_T$ corresponds to $\Psi_\lambda(\boldsymbol{r}_1) = \langle \lambda|\hat{\psi}(\boldsymbol{r}_1)|N\rangle$ and $E_\lambda = E_0 - E(N-1, \lambda)$.

With the spectral representation, we can once again breathe some physical meaning into the whole concept: The denominators of the Green's function (Eq. (3.19)) have zeroes, which are therefore poles. From a physical point of view, the poles represent, i.e. the electron addition/removal energies $\omega = E_\lambda$. The numerators represent amplitudes, which, while not normalized, adhere to the completeness relation $\sum_\lambda \Psi_\lambda(\boldsymbol{r})\Psi^*_\lambda(\boldsymbol{r}') = \delta(\boldsymbol{r} - \boldsymbol{r}')$ [54].

This brief introduction to Green's functions can be extended at will [23, 38, 45, 46]. However, we will only derive what we need for this work. Therefore, it should be emphasised that we cannot work with Green's function as it stands as equation (3.19) any more than we can work with the density-density response function. The common problem is, we also have to know the eigenstates and energies of the many-body Hamiltonian.

Confronted with this problem, we must address the question: *Can we derive an equation for the one-body Green's function that is independent of the eigenstates and energies of the many-body Hamiltonian?*

To answer this question, we need to delve deeper into the formalism of Green's functions, which leads us to the equation of motion for Green's functions in the following section.

3.3.3   Equation of Motion for Green's functions

As we have demonstrated in section 3.3.2, the one-body Green's function (Eq. (3.19)) is a nontrivial quantity, demanding the knowledge of many-body eigenenergies and states. Consequently, solving this problem in a general manner remains elusive. Therefore, we will pursue an alternative approach to tackle this challenge: The Equation of Motion (EoM) for the one-body Green's functions. For our journey, we will follow the thesis of M. Gatti [38] and S. Albrecht [54], as well as the book interacting electrons [46].

We start from the many-body Hamiltonian of the electrons (Eq. (3.1)), which, in second quantization, can be expressed as:

$$\hat{H} = \int d\boldsymbol{r}_1\, \hat{\psi}^\dagger(\boldsymbol{r}_1)h_0(\boldsymbol{r}_1)\hat{\psi}(\boldsymbol{r}_1) + \frac{1}{2}\int d\boldsymbol{r}_1 d\boldsymbol{r}_2\, \hat{\psi}^\dagger(\boldsymbol{r}_1)\hat{\psi}^\dagger(\boldsymbol{r}_2)v_c(|\boldsymbol{r}_1 - \boldsymbol{r}_2|)\hat{\psi}(\boldsymbol{r}_2)\hat{\psi}(\boldsymbol{r}_1).$$

Here $h_0(\boldsymbol{r}_1) = -\frac{1}{2}\partial_{\boldsymbol{r}_1} + v_{ext}(\boldsymbol{r}_1)$ describes the non-interacting system in an static external field.

Since the EoM describes the change in Green's function over time, so the temporal evolution of electron/hole propagation, we start with the time derivative of equation (3.18)

$$i\partial_{t_1}G(1,2) = \partial_{t_1}[\theta(t_1 - t_2)\langle|\hat{\psi}(1)\hat{\psi}^\dagger(2)|\rangle - \theta(t_2 - t_1)\langle|\hat{\psi}^\dagger(2)\hat{\psi}(1)|\rangle].$$

To work this equation further we use $\partial_{t_1}\theta(t_1-t_2) = -\partial_{t_1}\theta(t_2-t_1) = \delta(t_1-t_2)$ and the completeness relation $\langle|\hat{\psi}(\boldsymbol{r}_1)\hat{\psi}^\dagger(\boldsymbol{r}_2) + \hat{\psi}^\dagger(\boldsymbol{r}_2)\hat{\psi}(\boldsymbol{r}_1)|\rangle = \delta(\boldsymbol{r}_1 - \boldsymbol{r}_2)$, obtaining:

$$i\partial_{t_1}G(1,2) = \delta(1,2) - i\left\langle|T[i\partial_{t_1}\hat{\psi}(1)\hat{\psi}^\dagger(2)]|\right\rangle.$$

In order to derive an expression for the time evolution of the field operator within the time-ordered product, we utilize the equation of motion for the field operator $i\partial_{t_1}\hat{\psi}(1) = h_0(1)\hat{\psi}(1) + \int d2\, v_c(1,2)\hat{\psi}^\dagger(2)\hat{\psi}(2)\hat{\psi}(1)$. This allows us to rewrite the EoM of the Green's function as follows:

$$[i\partial_{t_1} - h_0(1)]G(1,2) = \delta(1,2) - i\int d3\, v_c(1,3)\left\langle|T[\hat{\psi}^\dagger(3^+)\hat{\psi}(3)\hat{\psi}(1)\hat{\psi}^\dagger(2)]|\right\rangle. \tag{3.20}$$

To simplify this expression for further use, we employ the freedome of permutations inside the time-ordered product and define the two-particle Green's function as:

$$G_2(1,2,3,4) := (-i)^2\left\langle|T[\hat{\psi}(1)\hat{\psi}(2)\hat{\psi}^\dagger(4)\hat{\psi}^\dagger(3)]|\right\rangle. \tag{3.21}$$

Substituting equation (3.21) into equation (3.20), we arrive at the EoM for the one-body Green's function, including the two-particle Green's function

$$[i\partial_{t_1} - h_0(1)]G(1,2) = \delta(1,2) - i\int d3\, v_c(1,3)G_2(1,3,2,3^+). \tag{3.22}$$

The physical interpretation of the EoM for the one-body Green's function is fascinating, as it offers insights into the physical processes occurring when an electron or hole passes through a material—the one-body term on the left-hand side of equation (3.22) describes the motion of electrons or holes, inducing polarization in the system. On the other hand, the two-particle Green's function G_2 captures the creation and annihilation of pairs of particles.

An intriguing result is achieved when one derives an EoM for the two-particle Green's function, which would involve the next higher-order, namely three-particle Green's functions, and so on. This leads to a cascade of interactions, illustrating how the original moving electron induces a complex chain reaction involving other electron-hole pairs. This intricate information is encoded in the second term on the right-hand side of equation (3.22).

Since we now have an equation of determination for one-body Green's functions (Eq. (3.22)) that is independent of the eigenfunctions and energies but depends on the two-particle Green's function, the question arises: *Where do we get the two-particle Green's function?*

In principle, different approaches can be employed, including diagrammatic techniques [45, 52] or Schwinger's functional derivative [53]. However, we will follow the path of correlation functions, as previously introduced. In this regard, we define the two-particle Green's function (Eq. (3.21)) function as [46]:

$$(-i)^2 C(1,2) \rightarrow G_2(1,2,1',2'),$$

with operators $\hat{A}_H(\boldsymbol{r}_1,t_1) \rightarrow \hat{\psi}^\dagger(\boldsymbol{r}_{1'},t_1^+)\hat{\psi}(\boldsymbol{r}_1,t_1)$ and $\hat{B}_H(\boldsymbol{r}_2,t_2) \rightarrow \hat{\psi}^\dagger(\boldsymbol{r}_{2'},t_2^+)\hat{\psi}(\boldsymbol{r}_2,t_2)$. This allows us to express the two-particle correlation function as:

$$L(1,2,1',2') = -G_2(1,2,1',2') + G(1,1')G(2,2'), \tag{3.23}$$

where we subtract the uncorrelated part GG from G_2. The choice of sign in equation (3.23) is merely a convention and should not lead to confusion. Applying the equation (3.23) to our purpose,

as shown in equation (3.22), we find:

$$L(1, 3, 2, 3^+) = iG(1, 2)n(3) - G_2(1, 3, 2, 3^+). \qquad (3.24)$$

Now we can insert equation (3.24) into the EoM for the one-body Green's function (Eq. (3.22)), obtaining:

$$[i\partial_{t_1} - h_0(1) - v_H(1)]G(1, 2) = \delta(1, 2) + \int d3 \, \Sigma_{xc}(1, 3)G(3, 2). \qquad (3.25)$$

For simplicity, we rewrote the equation by employing the identity relation of the Green's functions $\int d3 \, G^{-1}(4, 3)G(3, 2) = \delta(4, 2)$, and defining the exchange-correlation part of the self-energy

$$\Sigma_{xc}(1, 3) := i \int d4d5 \, v_c(1, 5)L(1, 5, 4, 5^+)G^{-1}(4, 3). \qquad (3.26)$$

The name exchange-correlation arises from the fact that it encapsulates the complete quantum mechanical information of the interaction. On the other hand, the full self-energy $\Sigma = v_H + \Sigma_{xc}$ represents the total potential experienced by the particle due to its environment.

With the expression (3.25), we have now found an equation of determination for the one-body Green's functions, independent of the eigenenergies and eigenstates of the many-body electron Hamiltonian. However, the problem remains that within the determination equation, there is a quantity that we do not know. In this case, it is the two-particle correlation function. The fascinating thing is that the quantity we are missing to calculate Green's functions is precisely the quantity we seek. This quantity describes the correlation function between two particles in an environment, similar to the density-density correlation function. So the crucial question arises: *Can we convert the two-particle correlation function into the density-density correlation function?*

3.3.4 Two-particle correlation function

As we have already seen, it is to be expected that the density-density and the two-particle correlation function belong together because by taking the limit for the Green's functions in time and space, one obtains the density.

If we now put the whole idea into an equation by starting with the definition of the two-particle correlation function (Eq. (3.23)) and taking the time and space limit, we obtain

$$L(1, 2, 1', 2') \rightarrow L(1, 2, 1^+, 2^+) := -G_2(1, 2, 1^+, 2^+) + G(1, 1^+)G(2, 2^+). \qquad (3.27)$$

The correlation function in time and space limit could be rewritten by using the connection between density and Green's function $G_2(1, 2, 1^+, 2^+) := (-i)^2 \langle |T[\hat{\psi}(1)\hat{\psi}(2)\hat{\psi}^\dagger(2^+)\hat{\psi}^\dagger(1^+)]| \rangle = \langle T[\hat{n}(1)\hat{n}(2)] \rangle$,
$(-i)G(1, 1^+) = n(1)$, in order to obtain:

$$L(1, 2, 1^+, 2^+) = \langle T[\hat{n}(1)\hat{n}(2)] \rangle - n(1)n(2). \qquad (3.28)$$

However, as we can see the obtained equation defines nothing else than a density-density correlation function, with the second term representing the uncorrelated part.

Comparing the definition of χ (Eq. (3.3)) with equation (3.28), we observe that by multiplying equation (3.28) with $(-i)$, we obtain the density-density correlation function. Since there is a time-order relation involved, we obtain the time-ordered density-density response function by accounting

for the uncorrelated part:

$$\chi^T(1,2) = -iL(1,2,1^+,2^+).$$

Since the quantity of interest is a retarded density-density response function (Eq. (3.7)), we can also express it as [46]:

$$\chi(1,2) := \chi^R(1,2) = -iL^R(1,2,1^+,2^+) = -i\theta(t_1 - t_2)\langle|[\hat{n}(1)\hat{n}(2)]|\rangle,$$

what exactly is what we were looking for. Since the time-ordered quantities χ^T and L are fundamental quantities in many-particle expansions and they are closely related to the retarded quantity (Eq. (3.16)), there is no problem in staying in the time-ordered framwork, which we will do in the following.

The link between the correlation functions shows that we can distil the essential information for the density-density correlation function from the complicated structure of the two-particle correlation function. Even though it is very practical, the crucial question now remains: *How can we calculate the two-particle correlation function?*

The idea now is to exploit the link between χ and L and remember that the density-density correlation function can be calculated as (Eq. (3.7)): $\chi(1,2) = \frac{\delta n(1)}{\delta v_{per}(2)}$.

This expression, together with the connection between the one-body Green's function and the density, allows us to write the link between χ and L^R in a more general form

$$\chi(1,2) = -iL^R(1,2,1^+,2^+) = \frac{\delta n(1)}{\delta v_{per}(2)} = -i\frac{\delta G_{v_{per}}(1,1^+)}{\delta v_{per}(2)}.$$

Starting from this equation, we can further elaborate the general structure for L. For this procedure we assume that the perturbative potential v_{per} is a non-local perturbative potential in both time and space, e.g. $v_{per}(1,2) := v_{per}(1)\delta(1,2)$, which allows us to construct the generalized form for the two-particle correlation function

$$L(1,3,2,4) = \frac{\delta G_{v_{per}}(1,2)}{\delta v_{per}(3,4)}. \tag{3.29}$$

Since we have shown a more argumentative derivation here, we would like to refer to more rigorous derivations in the literature [38, 46, 53].

As it is usually the case in retrospect, the answers to previously unclear questions seem trivial. The density-density correlation function describes how the density changes with the perturbation by an external field, and the two-particle correlation function describes how the propagation of a particle changes with the perturbation by an external field. Intuitively, it is immediately apparent that these two quantities must be related. The interesting question now arises: *Why did we even look for this functional derivative of the Green's function if we already know the functional derivative of density?*

Of course, in principle, it is possible to work with density in the so-called framework of time-dependent density functional theory [55], but its advantage is also its disadvantage. Density is a collective variable and contains information about many particles, making it very handy. On the other hand, it is very complicated to understand the physics of the process and, thus, to derive equations that would be helpful to obtain the dielectric susceptibility. For the one-body Green's function it is different. This quantity is physically very intuitive since it describes the propagation

of a particle in its environment. For this reason, it is also relatively intuitive to derive equations for the two-particle correlation function. However, the price is that we have more information than we need to describe the experiments.

Consequently, we stick to the Green's functions formalism, and our task for the next section will be to derive the determination equation for the two-particle correlation function from the functional derivative in equation (3.29). Going this way will lead us to the famous equation obtained by E. E. Salpeter and H. A. Bethe, the Bethe-Salpeter equation [19].

3.4   Bethe-Salpeter equation

In the previous section, we explored how to derive one- and two-particle correlation functions. We also established a connection between the density-density response function and the two-particle correlation function. Furthermore, we obtained an expression for the two-particle correlation function as the derivative of the one-body Green's function with respect to an external perturbation (Eq. (3.29)).

Now, we aim to further develop this function to obtain an expression that is more workable, known as the Bethe-Salpeter equation (BSE) [19]. To simplify our notation and avoid cumbersome expressions, we will adopt an even simpler notation than before. We define the integration $\int d_1 \, A(1)B(1)$ as $\int d_1 \, A(1)B(1) =: A(\bar{1})B(\bar{1})$.

3.4.1   Derivation of the Bethe Salpeter equation

From expression (3.29), we will now derive what is known as the BSE. The BSE is a Dyson-like equation, which is, in principle, invertible, making it a valuable tool for obtaining the two-particle correlation function L. For the derivation in this section, we follow [38].

To derive the BSE, we need to find an equation for the two-particle correlation function $L = \delta G/\delta v_{per}$ (Eq. (3.29)) that we can use to transform G. For this purpose, we use the identity $G^{-1}G = 1$ and apply the derivative concerning v_{per}, which, after employing the product rule and inverting one Green's function, leads to

$$\frac{\delta G_{v_{per}}(1,1')}{\delta v_{per}(2,2')} = -G_{v_{per}}(1,\bar{3})\frac{\delta G^{-1}_{v_{per}}(\bar{3},\bar{4})}{\delta v_{per}(2,2')}G_{v_{per}}(\bar{4},1'). \tag{3.30}$$

Next, we need an expression for the inverse of the one-body Green's function G^{-1}. Such an equation can be derived by relating the EoM of the one-body Green's function

$$\left[i\partial_{t_1} - h_0(1) - v_H(1) - v_{per}(1)\right]G_{v_{per}}(1,2) = \delta(1,2) + \Sigma_{xc}(1,\bar{3})G_{v_{per}}(\bar{3},2), \tag{3.31}$$

to the EoM of the independent one-body Green's function G_0

$$\left[i\partial_{t_1} - h_0(1) - v_{per}(1)\right]G_{0,v_{per}}(1,2) = \delta(1,2), \tag{3.32}$$

which describes the propagation of a free particle under external perturbation. Using the property $G_{v_{per}/0}(1,\bar{2})G^{-1}_{v_{per}/0}(\bar{2},3) = \delta(1,3)$ in both equations gives us the possibility to bring the equations together so that we obtain an expression for the inverse one-body Green's function

$$G^{-1}_{v_{per}}(1,2) = G^{-1}_{0,v_{per}}(1,2) - \Sigma(1,2). \tag{3.33}$$

Inserting this equation into equation (3.30) and using the chain rule $\delta\Sigma/\delta v_{per} = \delta\Sigma/\delta G_{v_{per}}\delta G_{v_{per}}/\delta v_{per} = \delta\Sigma/\delta G_{v_{per}}L$, will lead us to:

$$\frac{\delta G_{v_{per}}(1,1')}{\delta v_{per}(2,2')} = G_{v_{per}}(1,2)G_{v_{per}}(2',1') + G_{v_{per}}(1,\bar{3})G_{v_{per}}(\bar{4},1')\frac{\delta\Sigma(\bar{3},\bar{4})}{\delta G_{v_{per}}(\bar{5},\bar{6})}L(\bar{5},2,\bar{6},2').$$

The next step is to take the limit of weak perturbation $v_{per} \to 0$, obtaining:

$$L(1,2,1',2') = G(1,2)G(2',1') + G(1,\bar{3})G(\bar{4},1')\frac{\delta\Sigma(\bar{3},\bar{4})}{\delta G(\bar{5},\bar{6})}L(\bar{5},2,\bar{6},2').$$

Finally, we define the uncorrelated part GG as $L_0(1,2,1',2') := G(1,2)G(2',1')$, which gives us the famous BSE in theoretical spectroscopy:

$$L(1,2,1',2') = L_0(1,2,1',2') + L_0(1,\bar{4},1',\bar{3})\frac{\delta\Sigma(\bar{3},\bar{4})}{\delta G(\bar{5},\bar{6})}L(\bar{5},2,\bar{6},2'). \qquad (3.34)$$

With equation (3.34) it is now formally possible to calculate the two-particle correlation function L by inverting the equation $L = [1 - L_0(\delta\Sigma/\delta G)]^{-1}L_0$. As we will see later, however, this is not the most innovative way to solve this problem because we would have to invert a four-point function in which each index depends on time, position and spin.

For the moment, that is our minor problem. As we see in the equation, the BSE depends on the self-energy Σ, but this quantity itself depends on L (Eq. (3.26)). However, this means we need the final information L in advance in order to be able to solve the equation exactly at all. So before we venture into a more ingenious solution for BSE, we should ask ourselves the question: *Where do we get the self-energy for the BSE?*

3.4.2   The Random-Phase, Hatree-Fock and GW-approximation

Since the self-energy is exact but contains Green's function or the two-particle correlation function, which itself contains the self-energy, solving the problem necessitates an iterative approach. However, this iterative procedure would be numerically complex and the way how to do it is in general unclear. The more established way is to make approximations for the self-energy, and therefore also for all quantities, which include the self-energy, like the one-body Green's function G, the uncorrelated part L_0 of the BSE, and the functional derivative of the kernel $\delta\Sigma/\delta G$.

In the following, we will explore four commonly used approximations for the self-energy: The independent-particle approximation (IP), the Random-Phase approximation (RPA), the time-dependent Hartree-Fock (TDHF) and the *GW* approximation.

For the purpose of making approximations, we will start from the exact equation for the self-energy (Eq. (3.26)) [46]:

$$\Sigma(1,2) = v_H(1,2) - iv_c(1,\bar{4})G(1,\bar{3})\left[\frac{\delta G^{-1}(\bar{3},2)}{\delta v_{per}(\bar{4}^+)}\right]$$

$$= v_H(1,2) + \Sigma_x(1,2) + iG(1,\bar{3})\Xi(\bar{3},\bar{5},2,\bar{6})v_c(1^+,\bar{4})L(\bar{6},\bar{4},\bar{5},\bar{4}^+),$$

where we used the definintion of the interaction-kernel $\Xi := i\delta\Sigma/\delta G = i\delta(v_H + \Sigma_{xc})/\delta G$.

The last line of the self-energy offers us an ideal starting point for our analysis: The first term corresponds to the classical mean-field Hartree potential; the second term is the exact exchange term

(non-local exchange), defined as $\Sigma_x(1,2) = iG(1,2)v_c(1^+,2)$ and the last term represents a generalized induced non-local potential. Since the first two terms are less complicated to understand from a physical point of view, we will take a closer look at the last one.

To make the physical meaning of the induced non-local potential more visible we consider equation (3.8), where the classical induced potential was written as $v_{ind} = v_c n = v_c \chi v_{per} = v_c \chi v_c n$. If we compare the structure of the equation for the classical induced potential with the last equation and recall that $L \to \chi$ and $G \to n$, we can define $\Sigma_c = v_c L(i\Xi)G$ in order to work out the relation between the classical induced potential ($\Xi = v_c$) and the more general one ($\Xi = i\delta(v_H + \Sigma_{xc})/\delta G$). Having worked out the general structure, we can observe that Σ_c, expresses the change in the induced potentials due to the propagation of an extra particle. In other words, the extra particle polarizes the system and causes other particles to move, and this reaction self-consistently modifies the potentials that the extra particle experiences. Due to the description of the correlation between the particles, Σ_c is also called the correlation part.

After we have worked out the physical meaning of the individual terms of the self-energy, we can look at how they contribute to an absorption spectrum.

To do so, we proceed as follows:

1. We choose an approximation Σ^{apx} for the self energy $\Sigma \approx \Sigma^{apx}$

2. We use the approximated self-energy for the calculation of the one-body Green's function G^{apx}

3. We construct the uncorrelated part of the BSE $L_0^{apx} = G^{apx}G^{apx}$ using the Green's function from step 2

4. We approximate the functional derivative $\delta\Sigma/\delta G \approx \delta\Sigma^{apx}/\delta G^{apx}$

5. We calculate the absorption spectrum using the approximated quantities from steps 3 and 4

To familiarize ourselves with the approximations, we will start our journey with the most basic approximations, namely, where we consider the particles as independent particles (IP)

$$\Sigma^{IP}(1,2) = 0, \tag{3.35}$$

and the RPA, where we only include the classical mean-field Hartree potential in the self-energy

$$\Sigma^{RPA}(1,2) = v_H(1,2). \tag{3.36}$$

It is important to note that the RPA approximation means, as seen at the end of section 3.2 that we consider the crystal local fields in the calculations and only the classical part of the coulomb interaction for the interaction between the particles. Furthermore, we should remember that by both approximations, we neglect the exchange-correlation part entirely, i.e., all quantum mechanical parts in the calculations, both for Green's functions and the BSE. This means that what we consider in our calculations are classical particles that either interact via an averaged field (RPA) or do not interact at all (IP).

In the following, we will examine how these assumptions affect the absorption spectra for Si and LiF compared to experimental spectra.
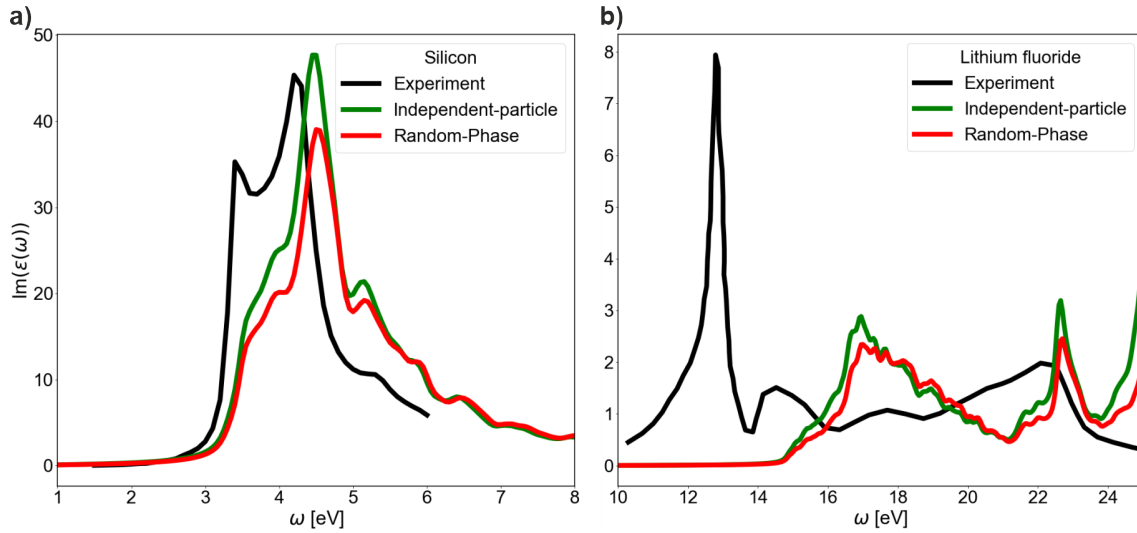
Figure 3.1: Comparison of the absorption spectra ($\mathrm{Im}(\epsilon(\omega))$) in the IP approximation (Eq. (3.35)) (green), RPA (Eq. (3.36)) (red) and experimental results (black) for **a)** silicon and **b)** lithium fluoride. It turns out that for materials with bound excitons close to the optical band edge (a)) and materials with strongly bound excitons (b)), the contribution of the local field primarily causes a reduction in the intensity of the spectrum. In both cases, no additional structures are visible due to the consideration of the local field effects. Numerical values are given in Appendix A; calculations were performed on an 8k grid (2048 k-points) using a Kohn-Sham band structure. Experimental data were taken from [40] for silicon and from [41] for lithium fluoride.

Comparing the RPA and IP approximations in figure 3.1, we see that the variation of the Hartree potential in the RPA lowers the overall magnitude of the absorption spectrum in both Si and LiF. The reason for a reduction in intensity is that the Hartree fraction represents a repulsive collective interaction of the electrons and, thus, contributes to a reduction in the excitation number of the electrons, which now repel each other compared to the independent particle spectrum.

However, it does not introduce additional features that would provide a good description of experimental results as we have seen them in the beginning. This suggests that classical collective interactions, given by the RPA, are not sufficient to explain the observed features in absorption experiments, so we need to include effects beyond, namely quantum-mechanical interactions.

Before we do so, however, it should be mentioned that, as we have already seen in the introductory chapter 2 (Fig. 2.2), a classical description of the interactions through the RPA is sufficiently good to describe EELS since we are dealing in this case with a collective oscillation of electrons and not with individual electron-hole pairs.

Having clarified this point, we will consider the quantum mechanical effects piece by piece.

In a system where correlation effects are less critical (meaning $L = 0$ or $\Xi = 0$), such as a system with very few electrons in a large space, the self-energy is approximately given by:

$$\Sigma(1,2) \approx v_H(1,2) + \Sigma_x(1,2) = -i\delta(1,2)v_c(1,\bar{3})G(\bar{3},\bar{3}^+) + iv_c(1^+,2)G(1,2).$$

The obtained expression is known as the Hartree-Fock approximation [46]. If we use this approximation for the self-energy, i.e. if we build up both the Green's function and thus L_0 with it as

described at the beginning and also approximate the interaction kernel with it, then we obtain the so-called time-dependent Hartree-Fock approximation of the BSE.

Physically, we can understand this approximation as follows: Suppose we have two non-interacting particles, but of quantum mechanical nature, then there are two parts we have to consider in the two-particle Green's function (Eq. (3.24)), namely firstly that the particles follow their path completely unaffected by each other or that the particles are indistinguishable and therefore they are exchanged. These two components are then reflected in the self-energy as Hartree potential (v_H) and exchange component (Σ_x).

If we consider only these two components in the functional derivation of the BSE, we obtain:

$$L^{TDHF}(1, 2, 1', 2') = L_0^{HF}(1, 2, 1', 2')$$

$$-iL_0^{HF}(1, \bar{4}, 1', \bar{3}) \left[v_c(\bar{3}, \bar{5})\delta(\bar{3}, \bar{4})\delta(\bar{5}, \bar{6}) - v_c(\bar{3}, \bar{4})\delta(\bar{3}, \bar{5})\delta(\bar{4}, \bar{6})\right] L^{TDHF}(\bar{5}, 2, \bar{6}, 2').$$

We see that both components are only described by the Coulomb component, but the indices are reversed so that these components do not cancel each other out, although they have opposite signs. The reason for the different signs is that the first part, which follows from the Hartree potential, describes a repulsive interaction, as we have stated in the case of RPA, and consequently the second part, which follows from the exchange part of self-energy, describes an attractive interaction.

Before we go into the physics of this interaction in more detail, let us consider the result from the work of Bruneval [56] for the absorption spectrum of Si (Fig. 3.6 in [56]). Using the TDHF approximation in the BSE, one can observe that it leads to a blue shift of the spectrum and to the generation of a single peak. The reason for this can be attributed to the fact that the attractive portion described by the exchange part is too strong and would, therefore, lead to strongly bound excitons in all materials [56].

In order to understand the whole thing more intuitively, let us go back to the two-particle Green's function. For the Hartree-Fock approximation, we assumed that we only consider two particles that do not interact with each other, and therefore, the self-energy only consists of an independent part and an exchange part. However, when we discussed the EoM of Green's function, we noted that the problem of the emerging two-particle Green's function could theoretically be solved iteratively, but this would result in the two-particle Green's function becoming a three-particle Green's function and this a four-particle Green's function, etc.. This shows us that the name two-particle Green's function is deceptive because it contains information about the interaction of all particles with each other. So, what we neglect by the Hartree-Fock approximation is the effect of the other electrons and holes in the BSE. For this reason, we can refer to the particles in the Hartree-Fock approximation as bare electrons and bare holes since we neglect the entire environment that interacts with them.

The fact that the TDHF BSE description of absorption spectra in bulk materials is inadequate means that we have to turn the bare particles into dressed particles, i.e. take their environment into account. Since the Hartree-Fock approximation also follows directly from the fact that we have neglected the correlation part $(\Sigma_c = 0)$ in the self-energy, we now have to take this part into account in order to turn the bare particles into dressed ones. The problem, however, is that we do not know the two-particle correlation function L, which means we have to make an approximation.

For this reason, we will assume that the interaction in the correlation part is mainly due to the Hartree part $\Xi = i\delta\Sigma/\delta G = i\delta(v_H + \Sigma_{xc})/\delta G \approx i\delta v_H/\delta G = v_c$. This assumption is because the Hartree part in many materials represents the most significant proportion in terms of the magnitude of the self-energy. The approximation allows us to write the exchange-correlation part of the self-energy as [46]:

$$\Sigma_{xc}(1,2) \approx \Sigma_{xc}^{GW}(1,2) = \Sigma_x(1,2) + v_c(1^+,\bar{3})G(1,2)L(\bar{4},\bar{3},\bar{4}^+,\bar{3}^+)v_c(2,\bar{4}^+).$$

Further elaborating the expression for the self-energy by taking into account that $-iL(\bar{4},\bar{3},\bar{4}^+,\bar{3}^+) = \chi(4,3)$ holds. We can rewrite the exchange-correlation part of the self-energy in the so called GW-approximation as [57]:

$$\begin{aligned}
\Sigma_{xc}^{GW}(1,2) &= iG(1,2)\left(v_c(1^+,2) + v_c(1^+,\bar{3})\chi(\bar{3},\bar{4})v_c(\bar{4},2)\right) = \\
&= iG(1,2)\epsilon^{-1}(1^+,\bar{4})v_c(\bar{4},2) := iG(1,2)W(1^+,2),
\end{aligned} \tag{3.37}$$

where W is the screened Coulomb interaction, defined as $W(1,2) = \epsilon^{-1}(1,\bar{3})v_c(\bar{3},2)$. Since we find in the BSE (Eq. (3.34)) the structure $\delta\Sigma_{xc}/\delta G$, we have to work out the equation (3.37) further

$$\frac{\delta\Sigma_{xc}(1,2)}{\delta G(3,4)} \approx \frac{\delta\Sigma_{xc}^{GW}(1,2)}{\delta G(3,4)} = i\frac{\delta G(1,2)}{\delta G(3,4)}W(1^+,2) + iG(1,2)\frac{\delta W(1^+,2)}{\delta G(3,4)}.$$

Before we continue with the derivation, we want to give the obtained expression a physical meaning: The first part is, of course, simply the screened exchange part, the second term, however, contains information about how screening changes when the system is perturbed [46], and it can be further evaluated as [58]:

$$\frac{\delta W(1^+,2)}{\delta G(3,4)} = \left[W(1^+,4)W(3,2) + W(1^+,3)W(4,2)\right]G(4,3).$$

Therefore, the full kernel in the GW approximation reads:

$$\frac{\delta\Sigma_{xc}^{GW}(1,2)}{\delta G(3,4)} = i\delta(1,3)\delta(2,4)W(1^+,2) + iG(1,2)\left[W(1^+,4)W(3,2) + W(1^+,3)W(4,2)\right]G(4,3).$$

Mostly, however, the second term is neglected for consistency reasons. To see this, we need to take a different route to derive the GW approximation, which is through a diagrammatic expansion of the self-energy, which up to the second order in W is given by [57]:

$$\Sigma_{xc}(1,2) \approx iG(1,2)W(1^+,2) - G(1,\bar{3})G(\bar{3},\bar{4})G(\bar{4},2)W(1^+,\bar{4})W(\bar{3},2) + \dots.$$

The first-order term in this expression is a second-order term in W, similar to what we obtained from the previous derivation of $\delta W/\delta G$. So, suppose we want to concentrate purely on the GW approximation, i.e., the first term in the diagrammatic expansion. We should also neglect the term $\delta W/\delta G$.

Under this assumption, we obtain the interaction kernel in the BSE in GW-approximation as:

$$\frac{\delta\Sigma_{xc}^{GW}(1,2)}{\delta G(3,4)} \approx i\delta(1,3)\delta(2,4)W(1^+,2),$$

which leads to the BSE in GW-approximation

$$L^{GW}(1,2,1',2') = L_0^{GW}(1,2,1',2')$$

$$-iL_0^{GW}(1,\bar{4},1',\bar{3})\left[v_c(\bar{3},\bar{5})\delta(\bar{3},\bar{4})\delta(\bar{5},\bar{6}) - W(\bar{3},\bar{4})\delta(\bar{3},\bar{5})\delta(\bar{4},\bar{6})\right]L^{GW}(\bar{5},2,\bar{6},2').$$

It is interesting to briefly analyse this equation physically. The expression shows that the interaction part that comes from the Hartree potential represents a repulsive interaction, as in the case of the TDHF approximation, and the attractive part, which was previously only described by the part coming from the exchange part of the self-energy, is now given by a screened version of the exchange part. Since the screening considers the influence of the interaction of all electrons and holes, i.e. the environment of an interacting electron-hole pair, this attractive interaction now describes the interaction between dressed particles.

To avoid confusion, the following should be briefly noted: An alternative definition of the BSE is often used, which has the same structure as above but with a positive sign instead of $(-i)$. This arises from the fact that if we start by defining L as $L = -i\delta G/\delta v_{per}$ and L_0 as $L_0 = iGG$, we end up with [21, 38, 59]

$$
\begin{aligned}
L^{GW}(1,2,1',2') &= L_0^{GW}(1,2,1',2') \\
+L_0^{GW}(1,\bar{4},1',\bar{3})&\left[v_c(\bar{3},\bar{5})\delta(\bar{3},\bar{4})\delta(\bar{5},\bar{6}) - W(\bar{3},\bar{4})\delta(\bar{3},\bar{5})\delta(\bar{4},\bar{6})\right]L^{GW}(\bar{5},2,\bar{6},2')
\end{aligned}
\tag{3.38}
$$

and the link between L and χ as $\chi(1,2) = L(1,2,1^+,2^+)$.

Now that we have decomposed the self-energy into its fundamentals and found a way to take into account both the classical part of the self-energy and the quantum mechanical part in the form of the exchange and correlation part in an approximate form in the BSE, we still lack information to make the equation useful in practice. The problem is that we do not know where to get ϵ in W, since it is the quantity, which we want to obtain at the end. For this reason, in the following section, we deal with the question: *How can we calculate the screening ϵ for W?*

3.4.3   W Random-Phase approximation

In practice, we face a challenging task. When examining equation ((3.37), (3.38)), we observe that the screened Coulomb potential W depends on the dielectric function ϵ^{-1} and therefore, according to equation (3.11) on the dielectric susceptibility χ. This implies that to solve our problem, we must first address it, which would lead to a self-consistent problem. As one might anticipate, this self-consistent approach can become computationally intensive since one must solve the problem iteratively. For this reason, researchers usually avoid employing this method and take a step back and calculate the dielectric function using the RPA

$$
\begin{aligned}
L^{RPA}(1,2,1',2') &= L_0(1,2,1',2') + L_0(1,\bar{4},1',\bar{3})\left[v_c(\bar{3},\bar{5})\delta(\bar{3},\bar{4})\delta(\bar{5},\bar{6})\right]L^{RPA}(\bar{5},2,\bar{6},2') \\
&= L_0(1,2,1',2') + L_0(1,\bar{3},1',\bar{3})v_c(\bar{3},\bar{5})L^H(\bar{5},2,\bar{5},2'),
\end{aligned}
\tag{3.39}
$$

where L_0 is constructed from Green's functions, which only consider the Hartree part in the self-energy. However, as we will see later (Ch. 4.4.2), for practical use, so-called Kohn-Sham orbitals and eigenvalues are used to construct L_0 or, to be precise, χ_0. If we now combine equation (3.38) with the connection between the dielectric susceptibility and the two-particle correlation function $(\chi(1,2) = L(1,2,1^+,2^+), \chi_0(1,2) = L_0(1,2,1^+,2^+))$, then we get for the dielectric susceptibility

$$\chi^{RPA}(1,2) = \chi_0(1,2) + \chi_0(1,\bar{3})v_c(\bar{3},\bar{4})\chi^{RPA}(\bar{4},2)$$

and for the dielectric function (Eq. (3.12)):

$$\left(\epsilon^{RPA}\right)^{-1}_{\boldsymbol{G},\boldsymbol{G}'}(\boldsymbol{q},\omega) = \delta_{\boldsymbol{G},\boldsymbol{G}'} + v_c(\boldsymbol{q}+\boldsymbol{G})\chi^{RPA}_{\boldsymbol{G},\boldsymbol{G}'}(\boldsymbol{q},\omega).$$

In this approximation, W in reciprocal space can be written as:

$$W^{RPA}_{\boldsymbol{G},\boldsymbol{G}'}(\boldsymbol{q},\omega) = \frac{4\pi\left(\epsilon^{RPA}\right)^{-1}_{\boldsymbol{G},\boldsymbol{G}'}(\boldsymbol{q},\omega)}{|\boldsymbol{q}+\boldsymbol{G}||\boldsymbol{q}+\boldsymbol{G}'|}.$$

Knowing now how we could approximate the screening, the question arises: *Why should this approximation work?* The reason for this is that spectroscopy calculations typically employ a GW calculation for the Green's function (Eq. (3.25)) and, therefore, for $L_0 \to L_0^{GW}$, which can be found in BSE, as well as for the interaction-kernel in the BSE. This approach leads to a partial error cancellation between the L_0^{GW} calculation (which opens the gap) and the BSE calculation (which closes the gap), as W has opposite signs in both cases [22]. Therefore, it is advisable to choose an entirely consistent W, denoted as ensuring uniform treatment for both the calculation of L_0 and BSE calculations [60].

However, not knowing the screening is not the only problem, we are also confronted with the fact that the screening is time dependent, as we have seen in the final BSE in GW-approximation and here in W. But this time dependency makes it much more complicated to solve the BSE since all times mixes and the two-particle correlation function would depend on three frequency parameters [59]. However if we could neglecte the time dependence of the W-part ($W(1,2) \to W(\boldsymbol{r}_1,\boldsymbol{r}_2)\delta(t_1 - t_2)$), than the BSE would reduce to a single frequency [59]. It is therefore worth asking the following question: *What is the physical meaning of this time dependence, and can we neglect it?*

We only want to clarify the reason phenomenologically here and refer to section 4.5.4 for a more detailed explanation, as we still lack some knowledge of a detailed explanation at this point.

As we have seen in the last section, screening comes into play when we consider the correlation part of the self-energy, which means that we consider correlations between electrons, holes and the entire electron gas. If we now perturb our system, e.g., creating electron-hole pairs, we change the environment of the particles, which leads to a rearrangement of them. This process, of course, is not instantaneous but retarded, which is why the screening is time-dependent and, therefore, W. If one would like to give an estimation of the time delay, one could assume that in an electron-gas, the correlation builds up in a period of the plasma oscillation $\sim 2\pi/\omega_p$ [23]. This estimation already tells us something important, which we will work out in section 4.5.4, it is the plasma frequency that determines whether a static approximation could work or not since if the plasma frequency is significant, then the retardation of the screening is small. For the moment, however, we assume that the static approximation is suitable for Si and LiF, leading us to

$$W^{RPA}_{\boldsymbol{G},\boldsymbol{G}'}(\boldsymbol{q},\omega) \to W^{RPA}_{\boldsymbol{G},\boldsymbol{G}'}(\boldsymbol{q},\omega=0) = \frac{4\pi\left(\epsilon^{RPA}_{\boldsymbol{G},\boldsymbol{G}'}\right)^{-1}(\boldsymbol{q},\omega=0)}{|\boldsymbol{q}+\boldsymbol{G}||\boldsymbol{q}+\boldsymbol{G}'|}$$

and therefore to the static or adiabatic BSE

$$\begin{aligned} L^{GW}(1,2,1',2') \to L_0^{AGW}(1,2,1',2') &= L_0^{AGW}(1,2,1',2') \\ + L_0^{AGW}(1,\bar{4},1',\bar{3})\left[v_c(\bar{3},\bar{5})\delta(\bar{3},\bar{4})\delta(\bar{5},\bar{6}) - W^A(\bar{3},\bar{4})\delta(\bar{3},\bar{5})\delta(\bar{4},\bar{6})\right]&L^{AGW}(\bar{5},2,\bar{6},2'), \end{aligned}$$

(3.40)

where the adiabatic W is defined as: $W^A(3,4) := W(\boldsymbol{r}_3, \boldsymbol{r}_4)\delta(t_3 - t_4)$.
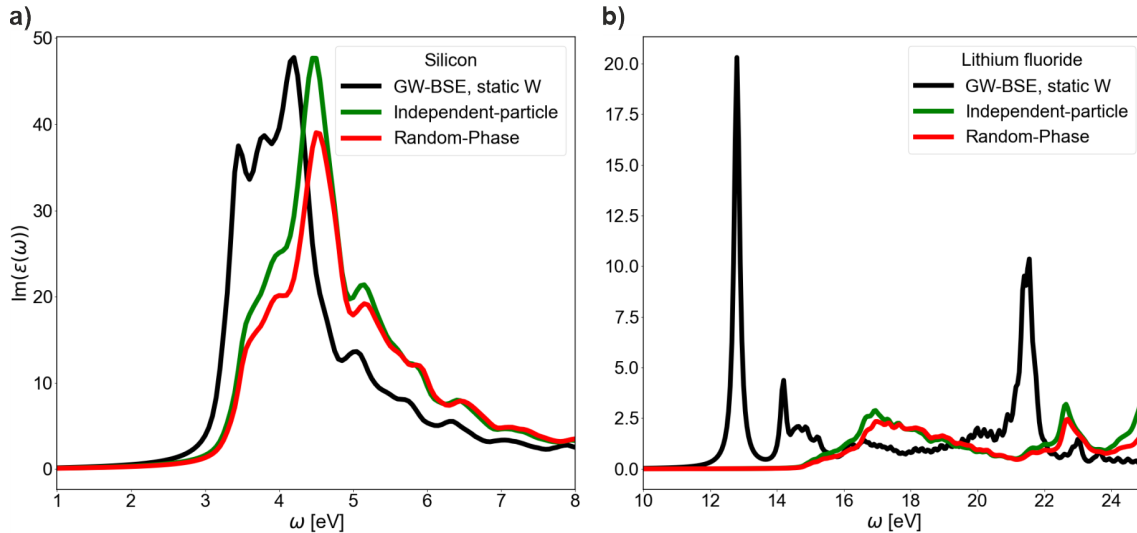

Figure 3.2: Plot of the absorption spectra ($\text{Im}(\epsilon(\omega))$) in the IP approximation (Eq. (3.35)) (green), RPA (Eq. (3.36)) (red) and static GW-BSE (Eq. (3.40)) (black) approximations for **a)** silicon and **b)** lithium fluoride. As can be seen in the case of bound excitons close to the band edge (a)), the spectrum is enhanced directly at the optical band edge due to the excitonic effects, but in the case of strongly bound excitons (b)), strong peaks form within the optical bandgap and close to the optical band edge. The reason for such behavior is the varying intensity of the excitonic effects in the two materials. Numerical values are given in Appendix A; calculations were performed on an 8k grid (2048 k-points) using a Kohn-Sham band structure.

A change in the spectrum can be observed by considering the exchange-correlation part in the static GW approximation. Generally speaking, taking these effects into account leads to additional enhancements near (Fig. 3.2 a)) or within the optical bandgap (Fig. 3.2 b)), which were not previously visible in a classical or independent-particle description. These come about because the exchange-correlation component contributes with a negative sign to the interaction, i.e. describes an attractive interaction between dressed electron-hole pairs as discussed before.

However, the difference between the materials is, at this point, not so noticeable. For this reason, we can turn to a simple description of the screening in order to make the discussion as simple as possible. For this purpose, we assume screening as a scalar quantity $\epsilon^{-1}_{\boldsymbol{G},\boldsymbol{G'}}(\boldsymbol{q}) \to \epsilon^{-1}_{mac}$ in W. If ϵ^{-1}_{mac} approaches 1, we immediately return the TDHF approximation, which leads to strongly bound excitons in all materials, as we have discussed before. The other limit is when ϵ^{-1}_{mac} becomes infinitely large; in this case, W disappears, and we are left with the RPA approximation.

In the case of Si (Fig. 3.2 a)), we are dealing with a material with a more substantial screening $\epsilon_{mac} = 11.4$ [61] i.e. the interaction between electrons and holes is more strongly shielded. Therefore, excitonic effects are less dominant. This is particularly evident when looking at the spectrum from around 4 eV, which looks very similar to the IP or RPA's structure.

However, this is not the only effect, which, in principle, must be taken into account since, according to this estimation, we would not expect such strong excitonic effects as we see them for Si. The strong excitonic effects could be traced back to the fact that Si has regions in which the bands are

almost parallel, meaning many IP transitions at almost the same energy mixes.

In the case of LiF (Fig. 3.2 b)), we see strong excitonic effects since excitations occur within the optical bandgap. This indicates that the screening is significantly weaker ($\epsilon_{mac} = 1.9$ [62]) than in Si.

Since we have already compared the static GW-BSE calculation at the beginning of the first chapter in figure 2.1 with the experimental data, we know that the BSE in static GW-approximation is sufficiently good to reproduce experimental absorption spectra for Si and LiF. For this reason, we will assume only the static GW approximation in the following and refer for a detailed discussion on this choice to section 4.5.4.

3.5   Intermediate summary

In this chapter, we have dealt intensively with obtaining the decisive quantity of spectroscopy χ, including all interactions in a many-electron system. To do this, we took two approaches. First, we considered the problem using the density-density-reponse formalism and the associated linear approximation. In this formalism, we derived χ, identified its structure as a density-density correlation function, and derived meaningful connections for later calculations. The problem with the density-density reponse formalism, however, was that to calculate the dielectric susceptibility. In principle, the many-body problem should have been solved in a first place, since we would at least have to know the oscillator strength matrix elements for the transitions from ground to excited states and the corresponding energies. For this reason, we have taken a second approach, the one via the Green's function formalism.

We have seen that the Green's function formalism is a practical and powerful tool to put the quantities we are really interested in into a form such that it is no longer necessary to solve the full many-body problem.

In this spirt, we derived the two-particle correlation function using the correlation functions formalism and directly connecting it to the density-density correlation function. This equation and the knowledge we received from density-density reponse formalism allowed us to derive the crucial equation of this chapter, the Bethe Salpeter equation.

The Bethe Salpeter equation theoretically allows the two-particle correlation function to be calculated and, thus, the density-density correlation function without having to solve the many-body problem beforehand. However, we have seen that the problem is not easy to solve, primarily due to the correlation part of the exchange-correlation self-energy. For this reason, we then delved deeper into some very important approximations. Among other things, we have derived the most crucial approximation in this field, the static GW approximation, representing the standard approximation for bulk materials.

Having arrived at this point, we are in a position to reformulate the BSE so that we can use it. So far it would have meant we would have to invert the four-point function L and then to take the diagonal part in order to get χ, which would have a very high numerical effort and was done by Hanke and Sham [63] using a small basis.

Chapter 4

The working framework

"I think it's tempting when the only tool you have is a hammer to treat everything as if it were a nail."

Abraham Maslow

So far, we have been working on obtaining a method to calculate the dielectric susceptibility quantum mechanically without it being necessary to know the eigenstates or eigenenergies of the electronic many-particle Hamiltonian (Eq. (3.1)). This is made possible by rephrasing the problem in terms of the Bethe Salpeter equation. However, the exact ingredients of this equation are not accessible in practice. Moreover, solving the BSE comes with a high numerical effort, requiring an inversion of a four-point function. For this reason, in this chapter, we will transform the static BSE in GW-approximation (Eq. (3.40)) into an effective eigenvalue equation for a two-particle Hamiltonian. In order to keep the notation as short as possible, we will always use L and L_0 in the following, whereby these naturally refer to the adiabatic GW approximation.

However, before introducing this method, we want to deal with the question: *What information do we need from the BSE to calculate an absorption spectrum?*

The answer to this question is intended to prevent us from spending unnecessary time and effort.

4.1 The macroscopic dielectric function for the absorption spectrum

The idea is now the following [64]: For absorption measurements, the interest lies primarily in the system's response to a macroscopic classical perturbation, which corresponds to the macroscopic limit ($\boldsymbol{q} \to \boldsymbol{0}$). For this reason, by solving the BSE, we are likely providing more information than we need to describe the macroscopic dielectric function. This is similar to the case where we obtain χ from L, or to the case where we calculate the one-body Green's function in order to obtain the density. For this reason, we should focus on what information we need. In order to make things not unnecessarily complicated, we present the following equations without specifying the space, time, or spin indices.

The equation we want to rewrite is the macroscopic dielectric function ($\boldsymbol{G} = \boldsymbol{G}' = \boldsymbol{0}$) in the limit $\boldsymbol{q} \to \boldsymbol{0}$, given by equation (3.12):

$$\frac{1}{\epsilon_M(\omega)} = \left[1 + v_c(0)\chi(0,\omega)\right]_{0,0} = \left[1 + v_c^0\chi(0,\omega)\right]_{0,0},$$

with the definition of the divergent long-range part of the Coulomb potential as $v_c^0 := v_c(\boldsymbol{q}+\boldsymbol{G} = \boldsymbol{0})$. To reformulate the equation for the macroscopic dielectric function, we define a new quantity $\bar{\chi}^{-1} := \chi^{-1} + v_c^0$, which we call the modified dielectric susceptibility $\bar{\chi}$, which allows us to rewrite the equation into

$$\frac{1}{\epsilon_M(\omega)} = \left[1 + v_c^0\chi(0,\omega)\right]_{0,0} = \left[1 + v_c^0 \left(\bar{\chi}^{-1}(0,\omega) - v_c^0\right)^{-1}\right]_{0,0} = \left[1 + v_c^0\bar{\chi}(0,\omega)\left(1 - v_c^0\bar{\chi}(0,\omega)\right)^{-1}\right]_{0,0} =$$

$$= \left[\left(1 - v_c^0\bar{\chi}(0,\omega)\right) + v_c^0\bar{\chi}(0,\omega)\right]_{0,0} \left(1 - v_c^0\bar{\chi}(0,\omega)\right)_{00}^{-1} = \left(1 - v_c^0\bar{\chi}(0,\omega)\right)_{00}^{-1}. \tag{4.1}$$

Up to this point, we have gained nothing except an exact algebraic transformation. The question that now arises is: *Is there a corresponding BSE for some $\bar{L}$ from which $\bar{\chi}$ can be derived?* For this purpose, we rewrite the BSE for L a little bit

$$L = L_0 + L_0[v_c + \Xi_{xc}]L = L_0 + L_0[\bar{v}_c + v_c^0 + \Xi_{xc}]L,$$

where $\Xi_{xc} := \Xi - v_c = i(\delta\Sigma_{xc}/\delta G)$ represents a general exchange-correlation interaction term, and $\bar{v}_c$ is defined in reciprocal space as:

$$\bar{v}_c(\boldsymbol{G}) = \begin{cases} 0, \text{ if } \boldsymbol{G} = \boldsymbol{0} \\ v_c(\boldsymbol{G}), \text{ if } \boldsymbol{G} \neq \boldsymbol{0}. \end{cases}$$

The next step is to invert the equation, obtaining

$$L_0^{-1} = L^{-1} + \bar{v}_c + v_c^0 + \Xi_{xc} =: \bar{L}^{-1} + \bar{v}_c + \Xi_{xc},$$

and to define $\bar{L}^{-1} := L^{-1} + v_c^0$, which is directly related to $\bar{\chi}$ via the equation $\bar{\chi}(1,2) = -i\bar{L}(1,2,1^+,2^+)$ [46]. With this definition, we can now set up the Dyson equation for $\bar{L}$ as

$$\bar{L} = L_0 + L_0\left[\bar{v}_c + \Xi_{xc}\right]\bar{L}, \tag{4.2}$$

which we call the modified BSE. The resulting equation shows us what we have gained, namely that it is not necessary to take into account the divergent contribution of the electron-hole exchange part v_c^0 for the calculation of the macroscopic dielectric function [46,64,65]:

$$\epsilon_M(\omega) = \frac{1}{\left[1 - \lim_{\boldsymbol{q}\to\boldsymbol{0}} v_c(\boldsymbol{q})\bar{\chi}(\boldsymbol{q},\omega)\right]_{00}^{-1}}, \tag{4.3}$$

for absorption spectroscopy.

Now that we have made this small but essential simplification, we will address the question: *How can we transform the BSE in the GW approximation, with a static W, into an effective eigenvalue equation?*

However, before we go into this question, we should not make the path unnecessarily complicated because up to this point, all statements about the modified as well as the regular Bethe Salpeter equation apply to arbitrary spin polarizations since we have always taken the spin in the magnitude $j = (\boldsymbol{r}_j, t_j, \sigma_j)$. But, this complicates the representation of the BSE in static GW in its effective eigenvalue equation. For this reason, we will be more specific and look directly at the effect of spin on the modified BSE.

4.2   Spin dependence of the modified Bethe Salpeter equation

Since the modified Bethe Salpeter equation depends on spin, it is advisable to start with the spin discussion first. For this reason, we will now consider the spin variable directly and explore the possible allowed transitions in absorption spectroscopy, following the approach outlined in [23, 66]. In general, there are some effects to consider at this point, for example, externally applied magnetic fields and intrinsic magnetic fields, such as those caused by spin-orbit coupling, as well as closed- (double occupied or empty states) and open-shell (unpaired electrons) systems. In the following discussion, we limit ourselves to the effect caused by spin-orbit coupling and closed-shell systems since this is the case to be found in most of the bulk materials. Dealing only with closed-shell systems means, in particular, that the total ground state spin is zero, and since only spin-conserving transitions are allowed, we deal with singlet transitions. Based on these specifications, we have two cases to consider, namely, the one where we can neglect the spin-orbit coupling and the one where we cannot.

If the spin-orbit interaction is negligible compared to the electron-hole interaction, the total spin projection, S_z, is a good quantum number. The single-particle states can then be classified as spin-up and spin-down (collinear magnetism). In this case, the Hilbert space of electron-hole pairs (vc) consists of four subspaces: $(v \uparrow c \uparrow)$, $(v \uparrow c \downarrow)$, $(v \downarrow c \uparrow)$, and $(v \downarrow c \downarrow)$ [67].

This allows us to factorize each single-particle spinor $\Psi_{m_\sigma, n, \boldsymbol{k}}(\boldsymbol{r}, \sigma)$ into a space-dependent orbital $\psi_{m_\sigma, n, \boldsymbol{k}}(\boldsymbol{r})$ and a single-particle fermionic $(\frac{1}{2})$ spin function $\chi_{\frac{1}{2} m_\sigma}(\sigma)$ with the quantum number $m_\sigma = \pm \frac{1}{2}$ and $\sigma = \uparrow \downarrow$ the spin variable [23]. Since the spin functions form an orthonormalized and complete set, they fulfil the conditions:

$$\sum_{\sigma} \chi^*_{\frac{1}{2}, m_\sigma}(\sigma) \chi_{\frac{1}{2} m'_\sigma}(\sigma) = \delta_{m_\sigma m'_\sigma},$$

$$\sum_{m_\sigma} \chi_{\frac{1}{2}, m_\sigma}(\sigma) \chi^*_{\frac{1}{2} m_\sigma}(\sigma') = \delta_{\sigma \sigma'}.$$

Before dealing with the modified BSE's spin dependence, we should define a few practical quantities that simplify the calculations. First of all, we define a basis in the spin space of the electron(σ_1)-hole(σ_2) pair, denoted as $\zeta_\alpha(\sigma_1 \sigma_2)$, where α takes values in the set $\{1, 2, 3, 4\}$, labelling the four independent pair states $\zeta_\alpha(\sigma_1 \sigma_2) := \langle \sigma_1 \sigma_2 | \alpha \rangle$:

$$\begin{aligned}
\zeta_1(\sigma_1 \sigma_2) &= \tfrac{1}{\sqrt{2}} \left[\chi_{\frac{1}{2} \frac{1}{2}}(\sigma_1) \chi^*_{\frac{1}{2} \frac{1}{2}}(\sigma_2) + \chi_{\frac{1}{2} -\frac{1}{2}}(\sigma_1) \chi^*_{\frac{1}{2} -\frac{1}{2}}(\sigma_2) \right] && (S = 0, M_S = 0) \\
\zeta_2(\sigma_1 \sigma_2) &= \left[\chi_{\frac{1}{2} \frac{1}{2}}(\sigma_1) \chi^*_{\frac{1}{2} -\frac{1}{2}}(\sigma_2) \right] && (S = 1, M_S = 1) \\
\zeta_3(\sigma_1 \sigma_2) &= \tfrac{1}{\sqrt{2}} \left[\chi_{\frac{1}{2} \frac{1}{2}}(\sigma_1) \chi^*_{\frac{1}{2} \frac{1}{2}}(\sigma_2) - \chi_{\frac{1}{2} -\frac{1}{2}}(\sigma_1) \chi^*_{\frac{1}{2} -\frac{1}{2}}(\sigma_2) \right] && (S = 1, M_S = 0) \\
\zeta_4(\sigma_1 \sigma_2) &= \left[\chi_{\frac{1}{2} -\frac{1}{2}}(\sigma_1) \chi^*_{\frac{1}{2} \frac{1}{2}}(\sigma_2) \right] && (S = 1, M_S = -1).
\end{aligned} \tag{4.4}$$

By means of equation (4.4), we see that the pair states are eigenstates of the total spin operator, with quantum numbers $S = 0$ (singlet), 1 (triplet) and magnetic quantum numbers $-S \leqslant M_S \leqslant S$. In order to later transform the BSE equation appropriately, we need the properties of the pair states:

1. They form a basis, satisfying:

$$\sum_{\sigma_1\sigma_2} \zeta_\alpha^*(\sigma_1\sigma_2)\zeta_\beta^*(\sigma_1\sigma_2) = \delta_{\alpha\beta},$$

$$\sum_\alpha \zeta_\alpha^*(\sigma_1\sigma_2)\zeta_\alpha^*(\sigma_{1'}\sigma_{2'}) = \delta_{\sigma_1\sigma_{1'}}\delta_{\sigma_2\sigma_{2'}}.$$

2. Since only singlet states ($\alpha = 1$) possess a non-vanishing trace, it holds [23, 68]:

$$\sum_\sigma \zeta_\alpha(\sigma\sigma) = \sum_\sigma \langle\sigma\sigma|\alpha\rangle = \frac{1}{\sqrt{2}}\sum_\sigma \left[\chi_{\frac{1}{2}\frac{1}{2}}(\sigma)\chi_{\frac{1}{2}\frac{1}{2}}^*(\sigma) + \chi_{\frac{1}{2}-\frac{1}{2}}(\sigma)\chi_{\frac{1}{2}-\frac{1}{2}}^*(\sigma)\right] = \sqrt{2}\delta_{\alpha 1}.$$

With these properties, we can now move on and transform the modified BSE into the pair state basis:

$$\bar{L}_{\alpha\beta}(1,2,1',2') := \langle\alpha|\bar{L}(1,2,1',2')|\beta\rangle = \sum_{\sigma_1\sigma_2\sigma_{1'}\sigma_{2'}} \langle\alpha|\sigma_1\sigma_2\rangle\langle\sigma_1\sigma_2|\bar{L}(1,2,1',2')|\sigma_{1'}\sigma_{2'}\rangle\langle\sigma_{1'}\sigma_{2'}|\beta\rangle.$$

After a few lines of calculation involving basis transformations and taking into account that only singlet states contribute, we arrive at the following equation for the modified BSE for spin-polarized and non-spin-polarized systems:

$$\bar{L}_{\alpha\beta}(1,2,1',2') = L_{\alpha\beta}^0(1,2,1',2')$$

$$+2L_{\alpha 1}^0(1,\bar{3},1',\bar{3})\bar{v}_c(\bar{3},\bar{4})\bar{L}_{1\beta}(\bar{4},2,\bar{4},2') - \sum_\gamma L_{\alpha\gamma}^0(1,\bar{4},1',\bar{3})W(\bar{3},\bar{4})\bar{L}_{\gamma\beta}(\bar{3},2,\bar{4},2').$$

Since we do not apply a magnetic field from the outside and are not interested in this work on materials in which the spin is aligned, we will only consider the case of non-spin-polarized systems and singlet states ($\alpha = \beta = 1$). Under this assumption, the modified BSE becomes:

$$\bar{L}(1,2,1',2') = L_0(1,2,1',2') + L_0(1,\bar{4},1',\bar{3})\left[2\bar{v}_c(\bar{3},\bar{5})\delta(\bar{3},\bar{4})\delta(\bar{5},\bar{6}) - W(\bar{3},\bar{4})\delta(\bar{3},\bar{5})\delta(\bar{4},\bar{6})\right]\bar{L}(\bar{5},2,\bar{6},2').$$
$$(4.5)$$

We see that under the assumptions made above, in the case of negligible spin-orbit coupling, we only need to multiply a factor of two in front of the electron-hole exchange term to account for the spin contribution. If we look at the modified dielectric susceptibility, the whole thing behaves very similarly:

$$\bar{\chi}(1,2) = \sum_{\sigma_1,\sigma_2}\sum_{\alpha,\beta} \zeta_\beta^*(\sigma_1\sigma_1)\bar{\chi}_{\alpha,\beta}(1,2)\zeta_\beta(\sigma_2\sigma_2) = \sum_{\sigma_1,\sigma_2} \zeta_1^*(\sigma_1\sigma_1)\bar{\chi}_{1,1}(1,2)\zeta_1(\sigma_2\sigma_2) = 2\bar{\chi}_{1,1}(1,2),$$

where $\bar{\chi}_{\alpha,\beta}(1,2) = \bar{\chi}_{\alpha,\beta}(1,2,1^+,2^+)$.

If we now consider the case in which spin-orbit coupling cannot be neglected, the classification of single-particle states into spin-up and spin-down states becomes inadequate. Consequently, the expressions mentioned earlier are no longer applicable. Therefore, due to the significant influence of the spin-orbit interaction, the mixing of singlet and triplet configurations arises. Therefore, in this case, we would have to forego a reduction, and the matter becomes more complicated [23, 37, 67]. Since we do not consider materials like transition-metal dichalcogenides (TMDs) [69, 70] or GaAs [35] in this work, which are materials with strong spin-orbit coupling, we will not discuss the more general case here. For this reason, we assume for the rest of the work that spin-orbit coupling does not play a role in our discussions and we are only interested in singlet configurations. As a consequence, we can neglect any spin dependency in the following and simply write a factor 2 in front of the electron-hole exchange component in the modified BSE.

4.3   The Hamiltonian formalism for the Bethe Salpeter equation in GW-approximation

This section aims to transform the complicated 4-point function $\bar{L}$ in the spin-singlet configuration (Eq. (4.5)) into a more straightforward matrix representation, a so-called two-particle equation [39, 54, 64, 66, 71]. For this purpose, we will make a basis representation of the 4-point function using the orthogonal orbitals $\Psi_{n_i}(\boldsymbol{r})$, of independent particles, given by the orthonormal and complete set of Bloch functions $\Psi_{n_i}(\boldsymbol{r}) := \frac{1}{\sqrt{\Omega_v}} u_{n_i}(\boldsymbol{r}) e^{i\boldsymbol{k}_i \boldsymbol{r}}$, where $\Omega_v = N_k V_0$ is the volume of the crystal, N_k the number of k-points and the indices $n_i = (\tilde{n}_i, \boldsymbol{k}_i)$ are defined as the merger of the band quantum number $\tilde{n}_i$ and a reciprocal vector $\boldsymbol{k}_i$ in the first Brillouin zone.

The first step towards our goal is to perform a transformation into Blochspace, which is nothing else than a usual basistransformation in quantum mechanics for $\bar{L}(1, 1', 2, 2')$, which depends on 4 variables [46]. However, since we assume that W is static, our problem is simplified to two time-variables $\bar{L}(1, 1', 2, 2') \rightarrow \bar{L}(r_1 t_1; r_{1'} t_1; r_2 t_2; r_{2'} t_2) = \bar{L}(r_1\, r_{1'}\, r_2\, r_{2'}; t_1 - t_2)$ instead of four. This leads to a spectral representation in the Bloch base of $\bar{L}$ as [59, 64]:

$$\bar{L}(1, 1', 2, 2') = \sum_{n_1, n_{1'}, n_2, n_{2'}} \int \frac{d\omega}{2\pi} \Psi_{n_1}^*(\boldsymbol{r}_1) \Psi_{n_{1'}}(\boldsymbol{r}_{1'}) \Psi_{n_2}(\boldsymbol{r}_2) \Psi_{n_{2'}}^*(\boldsymbol{r}_{2'}) \bar{L}_{(n_1 n_{1'}),(n_2 n_{2'})}(\omega) \exp(-i\omega(t_1 - t_2)),$$
(4.6)

with

$$\bar{L}_{(n_1 n_{1'}),(n_2 n_{2'})}(\omega) = \int d\boldsymbol{r}_1 d\boldsymbol{r}_{1'} d\boldsymbol{r}_2 d\boldsymbol{r}_{2'} \, \Psi_{n_1}(\boldsymbol{r}_1) \Psi_{n_{1'}}^*(\boldsymbol{r}_{1'}) \Psi_{n_2}^*(\boldsymbol{r}_2) \Psi_{n_{2'}}(\boldsymbol{r}_{2'}) \bar{L}(r_1\, r_{1'}\, r_2\, r_{2'}, \omega). \quad (4.7)$$

Furthermore, with equation (4.7), we could transform the whole BSE in static GW-approximation (Eq. (3.40)) into Bloch space [39]:

$$\bar{L}_{(n_1 n_{1'}),(n_2 n_{2'})}(\omega) = L^0_{(n_1 n_{1'}),(n_2 n_{2'})}(\omega) + L^0_{(n_1 n_{1'}),(n_3 n_4)}(\omega) \Xi^{AGW}_{(n_3 n_4),(n_5 n_6)} \bar{L}_{(n_5 n_6),(n_2 n_{2'})}(\omega). \quad (4.8)$$

Until here, we have not gained much with this presentation alone, but we can now make use of two characteristics:

1. The formal inversion of the 4-point equation, which gives us $\bar{L}$:

$$\bar{L} = \left[1 - L^0 \Xi^{AGW} \right]^{-1} \cdot L^0. \tag{4.9}$$

2. The fact that since we have used the wavefunctions of independent particles for the basis transformation and L_0 is in practice constructed with the independent-particle wavefunctions (discussion in Ch. 4.4), it follows that L_0 is diagonal in this representation:

$$L^0_{(n_1 n_{1'}),(n_2 n_{2'})}(\omega) = \frac{\left(f_{n_{1'}} - f_{n_1} \right) \delta_{n_1, n_2} \delta_{n_{1'}, n_{2'}}}{E^0_{n_{1'}} - E^0_{n_1} - \omega}, \tag{4.10}$$

where $E^0_{n_i}$ represent the quasi-particle energies, which we will discuss in section 4.4, and f_{n_i} the occupation numbers.

These two properties allows us to determine $\bar{L}$ as [39, 54]:

$$\bar{L}_{(n_1 n_{1'}),(n_2 n_{2'})}(\omega) = \left[\underline{\underline{H}}^{2p} - \omega \underline{\underline{1}} \right]^{-1}_{(n_1 n_{1'}),(n_2 n_{2'})} \left(E^0_{n_2} - E^0_{n_1} - \omega \right) \cdot \left(f_{n_2} - f_{n_{2'}} \right) \frac{\delta_{n_1 n_2} \delta_{n_{1'} n_{2'}}}{\left(E^0_{n_2} - E^0_{n_1} - \omega \right)}$$

$$= \left[\underline{\underline{H}}^{2p} - \omega\underline{\underline{1}}\right]^{-1}_{(n_1 n_{1'}),(n_2 n_{2'})} \left(f_{n_2} - f_{n_{2'}}\right),$$

with the components of the two-particle Hamiltonian $\underline{\underline{H}}^{2p}$ defined as:

$$H^{2p}_{(n_1 n_{1'}),(n_2 n_{2'})} := (E^0_{n_{1'}} - E^0_{n_1})\delta_{n_1 n_2}\delta_{n_{1'} n_{2'}} + (f_{n_1} - f_{n_{1'}})\Xi^{AGW}_{(n_1 n_{1'}),(n_2 n_{2'})}. \tag{4.11}$$

The equation shows two peculiarities: first $\bar{L}$ is determined by a matrix inversion of a Hamiltonian $\underline{\underline{H}}^{2p}$ minus a frequency component, and second, $\bar{L}$ depends on the difference of two occupation numbers $\left(f_{n_2} - f_{n_{2'}}\right)$. In order to avoid unnecessary quantities in the definition of the Hamiltonian, we will first turn to the difference of the occupation numbers.

Since we are working at zero temperature in a non-metal ($f_{n_i} = 0$ for unoccupied states or $f_{n_i} = 1$ for occupied states), we obtain for the differences between the occupation numbers

$$\left(f_{n_2} - f_{n_{2'}}\right) = \begin{cases} 0 & \text{if } f_{n_2} = f_{n_{2'}} \\ \pm 1 & \text{if } f_{n_2} \neq f_{n_{2'}}, \end{cases}$$

that nonzero contributions are only given by the combinations $(n_2, n_{2'}) = \{v, c\}$ or $(n_2, n_{2'}) = \{v, c\}$, where c represents the conduction bands and v the valence bands.

Furthermore, since our primary interest in this work applies to absorption spectroscopy $\boldsymbol{q} \to 0$ at $T = 0K$, we only take into account direct transitions and therefore neglect any momentum transfer on the resulting electron-hole pair. Under this restriction the two-particle Hamiltonian (Eq. (4.11)), which in general depends on four $\boldsymbol{k}$-vectors, will depend only on two $\boldsymbol{k}$-vectors, which we call $\boldsymbol{k}$ and $\boldsymbol{k}'$ [23, 39, 54]:

$$\underline{\underline{H}}^{2p} \to \underline{\underline{H}}^{2p,exc} = \begin{pmatrix} \overset{\{v'c'\boldsymbol{k}'\}}{H^{2p,reso}_{\mu\mu'}} & \overset{\{c'v'\boldsymbol{k}'\}}{\Xi^{AGW}_{\mu\mu'}} \\ -\left[\Xi^{AGW}_{\mu\mu'}\right]^* & -\left[H^{2p,reso}_{\mu\mu'}\right]^* \end{pmatrix} \begin{matrix} \{vc\boldsymbol{k}\} \\ \{cv\boldsymbol{k}\} \end{matrix}. \tag{4.12}$$

The two-particle Hamiltonian represents a non hermitian, but pseudo-hermitian matrix of the dimension $N = 2 \cdot (N_v N_c N_k)$, with the resonant Hamiltonian defined as:

$$H^{2p,reso}_{\mu\mu'} := \left(E^0_{c\boldsymbol{k}} - E^0_{v\boldsymbol{k}}\right)\delta_{v,v'}\delta_{c,c'}\delta_{\boldsymbol{k},\boldsymbol{k}'} + \Xi^{AGW}_{\mu\mu'} =: H^0_{\mu\mu'} + \Xi^{AGW}_{\mu\mu'}, \tag{4.13}$$

where the first part defines the independent-particle transition energies, and the second term denotes the interaction part, which is given as [39, 54]:

$$\Xi^{AGW}_{\mu\mu'} = \Xi^{AGW}_{(vc\boldsymbol{k})(v'c'\boldsymbol{k}')} = -\int d\boldsymbol{r}d\boldsymbol{r}'\, \Psi^*_{c,\boldsymbol{k}}(\boldsymbol{r})\Psi_{c',\boldsymbol{k}'}(\boldsymbol{r})W(\boldsymbol{r},\boldsymbol{r}')\Psi_{v,\boldsymbol{k}}(\boldsymbol{r}')\Psi^*_{v',\boldsymbol{k}'}(\boldsymbol{r}')$$

$$+ 2\int d\boldsymbol{r}d\boldsymbol{r}'\, \Psi^*_{c,\boldsymbol{k}}(\boldsymbol{r})\Psi_{v,\boldsymbol{k}}(\boldsymbol{r})\bar{v}_c(\boldsymbol{r},\boldsymbol{r}')\Psi_{c',\boldsymbol{k}'}(\boldsymbol{r}')\Psi^*_{v',\boldsymbol{k}'}(\boldsymbol{r}'). \tag{4.14}$$

A practical and conventional representation of the interaction kernel, which will be relevant for later applications, is its expression in Fourier space

$$\Xi^{AGW}_{\mu\mu'} := \Xi^{AGW}_{(vc\boldsymbol{k})(v'c'\boldsymbol{k}')} = -\frac{1}{\Omega_v}\sum_{\boldsymbol{G},\boldsymbol{G}'} \tilde{\rho}^*_{(c,\boldsymbol{k})(c',\boldsymbol{k}')}(\boldsymbol{q}+\boldsymbol{G})W_{\boldsymbol{G},\boldsymbol{G}'}(\boldsymbol{q})\tilde{\rho}_{(v',\boldsymbol{k}')(v,\boldsymbol{k})}(\boldsymbol{q}+\boldsymbol{G}')\delta_{\boldsymbol{q},\boldsymbol{k}-\boldsymbol{k}'}$$

$$+ \frac{2}{\Omega_v}\sum_{\boldsymbol{G}\neq\boldsymbol{0}} \tilde{\rho}^*_{(c,\boldsymbol{k})(v,\boldsymbol{k})}(\boldsymbol{G})\bar{v}_c(\boldsymbol{G})\tilde{\rho}_{(v',\boldsymbol{k}')(c',\boldsymbol{k}')}(\boldsymbol{G}), \tag{4.15}$$

with the representation of the transition matrix elements as:

$$\rho_{(i,\boldsymbol{k})(j,\boldsymbol{k}')}(\boldsymbol{q}+\boldsymbol{G}) = \langle \Psi_{i,\boldsymbol{k}}|e^{-i(\boldsymbol{q}+\boldsymbol{G})\boldsymbol{r}}|\Psi_{j,\boldsymbol{k}'}\rangle = \langle u_{i,\boldsymbol{k}}|e^{-i(\boldsymbol{q}+\boldsymbol{G})\boldsymbol{r}}|u_{j,\boldsymbol{k}'}\rangle\delta_{\boldsymbol{k}',\boldsymbol{k}-\boldsymbol{q}}$$

$=: \delta_{\boldsymbol{k}',\boldsymbol{k}-\boldsymbol{q}}\tilde{\rho}_{(i,\boldsymbol{k})(j,\boldsymbol{k}')}(\boldsymbol{q}+\boldsymbol{G}).$

Now that we have defined the two-particle Hamiltonian in the case of absorption spectroscopy and found a way to use it to determine $\bar{L}$, the question arises: *What is the physical significance of this Hamiltonian?*

The most important observation is that with this Hamiltonian and the structure for

$\bar{L}^{AGW}_{(n_1 n_{1'}),(n_2 n_{2'})}(\omega) = \left[\underline{\underline{H}}^{2p} - \omega\underline{\underline{1}}\right]^{-1}_{(n_1 n_{1'}),(n_2 n_{2'})}\left(f_{n_2} - f_{n_{2'}}\right)$, we must cover all the properties that we find in $\bar{\chi}$, which means in particular that this equation must already contain the resonant and the anti-resonant part. This property is why we find the term $H^{2p,reso}_{\mu\mu'}$ twice on the main diagonal but with the opposite sign. The contribution at the upper left of the Hamiltonian (Eq. (4.12)) represents the resonant term $H^{2p,reso}_{\mu\mu'}$, as it corresponds to transitions at positive absorption frequencies ω ($v \rightarrow c$). Conversely, the term in the lower right gives the same contribution as the upper left but at negative transition energies ($c \rightarrow v$), which is why it represents the anti-resonant term.

Staying with the contribution of the diagonal blocks $H^{2p,reso}_{\mu\mu'}$, we should understand their meaning. For this purpose, we first consider the independent-particle transition energies $H^0_{\mu\mu'} = \left(E^0_{c\boldsymbol{k}} - E^0_{v\boldsymbol{k}}\right)\delta_{v,v'}\delta_{c,c'}\delta_{\boldsymbol{k},\boldsymbol{k}'}$. As its name suggests, this term describes the energy difference between the conduction bands and the valence bands at the same $\boldsymbol{k}$ point. If only this component were present, the spectrum would be a sum of transitions between occupied and empty quasi-particle states.

If we now consider only the interaction part $\Xi^{AGW}_{\mu\mu'}$ of the resonant contribution, then assigning a physical meaning to this is easy. It contains all the interactions we consider in our system: the classical Hartree part (electron-hole exchange-term) and the exchange-correlation part (direct screened electron-hole Coulomb interaction).

The somewhat more complicated part is the off-diagonal blocks of the Hamiltonian (Eq. (4.12)). These terms are called coupling terms, and to get a feel for them, we will first look at what it means to neglect them. In this case we would have a block matrix containing the two decoupled systems of resonant- (positive energies) and anti-resonant- (negative energies) terms $\underline{\underline{H}}^{2p,exc} \rightarrow \underline{\underline{H}}^{2p,reso} \oplus \left[-\underline{\underline{H}}^{2p,reso}\right]^*$. This means, firstly, that the Hamiltonian becomes hermitian and its eigenvalues are real-valued and, secondly, that the information of the two systems is separate from each other and is not mixed. Suppose we now take into account the off-diagonal blocks. In that case, consequently, a mixing of the information between positive and negative energies occurs, and the Hamiltonian would become pseudo-hermitian.

At this point, it is rather tedious to assign a physical meaning to this difference between purely hermitian and pseudo-hermitian matrices. However, we will reach this point when we make an essential approximation to the Hamiltonian (Ch. 4.5.1).

After the brief discussion about the meaning of the individual elements of the Hamiltonian, we will get down to our real question: *How can we use the Hamiltonian structure to obtain the modified dielectric susceptibility?* For this purpose, we can now make use of the structure for $\bar{L}$, which allows us to use all the knowledge of linear algebra in order to simplify our task.

Since in order to solve the problem, we now have the problem of inverting a large matrix, which would be highly complex, we use instead the standard spectral representation of the inverse Hamil-

tonian for a pseudo-hermitian matrix in component notation:

$$\left[\underline{\underline{H}}^{2p,exc} - \omega\underline{\underline{1}}\right]^{-1}_{\mu\mu'} = \sum_{\lambda\lambda'} \frac{\bar{A}^{\mu}_{\lambda} S^{-1}_{\lambda\lambda'}\bar{A}^{*\mu'}_{\lambda'}}{E_\lambda - \omega}, \; S^{-1}_{\lambda\lambda'} = \sum_{\mu}\bar{A}^{*\mu}_{\lambda}\bar{A}^{\mu}_{\lambda}, \tag{4.16}$$

where $\underline{\bar{A}}_\lambda$ are the normalized eigenfunctions, $\underline{S}$ is an overlap matrix, and E_λ represent the eigenvalues, respectively, obtained from the eigenvalue equation:

$$\sum_{\mu'} H^{2p,exc}_{\mu\mu'} A^{\mu'}_{\lambda} = E_\lambda A^{\mu}_{\lambda}, \; E_\lambda \in \mathbb{R}. \tag{4.17}$$

The next step is to use the connection between $\bar{\chi}$ and $\bar{L}$, namely $\bar{\chi}(1,2) = \bar{L}(1,2,1^+,2^+)$. In order to obtain $\bar{L}(1,2,1^+,2^+)$, we have to replug equation (4.16) into the expression for $\bar{L}$, and then into the expression (4.8). The last two steps are taking the diagonal limit $\bar{L}(1,2,1',2') \to \bar{L}(1,2,1^+,2^+)$ and performing a Fourier transformation, which leads finally to the modified dielectric susceptibility in adiabatic GW-approximation

$$\lim_{\boldsymbol{q}\to 0}\bar{\chi}_{\boldsymbol{G},\boldsymbol{G}'}(\boldsymbol{q},\omega) = \lim_{\boldsymbol{q}\to 0}\lim_{\Gamma\to 0}\sum_{\lambda\lambda'}\sum_{\mu\mu'}\tilde{\rho}^*_\mu(\boldsymbol{q}+\boldsymbol{G})\frac{\bar{A}^{(\mu)}_\lambda S^{-1}_{\lambda\lambda'}\bar{A}^{*(\mu')}_{\lambda'}}{E_\lambda - (\omega + i\Gamma)}\tilde{\rho}_{\mu'}(\boldsymbol{q}+\boldsymbol{G}')f_{\mu'} =$$

$$= \lim_{\boldsymbol{q}\to 0}\lim_{\Gamma\to 0}\sum_{\mu\mu'}\tilde{\rho}^*_\mu(\boldsymbol{q}+\boldsymbol{G})\left[\underline{\underline{H}}^{2p,exc} - (\omega + i\Gamma)\underline{\underline{1}}\right]^{-1}_{\mu\mu'}\tilde{\rho}_{\mu'}(\boldsymbol{q}+\boldsymbol{G}')f_{\mu'}, \tag{4.18}$$

with $f_\mu := f_{n_2} - f_{n_1}$. Equation (4.18) represents the general expression for the modified dielectric susceptibility in the static GW approximation for the pseudo-hermitian Hamiltonian described in equation (4.12).

Thanks to this equation, we now know that we do not have to solve the BSE in static GW approximation as a 4-point equation but can solve the eigenvalue problem from equation (4.17) and use the obtained eigenvalues and eigenvectors to construct $\bar{\chi}$. However, it should be noted that a very similar procedure to the one we used to obtain $\bar{\chi}$ can also be used to determine χ. However, this is not part of this work and will not be discussed further.

In principle, we now have the most important information to calculate absorption spectra, but there is one point we need to clarify, namely: *Where do we get the quasi-particle energies $E^0_{n_i}$ and the associated wavefunctions Ψ_{n_i}?*

4.4   Ground state calculations

In general, to obtain the quantities Ψ_{n_i} and $E^0_{n_i}$ we are looking for, we would need to perform a full self-consistent calculation [38, 39, 46, 66]. We can illustrate this by considering that L_0 is built from Green's functions, which by using the static version of equation (3.19) can be expressed as [38]:

$$G(\boldsymbol{r}_1,\boldsymbol{r}_2,z) = \sum_\xi \frac{\Psi_\xi(\boldsymbol{r}_1,z)\Psi^*_\xi(\boldsymbol{r}_2,z)}{z - E_\xi(z)}.$$

With this assumption, one could find the formal solution of equation (3.25) for the dominant contribution to the poles $E^0_{n_i} = E_\xi(E^0_{n_i})$ leading to an eigenvalue equation [38]:

$$\left[h_0(\boldsymbol{r}) + v_H(\boldsymbol{r})\right]\Psi_{n_i}(\boldsymbol{r}) + \int d\boldsymbol{r}_2\,\Sigma_{xc}(\boldsymbol{r},\boldsymbol{r}_2,E^0_{n_i})\Psi_{n_i}(\boldsymbol{r}_2) = E^0_{n_i}\Psi_{n_i}(\boldsymbol{r}). \tag{4.19}$$

It is important to note that the eigenvalue equation for the so-called quasi-particle (QP) states Ψ_{n_i} and energies $E_{n_i}^0$ represents a single-particle equation, which does not correspond to the stationary one-particle eigenstates of the many-body electron Hamiltonian (Eq. (3.1)). However, they are formed by a macroscopically large number of degenerate, stationary eigenstates of the many-body electron Hamiltonian.

The problem with solving the eigenvalue equation above, as we already know, is that the exchange-correlation part is a functional of the Green's function (Eq. (3.26)). This means that the problem requires a self-consistent solution, which makes the whole thing numerically very expensive.

The crucial point for us, however, is that we need precisely these wave functions and energies to complete the Bloch representation of the modified BSE and thus calculate $\bar{\chi}$ (Eq. (4.18)). Since it is numerically very complex to solve the problem, it may be advisable to take a different route, namely that of density-functional theory (DFT), which, as we will see in the following, holds a similar equation as (4.19).

In order to avoid full self-consistency, we ask ourselves the question: *What is DFT and how can we use it to get the wavefunctions Ψ_{n_i} and the associated eigenenergies $E_{n_i}^0$, of the quasi-particle equation?*

4.4.1   Density-functional theory

DFT is well-documented in various references such as [23,46,72], as well as in PhD theses [38,39,56]. However, we focus here on presenting the fundamental concepts and key equations instead of showing lengthy calculations.

The central problem addressed by DFT is the many-body electron Hamiltonian, denoted as (Eq. (3.1)):

$$\hat{H}^{eMB} = \hat{T}_{kin} + \hat{V}_{ee} + \hat{V}_{ext}, \tag{4.20}$$

where the external potential includes only the electron-nuclei interaction $\hat{V}_{eN}$. As we already know, even for the simplified Hamiltonian, we have to consider about 10^{23} electrons, which makes an exact solution of the eigenvalues and eigenfunctions using classical storage impossible.

Having reached such a point, the exact solution to the problem will not be possible, so we must reflect and ask ourselves: *What information do we want to know?* In the case of Kohn, Hohenberg and Sham [73,74], the answer was the ground-state density and ground-state energy of the many-body electron Hamiltonian. In the following, we will follow their ideas. For this purpose, we will start with two theorems from Hohenberg and Kohn, which we formulate as follows [39]:

Theorem 1: The ground-state expectation value of any physical observable of a many-electron system is a unique functional of the electron density $n(r)$:

$$\langle \psi^0 | \hat{O} | \psi^0 \rangle = O[n].$$

Theorem 2: The total energy defines a universal functional that possesses a minimum, corresponding to the ground state energy E_0 and the ground state density n_0:

$$E_{v_{ext}}[n] = \langle \psi | \hat{T} + \hat{V}_{ee} + \hat{V}_{ext} | \psi \rangle$$

$$= \langle \psi | \hat{T} + \hat{V}_{ee} | \psi \rangle + \int d\boldsymbol{r} \, v_{ext}(\boldsymbol{r}) n(\boldsymbol{r}) = F_{HK}[n] + \int d\boldsymbol{r} \, v_{ext}(\boldsymbol{r}) n(\boldsymbol{r}),$$

where the universal functional (independent of the external potential) is defined as $F_{HK}[n] :=$ $\langle \psi | \hat{T} + \hat{V}_{ee} | \psi \rangle$. The minimization problem can be expressed as:

$$\min_{\psi \in \Psi} \langle \psi | \hat{T} + \hat{V}_{ee} + \hat{V}_{ext} | \psi \rangle = \langle \psi^0 | \hat{H}^{eMB} | \psi^0 \rangle = \min_{n \in N} E_{v_{ext}}[n],$$

where, Ψ represents the space of ground state wave functions, and N is the ensemble of electron densities.

So we see by theorem 2 that if we knew the universal functional $F_{HK}[n]$ of the ground state density, then we would also know the ground state energy of the system. The complicated part is to find the universal functional because, as formulated in theorem 2, we would still have to solve the many-particle problem. At this point comes the ingenious idea of Kohn and Sham, which was to replace the complex interacting system with a non-interacting auxiliary system that produces the same ground-state energy and density [72,74]. We will refer to this idea in the following as the Kohn-Sham (KS) approach.

The Kohn-Sham approach relies on two key assumptions:

1. **Non interacting V-representability [75]:** The exact ground state density can be represented by the ground state density of an auxiliary system of non-interacting particles.

2. **The auxiliary Hamiltonian** is defined as a non-interacting system, depending on the kinetic energy of one electron and an unspecified effective local potential $v_{eff}(\boldsymbol{r})$:

$$\hat{H}_{\text{aux}} = -\frac{1}{2} \nabla^2 + v_{eff}(\boldsymbol{r}). \tag{4.21}$$

Dealing with this effective non-interacting Hamiltonian, the static Schroedinger equation takes the form:

$$\left(-\frac{1}{2} \nabla^2 + v_{eff}(\boldsymbol{r}) \right) \phi_i(\boldsymbol{r}) = \epsilon_i \phi_i(\boldsymbol{r}),$$

where, ϵ_i represents the eigenvalues, and $\phi_i(\boldsymbol{r})$ represents the eigenstates of the non-interacting auxiliary system. The density of the auxiliary system is given as:

$$n(\boldsymbol{r}) = \sum_{i=1}^{occ} |\phi_i(\boldsymbol{r})|^2. \tag{4.22}$$

So, we see that according to the V-representability, we can choose the effective potential of the auxiliary system in such a way that we get the exact ground state density. So the crucial question is: *Which form takes the effective potential $v_{eff}(\boldsymbol{r})$?*

To answer this question, we need to derive an effective equation for the non-interacting particles based on the form of the ground state energy presented in theorem 2. For this purpose, we start with the definition of the universal functional and the fact that the ground-state energy of the system matches that of the auxiliary system:

$$\min_{n \in N} E_{v_{ext}}[n] = \min_{n \in N} E_{v_{eff}}[n]. \tag{4.23}$$

The next step is to rewrite the universal functional $F_{HK}[n] := \langle \psi | \hat{T} + \hat{V}_{ee} | \psi \rangle$ as follows:

$$\begin{aligned} F_{HK}[n] &= \langle \psi | \hat{T} | \psi \rangle - T_s + T_s + \langle \psi | \hat{V}_{ee} | \psi \rangle - E_H[n] + E_H[n] \\ &= T_s + E_H[n] + \left(\langle \psi | \hat{T} | \psi \rangle - T_s + \langle \psi | \hat{V}_{ee} | \psi \rangle - E_H[n] \right), \end{aligned} \tag{4.24}$$

where T_s represents the independent-particle kinetic energy, and $E_H[n]$ is the Hartree energy, the classical mean-field energy arising from densities. The remaining part will be defined in one quantity E_{xc}, which is defined as:

$$E_{xc}[n] := \left(\langle \psi | \hat{T} | \psi \rangle - T_s + \langle \psi | \hat{V}_{ee} | \psi \rangle - E_H[n] \right).$$

E_{xc} encapsulates the many-body effects of exchange and correlation, as it subtracts all classical information, including the independent-particle kinetic energy and the mean-field electron-electron interaction. This transformation allows us to define without loss of generality the ground-state energy of the system as:

$$E_{v_{eff}}[n] := T_s + E_H[n] + E_{xc}[n] + \int d\boldsymbol{r}\, v_{ext}(\boldsymbol{r})n(\boldsymbol{r}). \tag{4.25}$$

The next step will be now to derive a Kohn-Sham Schroedinger equation from the total energy. To do so, we employ the Lagrange multiplier method with the orthonormality constraint $\langle \phi_i | \phi_j \rangle = \delta_{ij}$ on the auxiliary eigenstates:

$$\delta \left[\langle \phi_i | \hat{H}_{\text{aux}} | \phi_i \rangle - \epsilon_i \left(\langle \phi_i | \phi_i \rangle - 1 \right) \right] = \left[\frac{\delta E_{v_{eff}}}{\delta \phi_i^*} - \epsilon_i \phi_i \right] = 0.$$

Building on this equation, after a few lines of calculation, we obtain the equation

$$H_{KS} \phi_i^{KS}(\boldsymbol{r}) = \epsilon_i \phi_i^{KS}(\boldsymbol{r}), \tag{4.26}$$

where the Kohn-Sham Hamiltonian is defined as:

$$H_{KS}(\boldsymbol{r}) = -\frac{1}{2}\nabla^2 + \frac{\delta E_H[n]}{\delta n}\bigg|_{n(\boldsymbol{r})} + \frac{\delta E_{xc}[n]}{\delta n}\bigg|_{n(\boldsymbol{r})} + v_{ext}(\boldsymbol{r}) =: -\frac{1}{2}\nabla^2 + v_H(\boldsymbol{r}) + v_{xc}(\boldsymbol{r}) + v_{ext}(\boldsymbol{r}).$$
$$\tag{4.27}$$

Therefore, our effective potential, as defined in the previous equation (4.21), is:

$$v_{eff}(\boldsymbol{r}) := v_H(\boldsymbol{r}) + v_{xc}(\boldsymbol{r}) + v_{ext}(\boldsymbol{r}).$$

We have now elucidated how we transition from the many-body Hamiltonian to an auxiliary system of non-interacting particles, the KS system. However, it is essential to emphasize that while the Kohn-Sham system is designed to provide the exact ground-state density and energy, it is not constructed to yield exact results for all energies, band structures, wave function, or other quantities, such as bandgaps in semiconductors or insulators.

Nevertheless, the topic is precisely this: we want the KS wave functions ϕ_i^{KS} and the associated eigenvalues ϵ_i as an approximation for the quasi-particle wave functions Ψ_{n_i} and energies $E_{n_i}^0$.

So the crucial question is: *How good would these approximations* $\Psi_{n_i} \approx \phi_{n_i}^{KS}$, $E_{n_i}^0 \approx \epsilon_{n_i}$ *be and where do we make a mistake?*

4.4.2   The GW correction

As discussed in the last section, the KS scheme in DFT is primarily designed to provide accurate ground-state densities and energies. However, it is not inherently suited for calculating other properties, such as quasi-particle eigenfunctions and energies. However, we want to use them

instead of solving equation (4.19). Therefore, the first question in this context is: *What mistake are we making regarding the quasi-particle energies?*

For this purpose, we will start by analyzing what we measure in a photoemission experiment, namely the fundamental gap (quasi-particle gap) of an N-electron system, defined as:

$$E_{gap}^{QP} = \left[E_{N+1}^0 - E_N^0 \right] - \left[E_N^0 - E_{N-1}^0 \right], \tag{4.28}$$

where the first term in brackets represents the energy of the lowest conduction band when an electron is added (N+1 electrons), and the second term in brackets corresponds to the energy of the highest valence band when one electron is removed (N-1 electrons) [46]. On the other hand, with the KS scheme, we could define a similar relation [46]:

$$E_{gap}^{QP} = \epsilon_{N+1}(N+1) - \epsilon_N(N), \tag{4.29}$$

where we consider the energy difference between the KS system of $N+1$ electrons with the ground-state energy $\epsilon_{N+1}(N+1)$ and the KS-system of N electrons with the ground-state energy $\epsilon_N(N)$. However the fundamental gap E_{gap}^{QP} is not the KS eigenvalue gap [46]

$$E_{gap}^{KS} := \epsilon_{N+1}(N) - \epsilon_N(N), \tag{4.30}$$

which we would obtain from the difference between the energies of the lowest unoccupied molecular orbital (LUMO) $\epsilon_{N+1}(N)$ and the highest occupied molecular orbital (HOMO) $\epsilon_N(N)$ of the N electron KS system.

A connection between the fundamental gap and the KS gap can be established by performing a zero addition $\epsilon_{N+1}(N) - \epsilon_{N+1}(N)$ and using the definition of the KS gap (Eq. (4.30)), in order to obtain the expression:

$$E_{gap}^{QP} = E_{gap}^{KS} + (\epsilon_{N+1}(N+1) - \epsilon_{N+1}(N)). \tag{4.31}$$

Equation (4.31) shows that the fundamental gap is composed of the KS gap and a correction given by the energy difference of the HOMO of the $N+1$ and the LUMO of the N KS electron system, which is called the derivative discontinuity. Of course, the question now arises: *What is the physical origin of the difference between the fundamental gap and the KS gap and what significance does it have for calculations?* The answer to this question will give us an estimation and physical explanation of the mistake we will make if we replace the quasi-particle energy with the KS energy. To understand the relationship between equation (4.30) and equation (4.29), we can utilize the definition of the total energy (Eq. (4.25)) in the KS schema [76]:

$$E = F_{HK}[n] = \sum_i f_i \epsilon_i - \left[E_H + \int d\boldsymbol{r}\, v_{xc}(\boldsymbol{r}) n(\boldsymbol{r}) - E_{xc} \right],$$

where f_i represents the occupation numbers of the KS orbitals. If we now add $(E(N+1))$ or remove $(E(N-1))$ an electron, the change in density, Δn, is on the order of $10^{-20} n$ and therefore in principle negligible. Nevertheless, the change in density leads to a finite change in the exchange-correlation potential, denoted as Δ_{xc}:

$$\Delta_{xc} = \lim_{\Delta n \to 0} v_{xc}[n + \Delta n] - v_{xc}[n - \Delta n].$$

When we relate the KS gap (Eq. (4.30)) and the quasi-particle bandgap (Eq. (4.29)), we immediately obtain that the change in density leads to an incorrect description of the gap within the KS scheme [76, 77]:

$$\lim_{\Delta n \to 0} E_{gap}^{QP} = E_{gap}^{KS} + (\epsilon_{N+1}(N+1) - \epsilon_{N+1}(N)) = E_{gap}^{KS} + \Delta_{xc}. \qquad (4.32)$$

So, the KS bandgap does not correspond to the fundamental bandgap we measure but is shifted, which, depending on the chosen functional for the exchange correlation potential, could be small or of the order of eV [78]. This observation is commonly referred to as the "band-gap problem" in DFT.

However, the final result (Eq. (4.32)) for the relation between the KS and the QP bandgap tells us that, at least in practical applications, if we use the KS energies as an approximation for the quasi-particle energies, we need to shift the bandgap and thus all the conduction bands by a constant Δ_{xc} in order to obtain an accurate description of the fundamental bandgap.

However, it is not clear at this point whether all bands are shifted by the same constant or whether the constant is not a constant but a band and $\boldsymbol{k}$ dependent quantity, which is why we have to ask the question: *How do the KS energies ϵ_i in general relate to the quasi-particle energies $E_{n_i}^0$?* For these purposes, we make a zero addition $\left[\hat{V}_{xc} - \hat{V}_{xc}\right]$ in equation (4.19) and multiply from the left by Ψ_{n_i}, obtaining:

$$\langle \Psi_{n_i} | \hat{h}_0 + \hat{V}_H + \hat{V}_{xc} | \Psi_{n_i} \rangle + \langle \Psi_{n_i} | \hat{\Sigma}_{xc}(E_{n_i}^0) - \hat{V}_{xc} | \Psi_{n_i} \rangle = E_{n_i}^0 \qquad (4.33)$$

in Dirac-notation. Next, we assume that the overlap between KS and quasi-particle wavefunctions is substantial ($\Psi_{n_i} \approx \phi_{n_i}^{KS}$), which implies that $\hat{\Sigma}_{xc}(E_{n_i}^0) - \hat{V}_{xc}$ is small compared to the rest of the Hamiltonian and corresponds therefore to a first-order perturbative approach [60]. Under this assumption, we can directly obtain the relation between the quasi-particle energies and the KS energies [38]:

$$E_{n_i}^0 = \epsilon_{n_i} + \langle \phi_{n_i}^{KS} | \hat{\Sigma}_{xc}(E_{n_i}^0) - \hat{V}_{xc} | \phi_{n_i}^{KS} \rangle, \qquad (4.34)$$

Equation (4.34) not only serves as a bandgap correction but also provides a more general energy correction. In general, however, equation (4.34) is further approximated for practical use by a linearization of the non-linear equation in $E_{n_i}^0$ by making a Taylor approximation of the self-energy [38, 39]

$$\langle \phi_{n_i}^{KS} | \hat{\Sigma}_{xc}(E_{n_i}^0) | \phi_{n_i}^{KS} \rangle \approx \langle \phi_{n_i}^{KS} | \hat{\Sigma}_{xc}(\epsilon_{n_i}) | \phi_{n_i}^{KS} \rangle + \langle \phi_{n_i}^{KS} | \frac{\delta \hat{\Sigma}_{xc}(\omega)}{\delta \omega} \Big|_{\omega = \epsilon_{n_i}} | \phi_{n_i}^{KS} \rangle \left(E_{n_i}^0 - \epsilon_{n_i} \right) + O\left((E_{n_i}^0 - \epsilon_{n_i})^2 \right),$$

in order to obtain

$$E_{n_i}^0 \approx \epsilon_{n_i} + \left(1 - \langle \phi_{n_i}^{KS} | \frac{\delta \hat{\Sigma}_{xc}(\omega)}{\delta \omega} \Big|_{\omega = \epsilon_{n_i}} | \phi_{n_i}^{KS} \rangle \right)^{-1} \langle \phi_{n_i}^{KS} | \hat{\Sigma}_{xc}(\epsilon_{n_i}) - \hat{V}_{xc} | \phi_{n_i}^{KS} \rangle. \qquad (4.35)$$

As we can see from equation (4.35), even if we can approximate the quasi-particle eigenstates with the KS eigenstates $\Psi_{n_i} \approx \phi_{n_i}^{KS}$, which is still to be discussed, we have to make a KS calculation first and then make additional effort to get the correction for the energies. Therefore, the question naturally arises: *Can we approximate the correction to the KS energies by a constant for all energies?*

At this point, we can only answer the question based on empirical findings. It is known that for simple semiconductors the assumption of a so-called scissor-operator $\langle \phi_{n_i}^{KS} | \hat{\Sigma}_{xc}(E_{n_i}^0) - \hat{V}_{xc} | \phi_{n_i}^{KS} \rangle \approx \Delta_{scissor}$ (Eq. (4.32)), which means that the correction does not depend on the band index or a specific $\boldsymbol{k}$ point, is mostly sufficiently reasonable and leads to minimal errors. The reason for the minor errors is that the KS wave functions already represent a good approximation for quasi-particle wave functions in bulk materials. One observes an overlap of 96% to 99.9% between the wave functions [79, 80]. In low-dimensional materials, however, one has to be much more careful with such approximations [81, 82].

Roughly speaking, depending on the purpose, it must be determined whether the result is sufficiently good for the need, e.g. band structure calculation or spectra calculation.

For simple semiconductors, however, as we will study them here in this work, i.e. Ge, Si, ZnS and LiF, a calculation through the KS scheme and using a scissor operator is sufficient. In the practical application, we will use the code Abinit [83] in this work in order to get the wave functions and energies from the KS scheme.

This section has shown us that we can use the KS scheme with a scissor to approximate the last missing elements, namely the quasi-particle energies and the quasi-particle eigenfunctions. Thus, we have reached the point where we can calculate $\bar{\chi}$ and thus the dielectric function.

However, before we do that, we will deal with possible approximations on the two-particle Hamiltonian. Because up to this point, we still have the problem that we are dealing with a very large matrix (Eq. (4.12)) of dimension $N = 2N_v N_c N_k$.

4.5   Approximations of the two-particle Hamiltonian

We have already seen very general approximations in section 3.4, especially the important static GW approximation of the self-energy. These approximations were especially necessary to build a functional and computable system at all. In the following however, the approximations will focus more on reducing the numerical costs, since these are still large and can be classified into two main blocks:

Firstly, the matrix dimension ($N = 2N_c N_v N_k$) of the Hamiltonian, which affects directly the numerical costs for the eigenvalue and eigenvector calculation.

Secondly, the construction costs of the matrix, which requires the setting up of the interaction kernel in reciprocal space and therefore contains two sums over the G-space (Eq. (4.15)). Further since the electron-hole exchange and the screened Coulomb part include different arrangements of indices, we have to constructe two types of transition dipole-matrix elements.

In the following, we want to deal with the first aspect more in-depth; for this, we go back to the off-diagonals of the two-particle Hamiltonian, which we have briefly mentioned before but we have not explained their physical meaning in detail. The interesting point is that if we could make further approximations, for example, neglecting the off-diagonal contributions, the matrix size would decrease. It therefore makes sense to take another close look at this contribution, which is why we ask ourselves the question: *Is it possible to neglect the coupling terms of the Hamiltonian?*

4.5.1   Tamm-Dancoff approximation

The approximation, which we discuss in this section, is one of the standard approximations in theoretical spectroscopy and is known under the name Tamm-Dancoff approximation (TDA) [84–86].

Since we do not know much about this approximation at this point, we will start by analyzing the coupling terms naively to see what we can achieve if we neglect them. We, therefore, assume that the coupling term tends to be smaller than the independent-particle transition energies $H^0 \gg \Xi$ of the two-particle Hamiltonian (Ch. 4.3). Therefore, naturally, we would assume that the contribution from the diagonal blocks given by the resonant Hamiltonian $H^{2p,reso}$ (Eq. (4.13)) is much more significant than the one from the off-diagonal one. Thus, we would assume that we can approximate the two-particle Hamiltonian as:

$$
\underline{\underline{H}}^{2p,exc} =
\begin{matrix}
\{v'c'\boldsymbol{k}'\} & \{c'v'\boldsymbol{k}'\} & \{v'c'\boldsymbol{k}'\} & \{c'v'\boldsymbol{k}'\} \\
\begin{pmatrix} H^{2p,reso}_{\mu\mu'} & \Xi^{AGW}_{\mu\mu'} \\ -\left[\Xi^{AGW}_{\mu\mu'}\right]^* & -\left[H^{2p,reso}_{\mu\mu'}\right]^* \end{pmatrix} & \approx & \begin{pmatrix} H^{2p,reso}_{\mu\mu'} & 0 \\ 0 & -\left[H^{2p,reso}_{\mu\mu'}\right]^* \end{pmatrix} & \begin{matrix} \{vc\boldsymbol{k}\} \\ \{cv\boldsymbol{k}\} \end{matrix}
\end{matrix} \quad .
\tag{4.36}
$$

This approximation would lead us further to a simplification of the eigenvalue equation (Eq. (4.17)), since the relationship $H^{2p,antires} = -[H^{2p,reso}]^*$ holds:

$$
\sum_{\mu'} H^{2p,reso}_{\mu\mu'} A^{\mu'}_\lambda = E_\lambda A^\mu_\lambda, \; E_\lambda \in \mathbb{R}
\tag{4.37}
$$

and therefore to a reduction of the dimension of the Hamiltonian from $N = 2N_v N_c N_k$ to $N = N_v N_c N_k$. This naive approach shows us that this approximation would lead to a significant reduction of the numerical cost, so we should better understand why the approximation should be valid. For this reason, we will discuss under which circumstances this approximation can work and what effects this approximation has. To make the whole thing more precise, we remember as we have deduced in section 4.3, the consideration of the coupling terms leads to a mixing of the positive (excitation energies of e-h pairs) and negative (excitation energies of e-h antipairs) energies. So the question arises: *Under which circumstances are these mixtures relevant?*

Mathematically, it can be shown by using perturbation theory that when the TDA is invalid, "the coupling strength between the resonant and antiresonant parts will be inversely proportional to the energy difference between the most prominent peaks in the resonant and antiresonant part" of the macroscopic dielectric function [65]. However, since the eigenenergies of the resonant and antiresonant terms are directly mirrored on the energy axis, this means that the greater the binding energy of an exciton, the more likely it is to be influenced by the associated energy on the antiresonant side. A similar argument applies, of course, to influences by plasmons, as we will discuss in section 4.5.4.

In summary, it can be said that the TDA for the calculation of $\bar{\chi}$ is appropriate if the following two points are fulfilled:

- There is a substantial separation between antiresonant and resonant transition energies

- An energetic gap exists between plasmonic and excitonic transitions.

Now that we have discussed the advantage and validity of the TDA, we should ask ourselves: *What influence does this approximation have on the eigenenergies, eigenvectors and the modified dielectric susceptibility $\bar{\chi}$?*

In the TDA, the two-particle Hamiltonian becomes hermitian, which leads to the following characteristics:

- Eigenvectors $\underline{A}_\lambda$ are orthogonal $\langle \bar{A}_\lambda | \bar{A}_{\lambda'} \rangle = \delta_{\lambda\lambda'}$

- The overlap matrix becomes the identity $S_{\lambda\lambda'}^{-1} = \sum_\mu \bar{A}_\lambda^{*\mu} \bar{A}_\lambda^\mu = \delta_{\lambda\lambda'}$

- The resonant and antiresonant parts of the modified dielectric susceptibility can be calculated separately and combined [54]

$$\bar{\chi}_{\boldsymbol{G},\boldsymbol{G}'}^{TDA}(\boldsymbol{q} \to 0, \omega) = \bar{\chi}_{\boldsymbol{G},\boldsymbol{G}'}^{TDA,reso}(\boldsymbol{q} \to 0, \omega) + \bar{\chi}_{\boldsymbol{G},\boldsymbol{G}'}^{TDA,reso}(\boldsymbol{q} \to 0, -\omega) \tag{4.38}$$

- The calculation of the resonant and thus also the anti-resonant part of the modified dielectric susceptibility is simplified to the expression:

$$
\begin{aligned}
\lim_{\boldsymbol{q}\to 0} \bar{\chi}_{\boldsymbol{G},\boldsymbol{G}'}^{TDA,reso}(\boldsymbol{q},\omega) &= \lim_{\boldsymbol{q}\to 0} \lim_{\Gamma\to 0} \sum_\lambda \sum_{\mu\mu'} \tilde{\rho}_\mu^*(\boldsymbol{q}+\boldsymbol{G}) \frac{\bar{A}_\lambda^{(\mu)} \bar{A}_\lambda^{*(\mu')}}{E_\lambda - (\omega + i\Gamma)} \tilde{\rho}_{\mu'}(\boldsymbol{q}+\boldsymbol{G}') f_{\mu'} \\
&= \lim_{\boldsymbol{q}\to 0} \lim_{\Gamma\to 0} \sum_{\mu\mu'} \tilde{\rho}_\mu^*(\boldsymbol{q}+\boldsymbol{G}) \left[\underline{\underline{H}}^{2p,reso} - (\omega + i\Gamma)\underline{1} \right]_{\mu\mu'}^{-1} \tilde{\rho}_{\mu'}(\boldsymbol{q}+\boldsymbol{G}') f_{\mu'}.
\end{aligned}
\tag{4.39}
$$

These results of the TDA should make us realize that, in principle, we should use this approximation where possible since it significantly reduces the numerical cost. It should be noted, however, that this approximation is only well suited for absorption spectra but not for measurements that require the real part of the dielectric function or a momentum dependent dielectric function ($\boldsymbol{q} \neq 0$). To see this, we first considering the connection between the imaginary part and the real part of the dielectric function through the Kramers-Kronig connection $\epsilon_1(\omega) = 1 + \frac{1}{\pi} P \left(\int_{-\infty}^\infty d\xi \frac{\epsilon_2(\omega)}{\xi - \omega} \right)$ [54]. In this case, both the positive and negative energies are automatically mixed together in the real part of the dielectric function, so neglecting the coupling terms is not possible in this case [39].

Secondly, only $\bar{\chi}$ is required to calculate the absorption spectra, which is independent of v_c^0, thereby drastically reducing the coupling terms. However, for the calculation of χ, which is necessary for EELS, this term cannot be neglected, making the coupling terms significantly stronger and thus making the TDA invalid. These two reasons lead to the fact that the TDA is unsuitable for, for example, EELS.

However, we will not consider other electron energy-loss spectra or materials with extremely strongly bound excitons in the following. So, we assume that the TDA is a reasonable approximation for our purposes. For this reason, we can now turn to the construction costs of the matrix and, in particular, take a closer look at the contribution of the local field effects and to the screening of the screened Coulomb potential.

4.5.2 The non local field approximation

The first idea in order to reduce the construction cost could be to neglect the entire electron-hole exchange part $\bar{v}$ in the interaction kernel

$$\Xi_{\mu,\mu'}^{AGW} \approx -\frac{1}{\Omega_v} \sum_{\boldsymbol{G},\boldsymbol{G'}} \rho_{c\boldsymbol{k},c'\boldsymbol{k'}}^*(\boldsymbol{q}+\boldsymbol{G})W_{\boldsymbol{G},\boldsymbol{G'}}(\boldsymbol{q})\rho_{v'\boldsymbol{k'},v\boldsymbol{k}}(\boldsymbol{q}+\boldsymbol{G'})\delta_{\boldsymbol{q},\boldsymbol{k}-\boldsymbol{k'}}$$

since in the case of absorption spectroscopy, it gives only contributions for $\boldsymbol{G} \neq \boldsymbol{0}$ and should be, therefore, relatively small.

The advantage of neglecting this portion would be that we can neglect the construction of the transition dipole matrix elements and the diagonal matrix of the electron-hole exchange part, which reduces the construction costs by $N_G N_v N_c N_k$.

This approximation would, from a physical point of view, mean that we neglect the local field effects, as we discussed in section 3.2. A proof of this statement can be found in [39], who traced back the origin of this effect to $\bar{v}_c$, arising from scattering effects at the periodic crystal potential. Since it is not easy to predict the effect of this approximation, we will use the absorption spectrum (Fig. 4.1) of our two prototypical materials, Si and LiF, to get a feeling for this approximation.
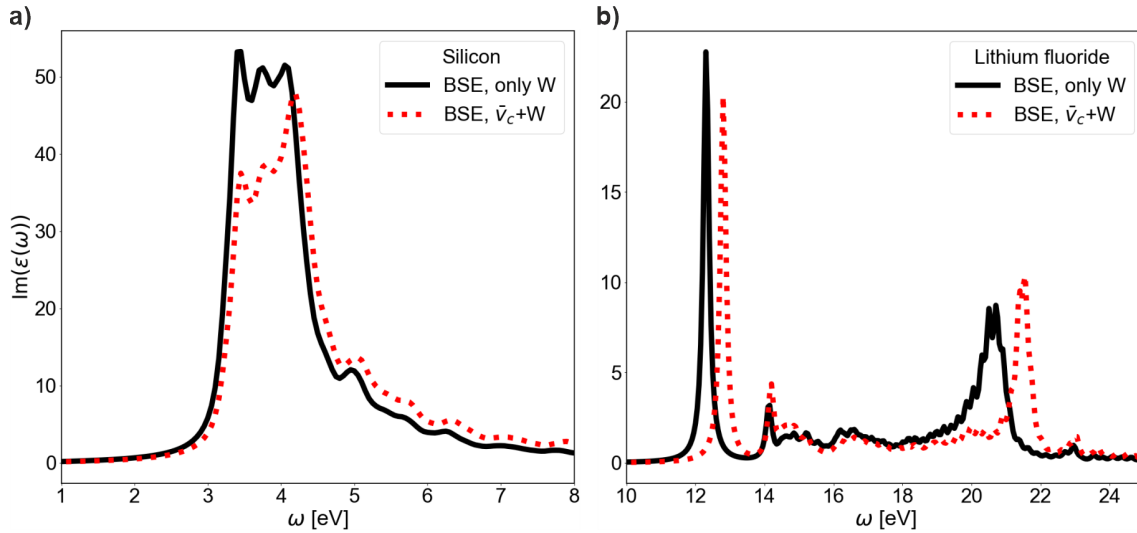

Figure 4.1: Comparison of the calculated absorption spectra ($\text{Im}(\epsilon(\omega))$) in TDA and static GW-BSE for **a)** silicon and **b)** lithium fluoride taking into account the local field effect (red) and neglecting it (black). It turns out that the approximation leads to a blue shift of the spectrum by around 0.5 eV and an overestimation of the strong bound exciton of 13 % for lithium fluoride (b)) and a minor shift (0.12 eV) of the spectrum but a more substantial increase of the excitonic effects near the optical band edge (30%) compared to lithium fluoride for silicon (a)). Numerical values are given in Appendix A; calculations were performed on an 8k grid (2048 k-points) using a Kohn-Sham band structure.

As we can see, neglecting the electron-hole exchange term, i.e. the local field effects, leads to a dramatic increase ($\sim 30\%$) in the spectrum for silicon (Fig. 4.1 a)) on the low energy side. This is probably because the excitonic effects in silicon tend to be less intense. Thus, a strong dominance of the electron-hole exchange term prevails, suppressing the attractive interaction of the screened

Coulomb component and, thus, the excitonic effects. In the case of lithium fluoride (Fig. 4.1 b)), on the other hand, the effect is about half ($\sim 13\%$) as strong as for silicon. The reason may be because the screened Coulomb fraction is much more substantial since the screening generally has to be small for strongly bound excitons to form. Thus, the electron-hole exchange part plays a minor role for lithium fluoride.

Even though the relationship between the magnitude of the electron-hole exchange and the screened Coulomb plays a crucial role, it is not the physical reason why this term can be neglected with minor restrictions for some materials and not for others. If we take a closer look at the physical origin of $\bar{v}_c$, we figure out that the strength of $\bar{v}_c$ is a measure of the inhomogeneity of the material, in particular the electronic density [87]. This can be seen if we consider what occurs in reciprocal space on large scales $\boldsymbol{G} \neq 0$ happens in real space on small scales. This means that the larger $\bar{v}_c$, the more the information varies in real space on small scales, i.e. the more localized the information must be considered. This means that the contribution of the electron-hole exchange component depends very much on the local structure of the material and, thus, on its inhomogeneity.

To summarize: If the inhomogeneity of the electronic density is negligible and the excitonic effects are strong, then it can be assumed under certain circumstances that neglecting the local field effects given by $\bar{v}_c$ will result in relatively minor errors. However, which error is to be expected and whether it meets the respective objective should always be checked.

4.5.3   Diagonal W-approximation

An expensive matter found in the interaction kernel is that the construction of the screened Coulomb component requires two sums over G-space (Eq. (4.15)). It is, of course, questionable whether this is necessary or whether the pure diagonal components of W, which provide the most substantial component in magnitude, is sufficient. To understand the physical meaning of the off-diagonal elements, we should go back to section 3.2 for a moment, where we discussed how the representation of spatial nonlocality is manifested through the off-diagonal elements of the inverse dielectric function. In the case of W, it is precisely the same. To simplify the discussion a bit, we can draw the following picture:

We imagine that we are looking at the interaction between two particles surrounded by many particles, which we do not take into account individually but put into a collective effect screening, which we can see as a cloud of particles. So in order to understand why a neglect of the spatial details in W is maybe possible, we can turn our focus on the two-body system dressed electron and hole.

If we assume that electron and hole can be located far away from each other, then only the macroscopic, long-range dielectric constant will play a crucial role [22, 54]. This is because the interaction takes place at a large distance, which leads to the fact that the exact information on small spatial distances becomes less important. In such a case, the off-diagonal elements of W hardly contribute to the interaction.

If, however, the electron and the hole are held together by strong interaction, then in particular, such effects which take place on short spatial scales, i.e. large G-values, play a decisive role, and

thus the entire structure of W must be considered.

This characteristic of the interaction between electrons and holes, depending on the screening, means that an exact spatial resolution will not always be necessary and, therefore, neglecting the off-diagonal elements of W:

$$W_{\boldsymbol{G},\boldsymbol{G'}}(\boldsymbol{q},\omega) \rightarrow W_{\boldsymbol{G},\boldsymbol{G}}(\boldsymbol{q},\omega), \tag{4.40}$$

which corresponds to the assumption of a homogeneous system

$$W(\boldsymbol{r},\boldsymbol{r'},\omega) \rightarrow W(|\boldsymbol{r}-\boldsymbol{r'}|,\omega),$$

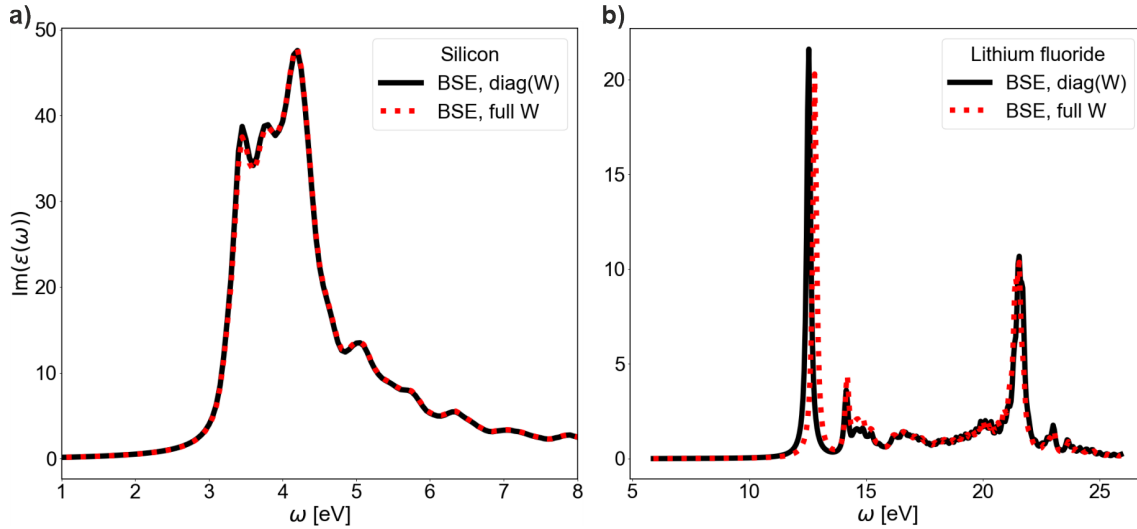could still lead to useful results.

Figure 4.2: Illustration of the difference between the full W matrix (red) and only its diagonal-part (Eq. (4.40)) (black) in TDA and static GW-BSE for the absorption spectrum ($\text{Im}(\epsilon(\omega))$). The plots show that in the case of **a)** silicon the influence of the approximation has only a marginal effect on the spectrum, but for **b)** lithium flouride it leads to a blue shift and intensity increase of the first peak by 7.4%. Numerical values are given in Appendix A; calculations were performed on an 8k grid (2048 k-points) using a Kohn-Sham band structure.

We see that in the case of Si (Fig. 4.2 a)), a material which exhibits weak excitonic interactions, we perceive a slight influence of the off-diagonal element of W. The reason for this can be attributed to the fact, as discussed in general above, that due to the weak interaction of the electrons and holes, the excitons are delocalised. Therefore, the diagonal contribution $\boldsymbol{G} = \boldsymbol{G'}$ gives the most significant contribution. In contrast, LiF (Fig. 4.2 b)), which exhibits strong electron-hole interactions, has strongly localised excitons, and thus, the short-range effects become significantly important. Neglecting the short-range part, therefore, leads to a blue shift in the spectrum and to an increase in the intensity by around 7.4 % of the spectrum. The general problem of averaging the short-range screening for strongly bound excitons is that it leads to an overestimation of screening and, consequently, an underestimation of electron-hole interaction and, thus, to even more strongly bound excitons.

To summarize, neglecting the off-diagonal elements is a good approximation for more homogeneous materials that exhibit strong screening and, thus, weak excitonic effects. In such a case, the

numerical construction cost for the screened Coulomb contribution (Eq. (4.15)) can be derived from a scaling of N_G, instead of N_G^2, without significant losses.

Now that we have discussed these crucial approximations, we should return to the time dependence of the screening. In section 3.4.3, we only mentioned the time dependence of screening phenomenology; in the following, we want to go into a little more detail and thus give the discussion more substance.

4.5.4 The static W-approximation

In section 3.4.3 we briefly traced the effects of the time dependence of the screening phenomenologically back to the temporal retardation of the rearrangement of the particles in the system after electron-hole pairs have been excited. In this context, we briefly mentioned that the plasma frequency plays a role in estimating the build-up time of the correlation effects.

This section, however, aims to add more substance to this discussion by going beyond the figurative description and understanding the dynamics through an equation. For this purpose, we will follow the calculations in [16], where an approximate formula was derived for the frequency-dependent, resonant part of the two-particle Hamiltonian [23]

$$\tilde{H}_{\mu\mu}^{2p,reso}(\omega) = \left(E_{c\mathbf{k}}^0 - E_{c\mathbf{k}}^0\right)\delta_{v,v'}\delta_{c,c'}\delta_{\mathbf{k},\mathbf{k}'} + 2\bar{v}_{\mu\mu'} - \tilde{W}_{\mu\mu'}(\omega)$$

using the Shindo approximation [23, 88].

In the analysis of the work, two approximations for the frequency-dependent screened Coloumb potential $\tilde{W}_{\mu\mu'}(\omega)$, namely a perturbative and plasmon-pole approximation from Hybertsen and Louie [67, 79, 89], were compared with the exact diagonalization of the Shindo-approximated dynamic BSE. The plasmon-pole model provided a reasonably good result for the absorption spectrum of the naphthalene organic crystal (Fig. 4 of [16]), except that the continuum was slightly overestimated compared to the exact diagonalization.

In order not to fill the discussion with unnecessarily complicated equations, we go directly over to the approximated form of the equation for the frequency-dependent W, which is more suitable for gaining a physical understanding:

$$\tilde{W}_{(vc\mathbf{k}),(c'v'\mathbf{k}')}(\omega) \to \tilde{W}_{(vc\mathbf{k}),(c'v'\mathbf{k}')}^{eff} = \frac{1}{\Omega_v}\sum_{\mathbf{q},\mathbf{G}} v_c(\mathbf{q}+\mathbf{G})\tilde{\rho}_{(c,\mathbf{k})(c',\mathbf{k}')}^*(\mathbf{q}+\mathbf{G})\tilde{\rho}_{(v',\mathbf{k}')(v,\mathbf{k})}(\mathbf{q}+\mathbf{G})$$

$$\left[1 - \frac{\omega_p}{2}[1 - \epsilon^{-1}(\mathbf{q}+\mathbf{G},\omega=0)]^{1/2} \cdot \left[\frac{2}{\omega_p(1 - \epsilon^{-1}(\mathbf{q}+\mathbf{G},\omega=0))^{-1/2} + E_b}\right]\right]\delta_{\mathbf{q},\mathbf{k}-\mathbf{k}'}.$$

In the obtained expression, $E_b = E_{gap} - E_{\lambda=0}^{stat}$ corresponds to the binding energy of the lowest excitonic state, where $E_{\lambda=0}^{stat}$ is the eigenenergy of the lowest exciton in the static W approximation. The effective frequency dependent equation for $\tilde{W}(\omega)$ opens the way for the interesting question: *What happens in the case of systems with weak and strong excitonic binding energies?*

We will start with the case of weak binding energies ($E_b \ll \omega_p(1 - \epsilon^{-1}(\mathbf{q}+\mathbf{G},\omega=0))^{-1/2}$). In this case, the entire equation for W goes over into the static equation (Eq. (4.15))

$$\tilde{W}_{(vc\mathbf{k}),(c'v'\mathbf{k}')} = \frac{1}{\Omega_v}\sum_{\mathbf{q},\mathbf{G}} \epsilon^{-1}(\mathbf{q}+\mathbf{G},\omega=0)v_c(\mathbf{q}+\mathbf{G})\tilde{\rho}_{v'v,\mathbf{k}'\mathbf{k}}(\mathbf{q}+\mathbf{G})\tilde{\rho}_{cc',\mathbf{k}\mathbf{k}'}^*(\mathbf{q}+\mathbf{G})\delta_{\mathbf{q},\mathbf{k}-\mathbf{k}'}.$$

Therefore, it follows immediately that in the case of weak excitonic effects, i.e. materials with primary continuum excitons or close to the optical bandgap, the formal dynamic equation for W changes into a static equation.

However, it is also known that dynamic effects can play a decisive role, up to deviations of 15% [16, 59, 90–92]. These cases then fall under the second category $E_b \geqslant \omega_p(1 - \epsilon^{-1}(\boldsymbol{q} + \boldsymbol{G}, \omega = 0))^{-1/2}$, where dynamic effects must be taken into account. The estimation shows very nicely what additional to strong excitonic binding energies must hold in the case of non-negligible dynamic effects, namely that firstly, the screening is very small ($\epsilon \to 1$) because then strongly bound excitons can form (E_b becomes large). Secondly, the right-hand side goes over into $\omega_p(1 - \epsilon^{-1}(\boldsymbol{q} + \boldsymbol{G}, \omega = 0))^{-1/2} \to \omega_p$, which means that, the plasma frequency must be very close to the binding energy of the excitons or smaller. Such a situation can be found for low-dimensional or organic semiconductors, where exciton binding energies can be on the order of 1 eV [16, 93–95].

It is important to note that some bulk semiconductors also have strongly bound excitons, such as LiF, but in the case of LiF, the plasma frequency is 25.3 eV [96], and the binding energy is slightly more than 1 eV. Thus, for the materials we study in this work, the scenario of instantaneous formation of the screening and, thus, a static GW approximation for the BSE is a reasonable assumption.

Now that we have established the working framework and analyzed the effects of the individual approximations analytically and empirically, we will finally delve a step deeper into the resonant part of the two-particle Hamiltonian. The reason for this is that so far, we have only become acquainted with continuum and bound excitonic states in relation to the absorption spectrum; here, we refer in particular to section 3.4.3. In the main part of this work, however, we will deal not only with absorption spectra but also with the eigenvalues and eigenvectors of bound and continuum excitonic states. For this reason, we want to conclude this chapter with a final section, which is already intended as a preview of the main part of this work. The next section will provide a brief history of the formation of bound and continuum excitons.

4.6   A short story about the formation of bound and continuum excitonic states

Since the aim of this work is not purely to develop methods and solutions but, in particular, to gain a deeper understanding of excitonic effects, in this section, we want to deal more intensively with the connection between the resonant part of the two-particle Hamiltonian and the emergence of continuum and bound excitonic states.

From linear algebra lectures, it is clear to us, that the structure of the Hamiltonian determines the eigenvectors and eigenvalues. What is more challenging to understand, however, is how bound states form and, in particular, how the excitonic effects influence the continuum. In this section, we will only be able to give a partial understanding of this. However, we will lay the foundation for a deeper understanding of bound and continuum excitons, which will be helpful for our later developments (Ch. 5).

In order to be more precise in the following, when we talk about different types of excitons, we will specify the definition of the various excitons. As before, we will use the optical bandgap of the IP spectrum for this purpose. In the case of the IP spectrum, the Hamiltonian is only defined by the independent-particle transition part

$$H^{2p,reso,IP}_{\mu\mu'} := \left(E^0_{c\boldsymbol{k}} - E^0_{v\boldsymbol{k}}\right)\delta_{v,v'}\delta_{c,c'}\delta_{\boldsymbol{k},\boldsymbol{k}'}$$

and thus in particular diagonally. In the following, we want to simplify the notation somewhat and define the Hamiltonian $H^0_{\mu\mu'} := H^{2p,reso,IP}_{\mu\mu'}$ and the corresponding eigenenergies of the IP Hamiltonian as $E^0_\mu := \left(E^0_{c\boldsymbol{k}} - E^0_{v\boldsymbol{k}}\right)\delta_{v,v'}\delta_{c,c'}\delta_{\boldsymbol{k},\boldsymbol{k}'}$, and the set of all eigenvalues as $M^{IP}(E^0_\mu)$. With this definition, we can now define mathematically more precisely what we mean by the various excitons. When we previously spoke of bound excitons, we meant excitons whose eigenenergy E_λ lay within the optical bandgap. Mathematically, we therefore want to define bound excitons in general as $E_\lambda < \min(M^{IP}(E^0_\mu))$ and, conversely, continuum excitons as $E_\lambda \geqslant \min(M^{IP}(E^0_\mu))$. Now that we have made these definitions, we want to take a brief phenomenological look at the variables responsible for the difference between the various excitons. In order to limit the discussion to the most important things, we will first look at what information we need to determine the emergence of weakly and strongly bound excitonic states. For this purpose it is worth looking at the structure of the Hamiltonian

$$H^{2p,reso}_{\mu\mu'} := \left(E^0_{c\boldsymbol{k}} - E^0_{v\boldsymbol{k}}\right)\delta_{v,v'}\delta_{c,c'}\delta_{\boldsymbol{k},\boldsymbol{k}'} + \Xi^{AGW}_{\mu\mu'} = H^0_{\mu\mu'} + \Xi^{AGW}_{\mu\mu'}. \qquad (4.41)$$

The first thing we see is that, on the one hand, we have an IP onset $\left(E^0_{c\boldsymbol{k}} - E^0_{v\boldsymbol{k}}\right)\delta_{v,v'}\delta_{c,c'}\delta_{\boldsymbol{k},\boldsymbol{k}'}$, which is determined by the independent-particle transition energies, and on the other hand, we have the interaction component $\Xi^{AGW}_{\mu\mu'}$, which in the context of the excitonic effects has to describe how these arise and what strength they have. Since we want to keep the discussion as simple as possible, we limit our analysis to the results we obtained for Si and LiF. In particular, we want to neglect both the electron-hole exchange component (Ch. 4.5.2) and the off-diagonal components (Ch. 4.5.3) of the screened Coulomb potential as well as limiting ourselves to the case where it is sufficient if we only consider the strongest components of the screened Coulomb interactions, namely the longrange proportion of the first BZ ($\boldsymbol{G} = \boldsymbol{G}' = 0$)

$$\Xi^{AGW}_{(vc\boldsymbol{k}),(v'c'\boldsymbol{k}')} = -\frac{1}{\Omega_v}\langle c\boldsymbol{k}|e^{i(\boldsymbol{k}-\boldsymbol{k}')\boldsymbol{r}}|c'\boldsymbol{k}'\rangle\frac{\epsilon^{-1}_{0,0}(\boldsymbol{k}-\boldsymbol{k}')}{|\boldsymbol{k}-\boldsymbol{k}'|^2}\langle v'\boldsymbol{k}'|e^{-i(\boldsymbol{k}-\boldsymbol{k}')\boldsymbol{r}}|v\boldsymbol{k}\rangle.$$

This is, of course, a drastic assumption, but we want to develop an understanding here, not make accurate calculations, and for this, we need structures that are as simple as possible. All in all, the Hamiltonian for our analyzes then looks similar to the Wannier-Mott exciton model [97,98]

$$H^{2p,reso}_{(vc\boldsymbol{k}),(v'c'\boldsymbol{k}')} \approx \left(E^0_{c\boldsymbol{k}} - E^0_{v\boldsymbol{k}}\right)\delta_{v,v'}\delta_{c,c'}\delta_{\boldsymbol{k},\boldsymbol{k}'} - \frac{1}{\Omega_v}\langle c\boldsymbol{k}|e^{i(\boldsymbol{k}-\boldsymbol{k}')\boldsymbol{r}}|c'\boldsymbol{k}'\rangle\frac{\epsilon^{-1}_{0,0}(\boldsymbol{k}-\boldsymbol{k}')}{|\boldsymbol{k}-\boldsymbol{k}'|^2}\langle v'\boldsymbol{k}'|e^{-i(\boldsymbol{k}-\boldsymbol{k}')\boldsymbol{r}}|v\boldsymbol{k}\rangle.$$

From the structure of our greatly simplified model, we have to consider two aspects in particular when discussing the strength of the excitonic effects. Firstly, the screening, as we have already discussed in section 3.4.3. The weaker the shielding by the other particles, the stronger the excitonic effects can generally be. Secondly, we also have to consider the transition matrix elements, which, depending on the strength of the overlap, can take on values in the interval $[0, 1]$. In general, this

means that the weaker the screening and the better the overlap between the bands, especially in our model between conduction and conduction bands, as well as between valence and valence bands, the stronger the excitonic effects are to be expected. This observation can be used to explain to a certain degree why we observe much more bound excitons for LiF than for Si. Since the screening is weak ($\epsilon_{mac} = 1.9$ [62]), and the overlap is expected to be significant due to partially parallel bands (Fig. 4 in [99]), the excitonic effects are large. Therefore, we can expect stronger excitonic effects in LiF. In the case of Si, we have strong shielding ($\epsilon_{mac} = 11.4$ [61]) and partially parallel bands (Fig. 1 in [100]), which means that the excitonic effects are stronger than usual for materials with such a high screening but still much weaker than for LiF.

Now that we have made this phenomenological explanation for the excitonic effects in Si and LiF, we will take a step towards the main part of this work and look at how the excitonic eigenenergies relate to the corresponding IP particle eigenenergies. In particular, we will take a look at how the difference between them $(E_\lambda - E_\lambda^0)$ changes as we increase the number of k-points bit by bit, this should represent a form of convergence into the thermodynamic limit $(N_k \to \infty)$ of the k-points, which is important because the interaction part of the Hamiltonian scales with the number of k-points like: $\Xi_{\mu\mu'}^{AGW} \sim \frac{1}{N_k}$.
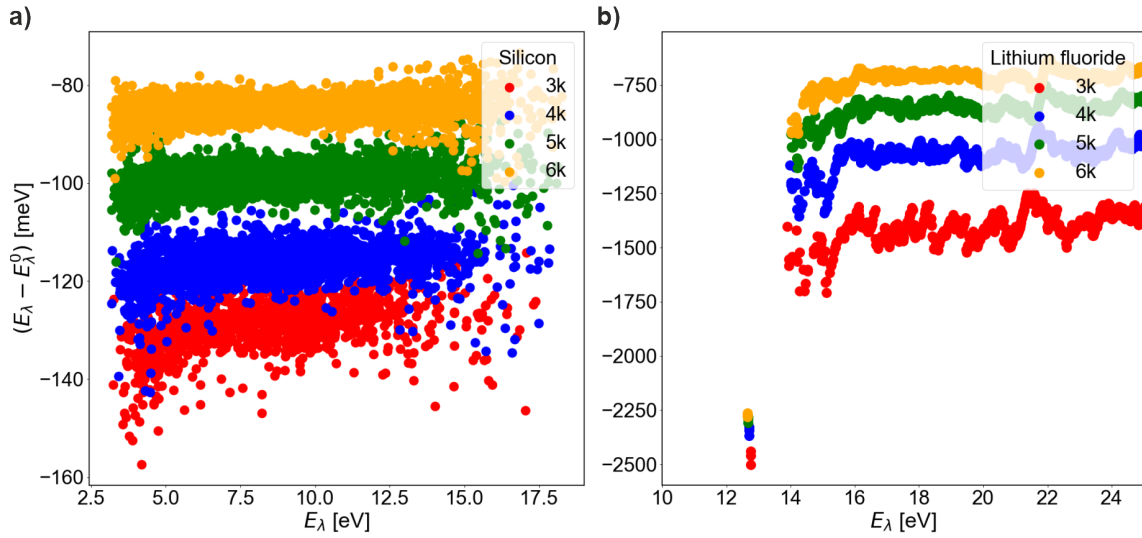

Figure 4.3: Difference between the independent-particle transition energies E_λ^0 and the $\underline{\underline{H}}^{2p,reso}$ Hamiltonian (Eq. (4.13)) eigenvalues E_λ. Results shown for increasingly dense k-grids of 3k (red), 4k (blue), 5k (green) and 6k (orange), for **a)** silicon and **b)** lithium fluoride . It can be seen that the continuum of the excitonic eigenenergies $E_\lambda \geqslant \min(M^{IP}(E_\mu^0))$ for a) Si $(\min(M^{IP}(E_\mu^0)) = 3.27$ eV, 6k) as well as for b) LiF $(\min(M^{IP}(E_\mu^0)) = 14.92$ eV, 6k) slowly moves towards the independent-particle transition energies. The same is true for both materials near the optical band edge, but not for the lowest three eigenenergies of LiF, which represent the strongly bound excitons. Numerical values are given in Appendix A; calculations were performed with 6 transition bands for silicon and lithium fluoride, using a Kohn-Sham band structure.

As figure 4.3 shows, it seems to exist for both materials in the continuum (Si: $(\min(M^{IP}(E_\mu^0)) = 3.27$ eV, 6k), LiF: $(\min(M^{IP}(E_\mu^0)) = 14.92$ eV, 6k)) a form of systemic approach of the eigenenergies to those of the independent-particle transition energies. The same can be observed for

the bound excitons slightly below the optical band edge. On the other hand, the strongly bound excitons in LiF (lowest three eigenvalues) converge only very slowly, if they converge at all. Of course, with a maximum number of 864 k-points we are far from talking about convergence in the thermodynamic limit, but we see a pattern in both materials: The continua of the two materials behave very similarly, although the strength of the excitonic effects that we observe in an absorption spectrum (Fig. 3.2) differ drastically. Furthermore, we see that we cannot distinguish the bound excitons close to the optical band edge in Si and also for LiF on the 6k-grid from the continuum, indicating that the behavior of these two exciton groups might be very similar. In the case of the strongly bound excitons, we observe that they are far away from the indedependent-particle transition energies.

However, for the moment, we want to leave these observations as observations, because for an analysis and understanding we need much more information than we have at the time of this chapter. Nevertheless, we want to formulate some questions which we will answer in the course of the main part: *Is the behavior we observe in the continuum of the two materials really a pattern or is it just coincidence? Can we explain this behavior physically and mathematically? Can we use this observation to make approximations for the eigenenergies? How is the strength of the excitonic effects related to the observations in the continuum?* These and some more questions, which are not directly mentioned again in the main part but will be answered, will help us make exciting discoveries and make surprisingly accurate approximations for eigenenergies and eigenvectors of the excitons.

4.7   Intermediate summary

In this chapter, we defined the work framework of the static BSE. We have shown that the inversion of the complex 4-point function of the original Dyson equation (Eq. (3.40)) can be transformed into an eigenvalue equation by representing it in the Bloch space (Eq. (4.17)). This means that an inversion of a highly complex matrix frequency by frequency is no longer necessary since it is sufficient to diagonalize a matrix once and for all. Further, we showed that in the case of weak spin-orbit interactions, a reduction of the rank of the matrix by 4 is possible.

We also examined how we get the quasi-particle energies and states for constructing the excitonic Hamiltonian. For this purpose, we discussed the concept of quasi-particle energies and wave functions and their relation to the Kohn-Sham scheme. It turned out that both the wave function and the energies can often be calculated sufficiently well using KS-DFT with a scissor-operator. The approximation significantly reduces the numerical costs for constructing the excitonic Hamiltonian. After these crucial findings, we looked further into possible approximations of the general excitonic Hamiltonian. We showed that the original task, namely constructing and diagonalizing the Hamiltonian from equation (4.17), can be significantly simplified under certain circumstances. We have seen that we can reduce the construction costs if we construct only the diagonal part of the screened Coulomb interaction instead of constructing the full matrix. This approximation is particularly successful when we work with materials that are determined by delocalized excitons. Further, we have dealt with one of the standard approximations in theoretical spectroscopy, the

TDA. The approximation neglects the coupling terms and thus reduces the previously $(2 \cdot N_v N_c N_k)$-dimensional problem to an $(N_v N_c N_k)$-dimensional one. So, a significant reduction in numerical costs is generated. Such an approach is particularly successful if plasmonic and excitonic effects are well separated, excitonic effects are not too strong, and the material does not exhibit excessive inhomogeneity.

Finally, we have already taken a small step beyond the working framework, namely at the point where we have become familiar with the different forms of excitons. In this context, an impressive observation was that in both Si and LiF, the continuum of the excitonic eigenenergies appear to converge towards the independent-particle transition energies in the thermodynamic limit. This observation is interesting because it suggests a pattern that we can explain physically and mathematically and, if we understand it, we may can use it for approximations. However, this will be part of the main part of this work.

Chapter 5

Analysis and perturbative approach to excitonic effects for absorption spectra, eigenvectors and eigenvalues

"We absolutely must leave room for doubt or there is no progress and there is no learning. There is no learning without having to pose a question."

Richard P. Feynman

After we have explained the theoretical foundations and discussed necessary approximations, we can now deal with the central topics of this work. However, before we do that, we will briefly recapitulate the essential information from the theoretical foundations and the working environment. This approach will help us to visualize the whole theoretical field and analyze different areas, posing new questions.

In the following, we assume that we work in the static GW approximation, so the BSE for the correlation function is given as:

$$\bar{L}(1,2,1',2') = L_0(1,2,1',2')+$$
$$L_0(1,\bar{4},1',\bar{3})\left[\bar{v}_c(\bar{3},\bar{5})\delta(\bar{3},\bar{4})\delta(\bar{5},\bar{6}) - W^A(\bar{3},\bar{4})\delta(\bar{3},\bar{5})\delta(\bar{4},\bar{6})\right]\bar{L}(\bar{5},2,\bar{6},2'). \tag{5.1}$$

As we have seen in section 4.3, we can transform the solution of this equation into an eigenvalue equation of an effective two-particle Hamiltonian. In the following, we only want to deal with the TDA version of the Hamiltonian (Eq. (4.36)), the case of negligible spin-orbit coupling and non-metal systems. In this case, the BSE can be mapped into an eigenvalue equation for the resonant part of the two-particle Hamiltonian (Eq. (4.37))

$$\sum_{\mu'} H^{2p,reso}_{\mu\mu'} A^{\mu'}_\lambda = E_\lambda A^\mu_\lambda, \tag{5.2}$$

for an effective two-particle Hamiltonian

$$H_{\mu\mu'}^{2p,reso} := \left(E_{c\boldsymbol{k}}^0 - E_{v\boldsymbol{k}}^0\right)\delta_{v,v'}\delta_{c,c'}\delta_{\boldsymbol{k},\boldsymbol{k}'} + \Xi_{(vc\boldsymbol{k})(v'c'\boldsymbol{k}')}^{AGW} = H_{\mu\mu'}^0 + \Xi_{\mu\mu'}^{AGW}, \tag{5.3}$$

with the interaction kernel given as:

$$\Xi_{\mu\mu'}^{AGW} := \Xi_{(vc\boldsymbol{k})(v'c'\boldsymbol{k}')}^{AGW} = -\frac{1}{\Omega_v}\sum_{\boldsymbol{G},\boldsymbol{G}'}\tilde{\rho}_{(c,\boldsymbol{k})(c',\boldsymbol{k}')}^*(\boldsymbol{q}+\boldsymbol{G})W_{\boldsymbol{G},\boldsymbol{G}'}(\boldsymbol{q})\tilde{\rho}_{(v',\boldsymbol{k}')(v,\boldsymbol{k})}(\boldsymbol{q}+\boldsymbol{G}')\delta_{\boldsymbol{q},\boldsymbol{k}-\boldsymbol{k}'}$$
$$\tag{5.4}$$
$$+\frac{2}{\Omega_v}\sum_{\boldsymbol{G}\neq\boldsymbol{0}}\tilde{\rho}_{(c,\boldsymbol{k})(v,\boldsymbol{k})}^*(\boldsymbol{G})\bar{v}_c(\boldsymbol{G})\tilde{\rho}_{(v',\boldsymbol{k}')(c',\boldsymbol{k}')}(\boldsymbol{G}),$$

For the construction of the Hamiltonian, we assume that KS-DFT with a scissor operator is sufficiently good for the computation of the quasi-particle energies $E_{n_i}^0$ and wave functions Ψ_{n_i}. Within these assumptions, we can then calculate the modified dielectric susceptibility using

$$\bar{\chi}_{\boldsymbol{G},\boldsymbol{G}'}^{TDA,reso}(\boldsymbol{q},\omega) = \sum_\lambda\sum_{\mu\mu'}\tilde{\rho}_\mu^*(\boldsymbol{q}+\boldsymbol{G})\frac{\bar{A}_\lambda^{(\mu)}\bar{A}_\lambda^{*(\mu')}}{E_\lambda - \Omega}\tilde{\rho}_{\mu'}(\boldsymbol{q}+\boldsymbol{G}')f_{\mu'}$$
$$\tag{5.5}$$
$$= \sum_{\mu\mu'}\tilde{\rho}_\mu^*(\boldsymbol{q}+\boldsymbol{G})\left[\underline{\underline{H}}^{2p,reso} - \Omega\underline{\underline{1}}\right]_{\mu\mu'}^{-1}\tilde{\rho}_{\mu'}(\boldsymbol{q}+\boldsymbol{G}')f_{\mu'}$$

and with this expression, we are then able to finally calculate the macroscopic dielectric function

$$\epsilon_M(\omega) = \frac{1}{\left[1 - \lim_{\boldsymbol{q}\to\boldsymbol{0}}v_c(\boldsymbol{q})\bar{\chi}(\boldsymbol{q},\omega)\right]_{00}^{-1}}. \tag{5.6}$$

After summarizing the essential equations, we can now ask ourself an infinite number of questions which one would like to consider for further work; as an example, the very general questions are: *Is our starting point (Eq. (5.1)) the only possible one? Can one optimize the construction of the matrix (Eq. (5.4))? Is diagonalization a useful way to solve the eigenvalue equation for the effective two-particle Hamiltonian (Eq. (5.2))?*

Much work has been invested in many of these questions over the decades. For example, attempts have been made to develop new alternative calculations for the two-particle correlation function [101]. Furthermore, work was also invested in reducing the construction costs of the matrix and its contents, for example, the screening through models [22, 102, 103] or machine learning [104]. Another significant development was the so-called double grid method [25, 35]. This method reduces the number of k-points required for the Hamiltonian construction, which is one of the biggest numerical issues.

Moreover, an enormous amount of work is invested in solving eigenvalue equations in a numerically more favorable way since this task occurs in almost all areas of applied mathematics. Examples of developed and used algorithms are the Jacobi-Davidson iteration method [33], the CG+SR algorithm [24], the QR algorithm [27–29], the inverse iteration [30, 32], the divide and conquer method [31], the Lanczos method [26], the MR^3 algorithm [34] and many more.

In this work, however, we will take a slightly different approach than developing clever algorithms or reducing construction costs. We want to understand the various elements of the BSE better and thus the effective two-particle Hamiltonian, especially concerning strongly bound and continuum excitons (Ch. 4.6). This approach will help us to provide new explanations for the emergence of the various excitonic effects and pave the way for new exact as well as simplified equations. As we will see, these equations will provide a deeper understanding, numerically favorable solutions, and new ways to calculate absorption spectra, eigenvalues, and eigenvectors.

For the further work, it is essential to note that we will only use partially converged results. We will carry out the calculations for our prototypical materials Si and LiF using a low band BSE calculation; for this, we will reduce the transition bands in both materials to 6 (Appendix C). Furthermore, we will forego convergence with the number of k-points and limit ourselves to grids of size 3k (108 k points), 4k (256 points), 5k (500 points) and 6k (864 points). We are doing this for ecological and economic reasons since even the unconverged Hamiltonian defines the problem.

5.1   The Haydock-Lanczos and Neumann algorithm

We start our journey by familiarizing ourselves with the standard procedure that one uses if one is only interested in the absorption spectrum but not in eigenvalues or eigenvectors, the so-called Haydock-Lanczos method [25, 26, 105–107].

This section will aim to answer the following questions: *How does the Haydock-Lanczos algorithm work? What are its limitations? Is it possible to use the Haydock-Lanczos algorithm to understand how the excitonic effects manifest themselves within the absorption spectrum?*

Answering these questions will enrich our knowledge and primarily support us when we venture deeper and deal with the eigenvalues and eigenvectors.

5.1.1   Haydock-Lanczos algorithm

In order to calculate, e.g. the absorption spectrum, one can use the macroscopic dielectric function (Eq. (5.6)) together with the modified dielectric susceptibility (Eq. (5.5)):

$$\bar{\chi}_{\boldsymbol{G},\boldsymbol{G}'}^{AGW,TDA,reso}(\boldsymbol{q},\omega) = \sum_{\mu\mu'} \tilde{\rho}_{\mu}^{*}(\boldsymbol{q}+\boldsymbol{G}) \left[\underline{\underline{H}}^{2p,reso} - \Omega\underline{\underline{1}}\right]_{\mu\mu'}^{-1} \tilde{\rho}_{\mu'}(\boldsymbol{q}+\boldsymbol{G}')f_{\mu'}. \qquad (5.7)$$

The crucial point now to be able to apply the Haydock-Lanczos algorithm [26, 105] is the structure of the equation $\left(\underline{\tilde{\rho}}^{*}\left[\underline{\underline{H}}^{2p,reso} - \Omega\underline{\underline{1}}\right]^{-1}\underline{\tilde{\rho}}\right)$, which allows solving the expensive problem of matrix inversion ($\sim O(N^3)$) or diagonalization of the Hamiltonian by an algorithm which scales with $O(m_{Hay}N^2)$, where m_{Hay} is the number of iteration steps used to achieve convergence and N is the dimension of the Hamiltonian $\underline{\underline{H}}^{2p,reso}$.

For the presentation of the algorithm, we will adopt its formulation from the publication by [107], as the latter provides a detailed breakdown of the algorithm.

As can be seen above (Eq. (5.7)), quite generally, the target is to calculate an expression like:

$$\langle i|(\Omega - \hat{H})^{-1}|j\rangle, \qquad (5.8)$$

where Ω is the complex frequency $\Omega = \omega + i\Gamma$, $\hat{H}$ is the Hamiltonoperator of the system, in our case the resonant two-particle Hamiltonian (Eq. (5.3)), and $|j\rangle, |i\rangle$ are vectors, which are known. The goal is now to bypass the expensive matrix inversion and to solve it with a suitable process. To achieve this, we define:

$$|\tilde{j}_{\Omega}\rangle = (\Omega - \hat{H})^{-1}|\tilde{j}\rangle, \ |\tilde{j}\rangle = |j\rangle/||j||. \qquad (5.9)$$

Subsequently, we substitute equation (5.9) into equation (5.8), which leads to the expression:

$$\langle i|(\Omega - \hat{H})^{-1}|j\rangle = \langle i|\tilde{j}_{\Omega}\rangle \, ||j||.$$

Since $|i\rangle, |j\rangle$ are known, the next step is to obtain $|\tilde{j}_\Omega\rangle$. To this end, we generate a set of orthogonal Lanczos vectors $\{|f_n\rangle\}$, with the starting vector $|f_0\rangle = |\tilde{j}\rangle$, and by using the recursion relations [108]:

$$b_{n+1}|f_{n+1}\rangle = \hat{H}|f_n\rangle - a_n|f_n\rangle - b_n|f_{n-1}\rangle,$$

with the definitions $a_n := \langle f_n|\hat{H}|f_n\rangle$ and $b_n := \langle f_{n-1}|\hat{H}|f_n\rangle$. The next step is to expand the solution vector $|\tilde{j}_\Omega\rangle$ in the Lanczos basis:

$$|\tilde{j}_\Omega\rangle = \sum_n c_n(\Omega)|f_n\rangle, \tag{5.10}$$

with the expansion coefficients given as $c_n(\Omega) = \langle f_n|\tilde{j}_\Omega\rangle$. This bases transformation allows us to express the initial equation (5.8) as:

$$\langle i|(\Omega - \hat{H})^{-1}|\tilde{j}\rangle\,||j|| = \langle i|\tilde{j}_\Omega\rangle\,||j|| = \sum_n \langle i|f_n\rangle c_n(\Omega)||j||.$$

With the equation obtained, we have already transformed the matrix inversion into a series expansion in the Lanczos basis. Now to make the expansion usable, we require an equation for the expansion coefficients c_n, which can be obtained by applying equation (5.8) to $|\tilde{j}_\Omega\rangle$ and using the reformulated equation (5.9) $\hat{H}|\tilde{j}_\Omega\rangle = \omega|\tilde{j}_\Omega\rangle - |\tilde{j}\rangle$ in order to obtain:

$$b_{n+1}c_{n+1}(\Omega) = \Omega c_n(\Omega) - \delta_{n0} - a_n c_n(\Omega) - b_n c_{n-1}(\Omega). \tag{5.11}$$

By introducing the relaxation function

$$\phi_n(\Omega) = [\Omega - a_n - b_{n+1}^2\phi_{n+1}(\Omega)]^{-1},$$

and rearranging the determination equation for c_n (Eq. (5.11)), we can obtain the final equation for the expansion coefficients as:

$$c_n(\Omega) = \phi_n(\Omega)b_n c_{n-1}(\Omega).$$

Now that we hold the general representation of the algorithm in our hands, we will assume that we are only interested in structures like $\langle j|(\Omega - \hat{H})^{-1}|j\rangle$, e.g. the $\boldsymbol{G} = \boldsymbol{G}' = \boldsymbol{0}$ component of the modified dielectric susceptibility, which we need in order to construct the macroscopic dielectric function (Eq. (5.6)). In this case, the equation simplifies to the well-known Haydock-Lanczos recursive formula

$$\langle \tilde{\rho}|(\Omega - \hat{H})^{-1}|\tilde{\rho}\rangle = \langle \tilde{\rho}|\tilde{\tilde{\rho}}_\Omega\rangle\,||\tilde{\rho}||^2 = c_0(\Omega)||\tilde{\rho}||^2, \tag{5.12}$$

where $\tilde{\rho}$ represents the transition dipoles and c_0 is given in a continued fraction series as:

$$c_0(\Omega) = \frac{1}{\Omega - a_0 - b_1^2\phi_1(\Omega)} = \frac{1}{\Omega - a_0 - \dfrac{b_1^2}{\Omega - a_1 - \dfrac{b_2^2}{\Omega - a_2 - b_3^2\phi_3(\Omega)}}}. \tag{5.13}$$

Before we move on to detailed analyses of the algorithm, we want to verify whether the algorithm is suitable for calculating the absorption spectra of materials with weak and strong excitonic effects. To do this, we calculate the spectra for Si and LiF using the Haydock-Lanczos algorithm.
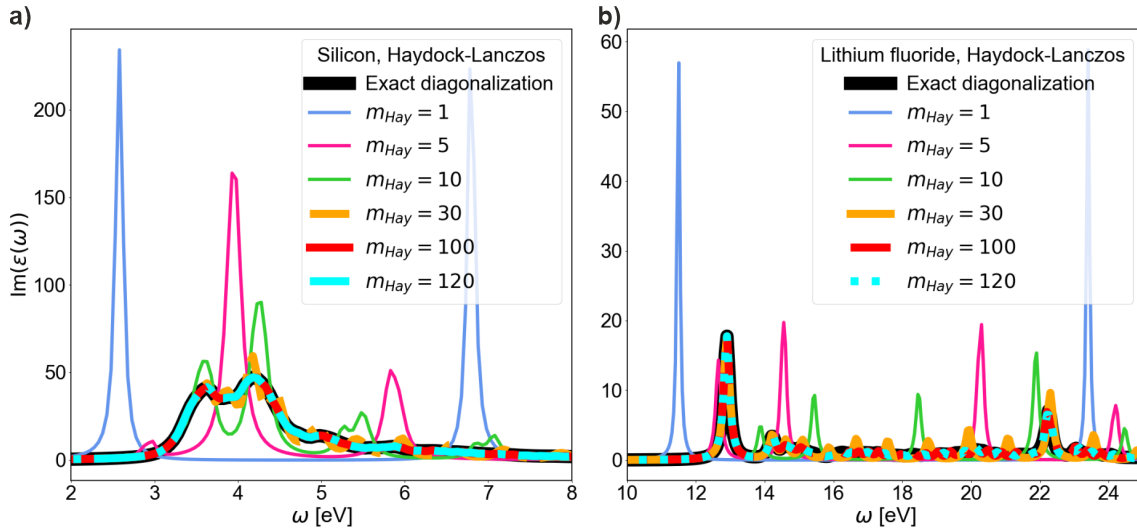
Figure 5.1: Haydock-Lanczos absorption spectrum ($\mathrm{Im}(\epsilon(\omega))$) (Eq. (5.12)) for **a)** silicon and **b)** lithium fluoride in static GW-BSE. Shown the exact diagonalization result in black and different iterations m_{Hay} (1, 5, 10, 30, 100, 120) of the Haydock-Lanczos algorithm. The result shows an approach of the algorithm for about 100 iterations to the exact diagonalization result for both materials. Numerical values are given in Appendix A; calculations were performed on a 5k grid (500 k-points) using a Kohn-Sham band structure, with 6 transition bands for optical calculations and a Lorentzian broadening of $\Gamma = 3.67 \cdot 10^{-3}$ Ha.

In figure 5.1, we see that we need at least 100-120 iterations to get an indistinguishable spectrum from the exact one on the scale of the figure for Si and LiF. Furthermore, we observe that the algorithm delivers the same spectrum obtained through the exact diagonalization for both materials with weak (Fig. 5.1 a)) and strong excitonic effects (Fig. 5.1 b)). Based on our empirical results and the numerous results and analyses from the literature [25, 107, 109, 110], we can assume that the Haydock-Lanczos algorithm is not limited to materials with weak or strong excitonic effects. What we also see, however, is that it is not easy to assign a physical meaning to the individual iteration steps or to say which iteration step leads to which result and why. The only thing we see is that in both spectra, low iteration steps (blue, pink) produce seemingly arbitrary peaks, which, with increasing iteration steps, split into several peaks (green) that eventually turn into a form of oscillation (orange) that is already close to the target spectrum. With further iteration steps, the oscillations flatten out and finally disappear (red, light blue). However, we cannot guess or read from the continued fraction representation (Eq. (5.13)) which part of the Hamiltonian makes which contribution.

However, the reason is not only the representation as a continued fraction, which stands in the way of using the algorithm to develop physical understanding, but also that we start with an arbitrary vector for the Lanczos basis ($|f_0\rangle = |\tilde{\rho}\rangle/||\tilde{\rho}||$), which per se has no direct relation to the Hamiltonian, such as an eigenvector or the same. To make this clear, we will illustrate the first iteration step. In this case, the spectrum is built up from:

$$\langle\tilde{\rho}|(\Omega - \hat{H})^{-1}|\tilde{\rho}\rangle = \langle\tilde{\rho}|\tilde{\tilde{\rho}}_\Omega\rangle\,||\tilde{\rho}||^2 = c_0(\Omega)||\tilde{\rho}||^2 = \frac{||\tilde{\rho}||^2}{\Omega - a_0} = \frac{||\tilde{\rho}||^2}{\Omega - \langle\tilde{\rho}|\hat{H}|\tilde{\rho}\rangle/||\tilde{\rho}||^2}.$$

Since, as mentioned, the vector $|f_0\rangle = |\tilde{\rho}\rangle/||\tilde{\rho}||$ has no direct physical or mathematical connection

to the Hamiltonian, the arising peaks have no physically meaning, like in example the position of eigenvalues of the independent Hamiltonian or something like that. This shows us that the algorithm is unsuitable for a deeper understanding of excitonic effects in absorption spectra. It can only provide us with absorption spectra.

Now that we have learned the algorithm in the form of equations and have seen that it gives the same result as the diagonalization of the resonant part of the two-particle Hamiltonian for materials with both weak and strongly bound excitons, we will turn to the question: *What is the numerical scaling of the algorithm, and what are its limitations?*

For this purpose we first examine the numerical scaling of the algorithm, which is relatively simple, because to calculate $\langle \tilde{\rho}(\boldsymbol{q} + \boldsymbol{G}) | (\Omega - \hat{H})^{-1} | \tilde{\rho}(\boldsymbol{q} + \boldsymbol{G}) \rangle$ one only has to calculate three quantities:

- a_n, $O(N^2)$,

- b_n, $O(N^1)$,

- $|f_{n+1}\rangle$, $O(N^2)$.

Since only c_0 needs to be generated and therefore a_n and b_n, which are similar for most of the the complex frequencies Ω. We can estimate the numerical scaling for the calculation of the macroscopic dielectric function as $O(m_{Hay}N^2)$. After we have worked out the numerical scaling of the algorithm, we need to determine what affects the number of iterations m_{Hay} required. One aspect that is important for the convergence speed and, therefore, for the number of needed iteration steps is the minimal distance $\min_\lambda |E_\lambda - \Omega|$ between the closest eigenvalue E_λ of the Hamiltonian to the value of the complex frequency Ω [111]. This statement, in reverse, means the convergence of the Haydock-Lanczos algorithm is fast if Ω is far from the eigenenergy spectrum. Since the complex frequency is defined as the frequency plus broadening ($\Omega = \omega + i\Gamma$), we can immediately recognize that broadening plays a central role in convergence. To evaluate more precisely, the broadening should be larger than the characteristic level spacing of the Hamiltonian ($E_\lambda - E_{\lambda+1} < \Gamma$) in order to guarantee that the iterative scheme is more efficient than standard diagonalization [112]. For this reason, we can understand broadening as a parameter that prevents Ω from coinciding with an eigenvalue, which would result in slow convergence and render the algorithm impractical.

This analysis shows us that the broadening and the choice of frequency primarily determine the convergence speed of the algorithm. The closer Ω is to the eigenvalues of the Hamiltonian, the slower the iterative procedure will converge. However, what does not emerge from this analysis is how the algorithm behaves when changing the k-grid density since, in practical application, we never have an infinitely dense grid and, therefore, never infinitely dense eigenvalues.

In order to understand the behavior of the algorithm with increasing k-grid density and also to see in a practical application how the Haydock-Lanczos algorithm behaves depending on the choice of broadening, we will look at the absorption spectra for Si depending on broadening, k-grid and the number of iteration steps needed to obtain a stable result.
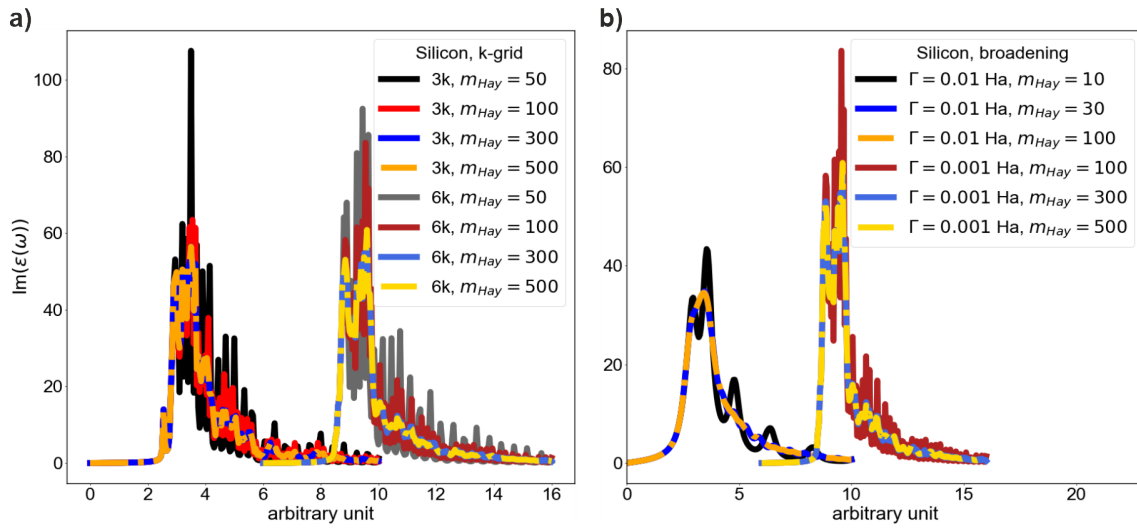
Figure 5.2: Convergence behavior of the Haydock-Lanczos algorithm (Eq. (5.12)) for the imaginary part of the dielectric function, $\text{Im}(\epsilon(\omega))$, of silicon in TDA and static GW-BSE. The convergence behavior is shown with respect to **a)** the k-grid density for a fixed Lorentzian broadening $\Gamma = 0.001$ Ha and **b)** the Lorentzian broadening Γ for a fixed k-grid (6k). It is shown in **a)** that the convergence speed with the number of iteration steps m_{Hay} of the Haydock-Lanczos algorithm hardly depends, if at all, on the density of the k-grid and in **b)** that there is a strong correlation between the number of iteration steps required for convergence within the scope of the observable change in the spectrum in the images and the chosen Lorentzian broadening. Calculation were done using a Kohn-Sham band structure and 6 transition bands for the optical calculations. Numerical values are given in Appendix A.

As we can see (Fig. 5.2 a)), for a finite grid, the Haydock-Lanczos algorithm behaves identically for a grid of 3k (108 k-points) and 6k (864 k-points) with constant broadening within the scope of the observable: after about 300 iteration steps there is no longer any optical visible change in the spectra to observe. The reason for this may be that the selected frequencies and broadening are in such a range that they sufficiently fulfill the conditions for the convergence speed discussed above. In the case of broadening, the spectrum behaves as we discussed above. As soon as the broadening becomes smaller, so frequencies move closer to the eigenvalues, we need more iteration steps until no visible change in the spectrum can be observed. For a broadening of $\Gamma = 0.01$ Ha, only about 30 iteration steps are necessary, and for a broadening of $\Gamma = 0.001$ Ha, ten times that number, namely 300 iteration steps, are required.

After discussing the Haydock-Lanczos algorithm, we understand that the algorithm is well-suited for calculating absorption spectra. However, on the other side, it gives us minimal opportunity to analyze and understand the excitonic effects more profoundly. The reason for this lies in the complicated structure of $c_0(\Omega)$ (Eq. (5.13)), which represents a continued fraction and therefore, much information is nested. Furthermore, the ability to develop physical intuition and understanding is further complicated by how we construct the Lanczos base (Eq. (5.10)), especially how we have to choose the starting vector. It is, therefore, crucial that we find an algorithm that allows us more profound insight.

5.1.2   Neumann algorithm

As our primary focus is on understanding, which is not easy based on the Haydock-Lanczos algorithm, we want to use a straightforward algorithm to take the inversion apart piece by piece. One such algorithm is the Neumann algorithm. The algorithm corresponds to what can be described as matrix perturbation theory, and detailed expressions can be found in various mathematical textbooks on functional analysis, such as the book by D. Werner [113]. If such an algorithm were to be applicable, it would be extremely practical because a perturbation theory has always been a proven method in physics for making calculations and generating understanding.

To explain the idea of the algorithm as simply as possible, we begin with the scalar case since transferring the idea to the matrix case is then straightforward: Suppose we wish to expand the expression $(1-x)^{-1}$ into a Taylor series; the condition $|x| < 1$ must be satisfied. If the condition is fulfilled, we can initiate a simple Taylor series:

$$(1-x)^{-1} \approx 1 + \frac{1}{1!}(-1)^2 x + \frac{2}{2!}(-1)^4 x^2 + \frac{2 \cdot 3}{3!}(-1)^6 x^3 + \cdots = \sum_{k=0}^{m_{PT}} x^k.$$

Now, envisioning the same structure as a matrix equation, we obtain:

$$\left(\underline{\underline{1}} - \underline{\underline{X}}\right)^{-1} = \sum_{k=0}^{N} \underline{\underline{X}}^k,$$

with the condition that the operator norm fulfills $||\underline{\underline{X}}|| < 1$. Since the operator norm on a real or complex Hilbert space is equivalent to the spectral radius $S_r(\underline{\underline{X}}) = \max |\lambda(\underline{\underline{X}})|$, i.e., the largest eigenvalue in magnitude, one has to deal with the condition $S_r(\underline{\underline{X}}) < 1$ for the expansion.

However, practical situations often involve a different structure, such as $(\underline{\underline{A}} - \underline{\underline{B}})^{-1}$, where one refers to $\underline{\underline{B}}$ as the perturbation of $\underline{\underline{A}}$. In such a case, a short transformation helps to find the expansion of the inversion as

$$(\underline{\underline{A}} - \underline{\underline{B}})^{-1} = \underline{\underline{A}}^{-1}(\underline{\underline{1}} - \underline{\underline{B}}\,\underline{\underline{A}}^{-1})^{-1} = \underline{\underline{A}}^{-1} \sum_{k=0}^{m_{PT}} \left(\underline{\underline{B}}\,\underline{\underline{A}}^{-1}\right)^k, \tag{5.14}$$

with the condition $S_r(\underline{\underline{B}}\underline{\underline{A}}^{-1}) < 1$.

Since, in addition to generating understanding, our goal is also to provide an efficient way of calculation, we need to consider that this algorithm, in its current form, cannot be numerically more efficient than $\sim O(N^3)$. The reason for the inefficiency of the Neumann series arises from the necessity to invert $\underline{\underline{A}}$ initially, followed by the performance of m_{PT} matrix multiplications, each incurring a cost of $\sim O(N^3)$. Therefore, the algorithm can only offer advantages when two conditions are met: firstly, $\underline{\underline{A}}$ should be a diagonal matrix ($\underline{\underline{A}}^{-1} = 1/\underline{\underline{A}}$), and secondly, matrix-vector multiplication must be executed from the right- or left-hand side. This modification implies that we require the following structure resembling:

$$(\underline{\underline{A}} - \underline{\underline{B}})^{-1}\underline{v} = \frac{1}{\underline{\underline{A}}} \sum_{k=0}^{m_{PT}} \left(\underline{\underline{B}}\frac{1}{\underline{\underline{A}}}\right)^k \underline{v} \approx \frac{1}{\underline{\underline{A}}}\underline{v} + \frac{1}{\underline{\underline{A}}}\underline{\underline{B}}\frac{1}{\underline{\underline{A}}}\underline{v} + \frac{1}{\underline{\underline{A}}}\underline{\underline{B}}\frac{1}{\underline{\underline{A}}}\underline{\underline{B}}\frac{1}{\underline{\underline{A}}}\underline{v} + \cdots,$$

which would lead to a reduction of the computational cost from $O(N^3)$ to $O(m_{PT}N^2)$. The advantageous aspect for us is that the structure of the response function (Eq. (5.4)) aligns perfectly with the efficient utilization of the Neumann series. However, we must bear in mind what the

limitation of the algorithm is, namely the size of the spectral radius. For this reason, we ask ourselves the question: *Can we choose the matrices $\underline{\underline{A}}$ and $\underline{\underline{B}}$ such that the spectral radius is less than one?* To check this, we will restrict ourselves to the macroscopic dielectric function (Eq. (4.1)). For this purpose, we combine the equation for the macroscopic dielectric function (Eq. (4.1)) and the resonant part of the modified dielectric susceptibility (Eq. (4.38)), in order to obtain the expression:

$$\epsilon(\omega) = 1 + \lim_{\boldsymbol{q}\to\boldsymbol{0}} v_c(\boldsymbol{q})\tilde{\underline{\varrho}}^{\dagger}(\boldsymbol{q})\left[\left(\Omega\underline{\underline{1}} - \underline{\underline{H}}\right)^{-1}\right]\tilde{\underline{\varrho}}(\boldsymbol{q})$$

or in a more general representation

$$\epsilon(\omega) = 1 + \lim_{\boldsymbol{q}\to\boldsymbol{0}} v_c(\boldsymbol{q})\tilde{\underline{\varrho}}^{\dagger}(\boldsymbol{q})\left[\underline{\underline{A}}^{-1}(\Omega)\left(\underline{\underline{1}} - \underline{\underline{B}}\,\underline{\underline{A}}^{-1}(\Omega)\right)^{-1}\right]\tilde{\underline{\varrho}}(\boldsymbol{q}). \tag{5.15}$$

Due to the limitation that $\underline{\underline{A}}$ must be a diagonal matrix, there exist only four straightforward possibilities:

1. $\underline{\underline{A}}(\Omega) = \Omega\underline{\underline{1}}$, $\underline{\underline{B}} = \underline{\underline{H}}$,

2. $\underline{\underline{A}}(\Omega) = \Omega\underline{\underline{1}} - \underline{\underline{H}}^0$, $\underline{\underline{B}} = \underline{\underline{H}} - \underline{\underline{H}}^0$,

3. $\underline{\underline{A}}(\Omega) = \Omega\underline{\underline{1}} - \mathrm{diag}(\underline{h})$, $\underline{\underline{B}} = \underline{\underline{H}} - \mathrm{diag}\left(\underline{h}\right)$,

4. $\underline{\underline{A}}(\Omega) = \Omega\underline{\underline{1}} - \mathrm{diag}\left(\underline{\underline{H}}\right)$, $\underline{\underline{B}} = \underline{\underline{H}} - \mathrm{diag}\left(\underline{\underline{H}}\right)$.
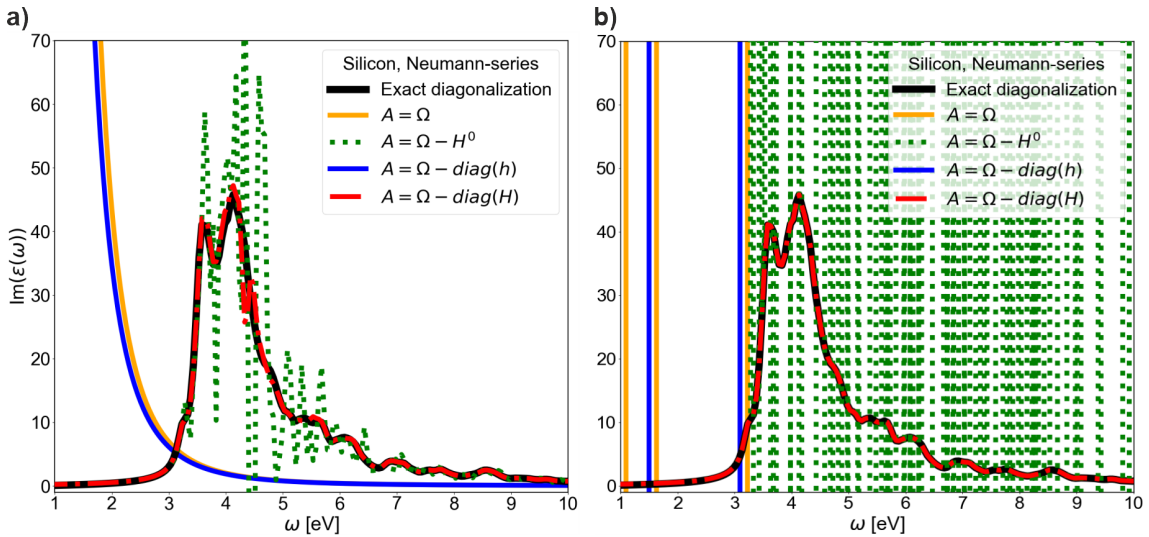

Figure 5.3: Neumann series expansion (Eq. (5.14)) for the imaginary part of the dielectric function $\mathrm{Im}(\epsilon(\omega))$ (Eq. (5.15)) of silicon in static GW-BSE. Shown is the behavior in **a)** third and **b)** one hundered order for different approaches of the unperturbed matrix $\underline{\underline{A}}$, $\underline{\underline{A}} = \Omega\underline{\underline{1}}$ (orange), $\underline{\underline{A}} = \Omega\underline{\underline{1}} - \mathrm{diag}(\underline{h})$ (blue), $\underline{\underline{A}} = \Omega\underline{\underline{1}} - \underline{\underline{H}}^0$ (green) and $\underline{\underline{A}} = \Omega\underline{\underline{1}} - \mathrm{diag}(\underline{\underline{H}})$ (red). Only the choice of the unperturbed matrix as $\underline{\underline{A}} = \Omega\underline{\underline{1}} - \mathrm{diag}(\underline{\underline{H}})$ approaches to the exact result (black). Numerical values are given in Appendix A; calculations were performed on a 3k grid (108 k-points) using a Kohn-Sham band structure, with 6 transition bands for the optical calculation and a Lorentzian broadening of $\Gamma = 3.67 \cdot 10^{-3}$ Ha.

From figure 5.3, we observe that the ansatz based on Ω (orange) gives a function that is proportional to $1/\Omega$ and is incapable of reproducing a spectrum. A similar outcome is seen for $\Omega - \mathrm{diag}(h)$ (blue),

representing only a slight shift compared to Ω. We can already discern an absorption spectrum if we exclude the independent-particle transition energies $\Omega - H_0$ (Fig. 5.3 a) green). However, the result exhibits significant oscillations around the target spectrum, and it diverges with increasing orders (Fig. 5.3 b) green). Nevertheless, for small ω ($\omega < 3$ eV), the result approaches the exact result. Going one step further and excluding the entire diagonal part of the matrix, $\Omega - \text{diag}(H)$ (red), which is equivalent to taking the off-diagonal part of the Hamiltonian as the perturbation, shows that after three orders (Fig. 5.3 a) red), the spectrum already closely matches the target spectrum and the spectrum is stable even at one hundred orders (Fig. 5.3 b) red).

The result indicates that the Neumann algorithm can reproduce the matrix inversion accurately for calculating the absorption spectrum of Si. However, at this point, we do not understand why the different approaches lead to the results as they do. Since we have already stated that the spectral radius is a limiting factor for the algorithm's functionality, we can use this quantity to gain some inside. To do so, it is necessary to provide a detailed analysis of the spectral radius depending on the choice of the matrix $\underline{\underline{A}}$:

$$S_r(\underline{\underline{B}}\,\underline{\underline{A}}^{-1}(\Omega)) := \max|\lambda(\underline{\underline{B}}\,\underline{\underline{A}}^{-1}(\Omega))| = \sqrt{\text{Re}(\max(\lambda(\underline{\underline{B}}\,\underline{\underline{A}}^{-1}(\Omega)))^2 + \text{Im}(\max(\lambda(\underline{\underline{B}}\,\underline{\underline{A}}^{-1}(\Omega)))^2}.$$

To keep the calculation as general as possible, we assume an arbitrary diagonal choice $\underline{\underline{D}}$, $\underline{\underline{A}} = \Omega\underline{\underline{1}} - \underline{\underline{D}}$, which consequently leads to the definition of the perturbation part as $\underline{\underline{B}} = \underline{\underline{H}} - \underline{\underline{D}}$. With these choices, we can rewrite the equation for the spectral radius and make an estimate:

$$S_r\left(\underline{\underline{B}}\,\underline{\underline{A}}^{-1}(\Omega)\right) \leqslant S_r\left(\underline{\underline{B}}\right) \cdot S_r\left(\underline{\underline{A}}^{-1}(\Omega)\right) = S_r\left(\underline{\underline{H}} - \underline{\underline{D}}\right) \cdot S_r\left((\underline{\underline{\Omega}} - \underline{\underline{D}})^{-1}\right), \tag{5.16}$$

$$S_r\left((\underline{\underline{\Omega}} - \underline{\underline{D}})^{-1}\right) = \max\left|\lambda\left(\frac{1}{\underline{\underline{\Omega}} - \underline{\underline{D}}}\right)\right| = \frac{1}{\min|\lambda\left(\underline{\underline{\Omega}} - \underline{\underline{D}}\right)|}, \tag{5.17}$$

$$S_r\left((\underline{\underline{\Omega}} - \underline{\underline{D}})^{-1}\right) \geqslant \left(S_r\left(\underline{\underline{\Omega}}\right) + S_r\left(\underline{\underline{D}}\right)\right)^{-1} = \left(\max|\lambda(\underline{\underline{\Omega}})| + \max|\lambda(\underline{\underline{D}})|\right)^{-1} = \frac{1}{\sqrt{\omega^2 + \Gamma^2} + \max|\lambda(\underline{\underline{D}})|}. \tag{5.18}$$

The estimation shows that the spectral radius is divided into two parts, one $S_r\left((\underline{\underline{\Omega}} - \underline{\underline{D}})^{-1}\right)$, which is directly influenced by the choice of the frequency and the broadening, and one $S_r\left(\underline{\underline{H}} - \underline{\underline{D}}\right)$ determined by the Hamiltonian and the choice of the diagonal part. With the obtained estimation we can identify why the choices $\underline{\underline{A}}(\Omega) = \Omega\underline{\underline{1}} \Leftrightarrow \underline{\underline{D}} = 0 \cdot \underline{\underline{1}}$ and $\underline{\underline{A}}(\Omega) = \Omega\underline{\underline{1}} - \text{diag}(\underline{h}) \Leftrightarrow \underline{\underline{D}} = \text{diag}(\underline{h})$ cannot work: If ω becomes too small, e.g. $\omega = 0$ eV, and $\max|\lambda(\underline{\underline{D}})|$ is either zero ($\underline{\underline{D}} = 0$) or very small ($\underline{\underline{D}} = \text{diag}(\underline{h})$), then the frequency dependent part of the spectral radius is only determined by broadening as we can immediately see from equation (5.17):

$$S_r\left((\underline{\underline{\Omega}} - \underline{\underline{D}})^{-1}\right) = \frac{1}{\min|\lambda\left(\underline{\underline{\Omega}} - \underline{\underline{D}}\right)|} \xrightarrow{D \to 0,\, \omega \to 0} \frac{1}{\Gamma}$$

However, since the broadening is in the order of magnitude of 10^{-3}, the spectral radius will become very large, and the series expansion will diverge, which we observed in figure 5.3. In order to give a general estimate of when the spectral radius is as small as possible, independently of the chosen broadening or frequency, we want to reformulate the estimation into three conditions:

1. $S_r\left((\underline{\underline{\Omega}} - \underline{\underline{D}})^{-1}\right)$ must be small, implying $\underline{\underline{D}}$ is large, since $\max|\lambda(\underline{\underline{D}})|$ must be large,

2. Ω must be far away from the values of the diagonal matrix $\underline{\underline{D}}$ so that $\min|\lambda\left(\underline{\underline{\Omega}} - \underline{\underline{D}}\right)|$ is large

3. $S_r \left(\underline{H} - \underline{D} \right)$ must be small, meaning $\underline{D}$ must represent a significant portion of $\underline{H}$.

These conditions define a framework within which we can choose the D-matrix and which frequencies or broadening we can choose. In order to satisfy condition 1 and 3 optimally, a large proportion of the diagonal part of the Hamiltonian must be extracted. The reason for this choice is that we have to keep $S_r \left(\underline{H} - \underline{D} \right)$ as small as possible, which means that we have to take out a large part of the Hamiltonian and additionally for all Ω, the spectral radius should be as small as possible, which requires that $\underline{D}$ is as large as possible, especially more significant than the broadening, if the broadening is small so that no divergence occurs in the low-frequency range. Condition 2 imposes another additional condition, which concerns the choice of frequencies and broadening. Ω must be as far away as possible from the values of the diagonal matrix $\underline{D}$, which in turn means that the broadening should be as large as possible and/or ω should be as far away as possible from the values of the diagonal matrix.

However, whether choice $\underline{D} = \underline{H}^0$ or $\underline{D} = \text{diag}(\underline{H})$ is the better decision is not directly evident from these conditions, since $\underline{H}^0 > \text{diag}(\underline{H})$ holds in the case of the resonant part of the effective two particle Hamiltonian (Eq. (5.3)). Therefore, we have an interplay between the two conditions (1 and 3): If we chose $\underline{D} = \underline{H}^0$ than condition 1 will be better fulfilled and with the choice $\underline{D} = \text{diag}(\underline{H})$ condition 3.
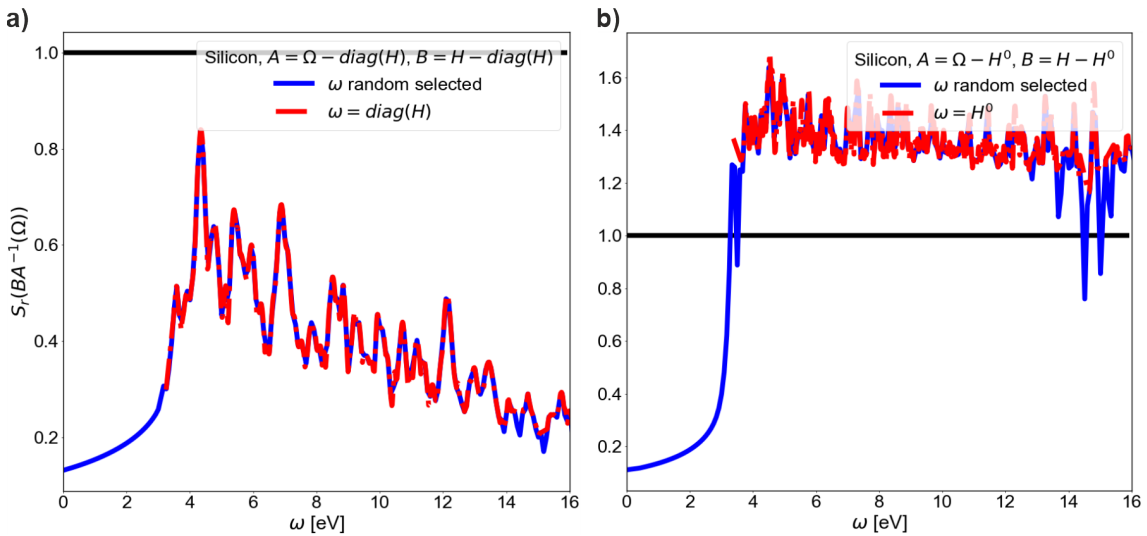

Figure 5.4: Spectral radius S_r for the matrix $\underline{B}\underline{A}^{-1}$ inside the imaginary part of the macroscopic dielectric function $\text{Im}(\epsilon(\omega))$ (Eq. (5.15)) for silicon, for the choice **a)** $\underline{A} = \Omega\underline{1} - \text{diag}(\underline{H})$, $\underline{B} = \underline{H} - \text{diag}(\underline{H})$ and **b)** $\underline{A} = \Omega\underline{1} - \underline{H}^0$, $\underline{B} = \underline{H} - \underline{H}^0$. Results are shown for both cases for the choice of frequency: blue random choice and red $\omega = \text{diag}(H)$ (a)) or $\omega = H^0$ (b)). It turns out that in both cases the choice of frequency is insignificant, but the choice for the diagonal component as $\underline{D} = \text{diag}(\underline{H})$ provides a significantly lower spectral radius than the choice $\underline{D} = \underline{H}^0$. Numerical values are given in Appendix A; calculations were performed on a 3k grid (108 k-points) using a Kohn-Sham band structure, with 6 transition bands for the optical calculation and a Lorentzian broadening of $\Gamma = 3.67 \cdot 10^{-3}$ Ha.

Figure 5.4 shows that choice $\underline{D} = \text{diag}(\underline{H})$ provides a significantly lower spectral radius. How-

ever, we could not get this result directly from our conditions for choosing the diagonal matrix. Furthermore, the results show us that the choice of the frequency spectrum is not crucial because even if we choose the frequency identical to the diagonal matrix elements, we still get almost the same result as in the case of a randomly chosen frequency spectrum. This may be attributed to the non-disappearing broadening, which ensures that $\underline{\underline{\Omega}} \neq \underline{\underline{D}}$ holds. Therefore broadening serves a similar purpose for the Neumann algorithm as it did in the Haydock-Lanczos; it prevents the frequency from getting too close to a certain value.

Based on our mathematical justification and illustration of the different approaches, we will focus solely on what we will refer to as the correct ansatz:

$$\epsilon(\omega) = 1 + \lim_{\boldsymbol{q} \to \boldsymbol{0}} v_c(\boldsymbol{q}) \tilde{\underline{\rho}}^\dagger(\boldsymbol{q}) \left[\underline{\underline{\Delta E}}_\Omega^{-1} \left(\underline{\underline{1}} - \tilde{\underline{\underline{h}}} \, \underline{\underline{\Delta E}}_\Omega^{-1} \right)^{-1} \right] \tilde{\underline{\rho}}(\boldsymbol{q}), \tag{5.19}$$

where $\tilde{\underline{\underline{h}}} := \underline{\underline{H}} - \text{diag}(\underline{\underline{H}}) = \underline{\underline{h}} - \text{diag}(\underline{\underline{h}})$ and $\underline{\underline{\Delta E}}_\Omega^{-1} := (\Omega \underline{\underline{1}} - \text{diag}(\underline{\underline{H}}))$. Now that we have figured out how we have to rewrite the matrix inversion in order to be able to set up a perturbation series, we want to deal with the same question as in the case of the Haydock-Lanczos algorithm: *What role does broadening play for the algorithm, and is the algorithm independent of the k-grid density?* If we take a look at our estimation for the spectral radius (Eq. (5.16), (5.18)), we already know that the broadening factor will play a decisive role for the size of the spectral radius, especially the smaller it gets, the larger the spectral radius should become. To give this assumption more substance, we want to look at the spectral radius for Si depending on the k-grid density and the broadening.
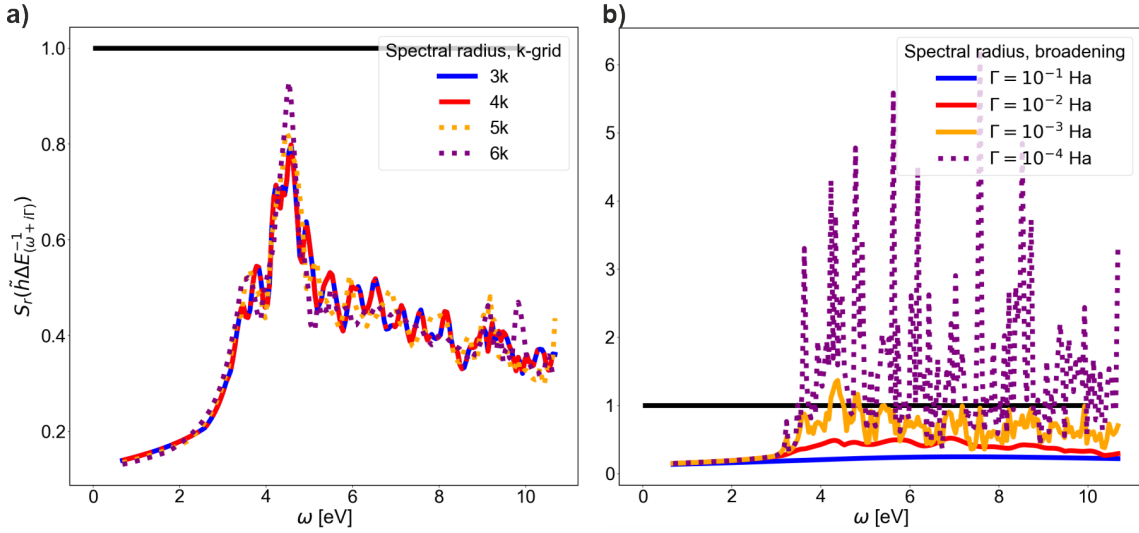

Figure 5.5: Behavior of the spectral radius S_r for the matrix $\tilde{\underline{\underline{h}}} \, \underline{\underline{\Delta E}}_\Omega^{-1}$ inside the imaginary part of the dielectric function $\text{Im}(\epsilon(\omega))$ (Eq. (5.19)). Results for silicon in static GW-BSE shown for **a)** the k-grid density for a fixed Lorentzian broadening $\Gamma = 3.67 \cdot 10^{-3}$ Ha and **b)** the Lorentzian broadening Γ for a fixed k-grid (3k). The results indicate that the spectral radius is almost independent of the k-grid density but strongly depends on the broadening. Numerical values are given in Appendix A; calculations were performed with 6 transition bands for the optical calculation, using a Kohn-Sham band structure.

As we can see from figure 5.5 a), the Neumann algorithm for Si is almost independent of the

density of the k-grid over the entire energy range, as it was the case for the Haydock-Lanczos (Fig. 5.2 a)). However, the same cannot be said about broadening. We can observe (Fig. 5.5 b)) that if we change the broadening, we see precisely the same as for the Haydock-Lanczos (Fig. 5.2 b)). The convergence speed becomes slower with decreasing broadening (larger spectral radius). In the case of the Neumann series, however, we have another disadvantage to contend with. If the spectral radius is larger than one (orange, purple), the series will not converge anymore. This leads to a limitation for the application possibilities of a perturbative series expansion. The reason for the strong broadening dependency of the algorithm can be explained by the analysis we have already carried out regarding the dependence of the spectral radius on the selected frequency spectrum while maintaining the same broadening (Fig. 5.4). As the broadening becomes smaller, the frequencies lie closer to the values of the diagonal matrix $\underline{\underline{D}}$, and a divergence of the contribution $S_r(\underline{\underline{\Omega}} - \underline{\underline{D}})$ in the spectral radius occurs whenever Ω comes too close to a value of $\underline{\underline{D}}$. For this reason, the frequency spectrum choice becomes crucial for a series expansion, as it actively determines the spectral radius.

Now that we have dealt extensively with the possible approaches to a perturbation series, particularly the limiting factors, it is time to bring more physics into our analysis. To do this, we want to look at how the algorithm behaves in practice, i.e., how an absorption spectrum is built up piece by piece and what we can expect for a perturbation series depending on the strength of the excitonic effects.
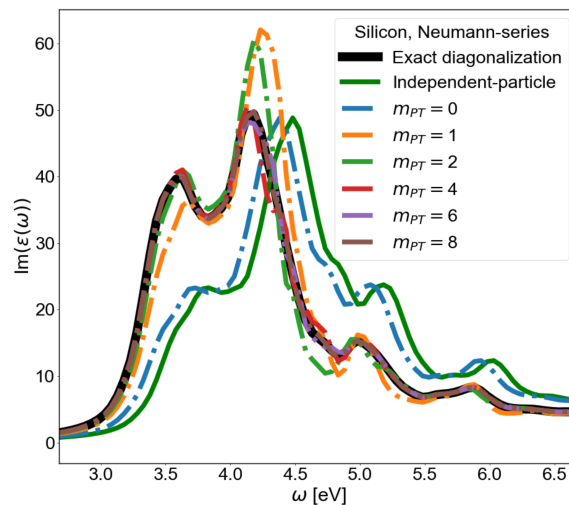

Figure 5.6: Neumann series expansion (Eq. (5.14)) for the imaginary part of the dielectric function $\text{Im}(\epsilon(\omega))$ (Eq. (5.19)) of silicon in static GW-BSE. Shown are the IP approximation (Eq. (3.35)) (darkgreen) and the exact result (black), as well as the Neumann series up to the eighth order. An approach of the series towards the exact result can be seen between the sixth and eighth order. Numerical values are given in Appendix A; calculations were performed on a 5k grid (500 k-points) using a Kohn-Sham band structure, with 6 transition bands for the optical calculation and a Lorentzian broadening of $\Gamma = 3.67 \cdot 10^{-3}$ Ha.

As seen in figure 5.6, there is no question that the Neumann series for Si approaches the exact result, and after about six (purple) to eight (brown) orders, there is almost no difference to the

exact absorption spectrum. However, the more exciting result only emerges from a closer look at the individual orders. In order to support the discussion on mathematical foundations, we want to write out the macroscopic dielectric function (Eq. (5.19)) up to the second order in the perturbation series:

$$
\epsilon(\omega) = 1 + \lim_{\boldsymbol{q} \to \boldsymbol{0}} v_c(\boldsymbol{q}) \underline{\tilde{\rho}}^{\dagger}(\boldsymbol{q}) \left[\underline{\underline{\Delta E}}_{\Omega}^{-1} + \underline{\underline{\Delta E}}_{\Omega}^{-1} \underline{\underline{\tilde{h}}}\, \underline{\underline{\Delta E}}_{\Omega}^{-1} + \underline{\underline{\Delta E}}_{\Omega}^{-1} \underline{\underline{\tilde{h}}}\, \underline{\underline{\Delta E}}_{\Omega}^{-1} \underline{\underline{\tilde{h}}}\, \underline{\underline{\Delta E}}_{\Omega}^{-1} + \dots \right] \underline{\tilde{\rho}}(\boldsymbol{q}).
$$

If we compare the equation with the respective orders in figure 5.6, we see the following behavior: The zeroth order (blue) represents nothing other than a shifted IP spectrum and is only constructed by the diagonal matrix $\underline{\underline{\Delta E}}_{\Omega}^{-1}$. The reason for the commonality between the zeroth order (blue) and the IP spectrum (dark green), lies in the fact that the only difference is that we have $(\Omega \underline{\underline{1}} - \mathrm{diag}(\underline{\underline{H}}))$ instead of $(\Omega \underline{\underline{1}} - \underline{\underline{H}}^0)$. Therefore the difference in including the diagonal part of the interaction Hamiltonian $(\mathrm{diag}(\underline{\underline{h}}) = \mathrm{diag}(\underline{\underline{\Xi}}^{AGW}))$ or not creates a shift towards lower energies. The next, the first order of the perturbation series (orange), no longer describes a diagonal matrix but instead brings pure off-diagonal elements into play. By taking the off-diagonal elements into account, the characteristic peak structures that arise from the excitonic effects become visible. However, the spectrum has not yet been correctly described. The continuum part of the spectrum tends to be overestimated, and the bound states (first peak) near the optical edge tend to be described too weakly. For this reason, including the first order still leads to an absorption spectrum, which looks like an increased IP spectrum.

By taking further orders into account (diagonal and off-diagonal elements), this imbalance is eliminated by weakening the spectrum's continuum and increasing that of the bound states.

Unlike the Haydock-Lanczos algorithm, the Neumann series allows a relatively precise analysis of the excitonic effects and which parts of the Hamiltonian are responsible for what.

However, before we compare the two algorithms in more detail, we will look at how a perturbation series behaves for other materials and how the series expansion depends on the strength of the excitonic effects. For this purpose, we only look at the spectral radius of the materials, as this tells us whether a series development is possible or not.

In the following, we want to look at the four materials: Ge, Si, ZnS and LiF, since with the choice of the four materials, we cover the range from very weak to very strong excitonic effects. This is because they exhibit increasing excitonic effects in ascending order, which stems from the fact that, among other things, they have weaker screening in ascending order ((Ge, $\epsilon_{mac} \sim 16$ [114]), (Si, $\epsilon_{mac} = 11.4$ [61]), (ZnS, $\epsilon_{mac} = 4.72 - 5.1$ [115, 116]), (LiF, $\epsilon_{mac} = 1.9$ [62])), and therefore, stronger excitonic effects are possible (Ch. 4.6).
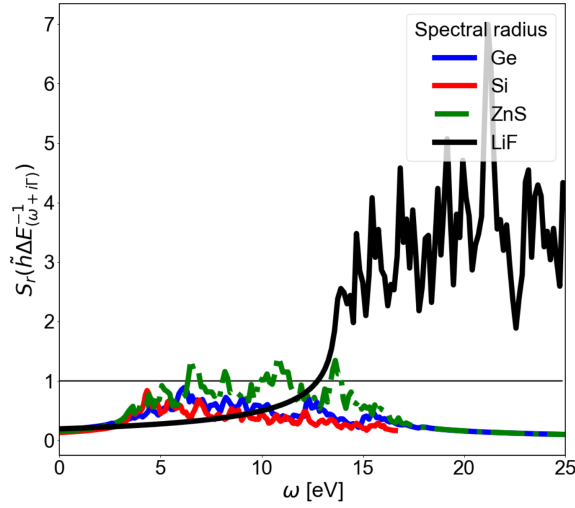
Figure 5.7: Spectral radius S_r for the matrix $\underline{\tilde{\underline{h}}}\Delta\underline{E}_\Omega^{-1}$ inside the imaginary part of the dielectric function $\text{Im}(\epsilon(\omega))$ (Eq. (5.19)). Results are shown for germanium (blue), silicon (red), zinc sulphide (green) and lithium fluoride (black) in static GW-BSE. It can be seen that with increasing strength of the excitonic effects from germanium to lithium fluoride, the spectral radius increases and exceeds one by far for lithium fluoride. Numerical values are given in Appendix A; calculations were performed on a 3k grid (108 k-points) using a Kohn-Sham band structure, with a Lorentzian broadening of $3.67 \cdot 10^{-3}$ Ha, and 6 transition bands for the optical calculation.

As we can see in figure 5.7, as the strength of the interaction increases, so does the spectral radius. The spectral radius for Ge (blue) and Si (red) is below one for all energies. For ZnS (green), some energies already exceed this limit, and for LiF (black), the spectral radius is clearly above one. The behavior of the spectral radius can be explained, if we take another look at our previous estimate of the spectral radius (Eq. (5.16)):

$$S_r\left(\left(\underline{H} - \text{diag}(\underline{H})\right)\left((\underline{\Omega} - \text{diag}(\underline{H}))^{-1}\right)\right) \leqslant S_r\left(\underline{H} - \text{diag}(\underline{H})\right) \cdot S_r\left((\underline{\Omega} - \text{diag}(\underline{H}))^{-1}\right).$$

Since the second frequency-dependent component of the spectral radius is independent of the strength of the excitonic effects because only the minimum distance (Eq. (5.17)) $\min|\underline{\Omega} - \text{diag}(\underline{H})|$ between the complex frequency and the diagonal matrix elements matters for the spectral radius. It, therefore, is the first term $(\underline{H} - \text{diag}(\underline{H}))$ which changes with the excitonic effects.

In order to make the whole thing more transparent to use, we will play around with the simple Gedankenexperiment: Suppose we could represent the strength of the excitonic effects by a single constant c so that we can rewrite our estimate as:

$$S_r\left(\left((\underline{H}^0 + c \cdot \underline{h}) - \text{diag}\left(\underline{H}^0 + c \cdot \underline{h}\right)\right)\left(\underline{\Omega} - \text{diag}\left(\underline{H}^0 + c \cdot \underline{h}\right)\right)^{-1}\right) \leqslant$$

$$S_r\left(c \cdot \text{diag}(\tilde{\underline{h}})\right) \cdot S_r\left(\left(\underline{\Omega} - \text{diag}\left(\underline{H}^0 + c \cdot \underline{h}\right)\right)^{-1}\right).$$

For the second term, the frequency-dependent part of the spectral radius, we can observe that only the frequency range at which the minimum difference lies has now shifted. However, the value itself for the minimum does not shift, since for a given c, we can always find an $\tilde{\Omega}$, so that:

$$\min|\underline{\Omega} - \text{diag}(\underline{H})| = \min|\tilde{\underline{\Omega}} - \text{diag}\left(\underline{H}^0 + c \cdot \underline{h}\right)|.$$

However, the same does not hold in the case of the first component, as we can see if we rewrite it:

$$S_r \left(\left(\underline{\underline{H}}^0 + c \cdot \underline{\underline{h}} \right) - \mathrm{diag} \left(\underline{\underline{H}}^0 + c \cdot \underline{\underline{h}} \right) \right) = \cdot S_r \left(c \cdot \mathrm{diag}(\underline{\underline{\tilde{h}}}) \right) = c \cdot S_r \left(\mathrm{diag}(\underline{\underline{\tilde{h}}}) \right).$$

The proportion scales with c, which in turn means that with increasing strength of the excitonic effects, i.e. increasing c values, the spectral radius will increase. In summary, the observation could be plugged into a simple rule:

The stronger the excitonic interaction, the stronger the off-diagonal elements, and the larger the spectral radius.

However, this is not the only important observation we can make in figure 5.7. We can also see a pattern between the materials namely, that the spectral radius drops drastically below the value one for frequencies within the optical bandgap ($\omega < \min(E_\lambda)$. The reason for the pattern can be traced back to the fact that: $\min |\underline{\underline{\Omega}} - \mathrm{diag}(\underline{\underline{H}})| \rightarrow \min |\lambda \left(\mathrm{diag}(\underline{\underline{H}}) \right) |$, with $\omega \rightarrow 0$ and $\Gamma \ll \min |\lambda \left(\mathrm{diag}(\underline{\underline{H}}) \right) |$ holds. In turn, this means that the spectral radius runs towards a constant value:

$$S_r \left(\left(\underline{\underline{H}} - \mathrm{diag}(\underline{\underline{H}}) \right) \cdot \mathrm{diag}(\underline{\underline{H}})^{-1} \right) \leqslant S_r \left(\underline{\underline{H}} - \mathrm{diag}(\underline{\underline{H}}) \right) \cdot S_r \left(\mathrm{diag}(\underline{\underline{H}})^{-1} \right) = \frac{S_r \left(\underline{\underline{\tilde{h}}} \right)}{\min |\lambda \left(\mathrm{diag}(\underline{\underline{H}}) \right) |}.$$

If we calculate the spectral radius for Si, ZnS and LiF we observe what we expected: An increase in the interaction leads to an increase in the spectral radius. However, for Ge, it does not, as can be seen in the table below. The reason for this may be a relic of the fact that we work with a low-band calculations.

$\omega = 0$ eV	Ge	Si	ZnS	LiF
$S_r \left(\underline{\underline{\tilde{h}}} \underline{\underline{\Delta E}}^{-1}_{(\omega + i\Gamma)} \right)$	0.18	0.13	0.15	0.19

Before we delve deeper into the subject of excitons, we will briefly compare the Haydock-Lanczos with the Neumann series. As we have seen, both algorithms are, in principle, suitable for correctly predicting at least absorption spectra for materials with weak excitonic effects.

5.1.3   Haydock-Lanczos vs Neumann algorithm

As we saw in the last section, in the case of Si and presumably for many materials with weak excitonic effects, the Neumann series will likely approach the exact result after only a few orders. Furthermore, expanding the modified dielectric susceptibility in a perturbation series allows us to take a deeper look at the excitonic effects and better understand how they are built up, especially which information is essential.

However, it is not the Neumann series, which is generally used in the community, but the Haydock-Lanczos algorithm. In the following, we will take a closer look at the numerics and see which algorithm has a practical application in which case. For this reason, we first want to ask ourselves: *What is the numerical scaling of the two algorithms?* Instead of restricting ourselves as before to the macroscopic dielectric function, we want to estimate the numerical complexity for the modified dielectric susceptibility $\bar{\chi}_{\boldsymbol{G},\boldsymbol{G}'}(\boldsymbol{q}, \omega)$, respectively the dielectric susceptibility $\chi_{\boldsymbol{G},\boldsymbol{G}'}(\boldsymbol{q}, \omega)$, which is linked to the modified dielectric susceptibility by $\bar{\chi}^{-1} = \chi^{-1} + v_c^0$ (Ch. 4.1).

For the estimation of the numerical complexity, we will only focus on the highest power in the numerical scaling, meaning $O(N^n + N^{n-1}) \rightarrow O(N^n)$, in order not to make things unnecessarily complicated. Furthermore, we assume that we are only interested in single values of the spectrum, for example the case $\omega = 0$ eV, which is more often needed, e.g. for the calculation of the static dielectric function for self-consistent BSE, i.e. beyond the RPA approximation (Ch. 3.4.3) or for GW calculations in the plasmon pole model (Ch. 4.5.4). Additionally, we assume that we need for all elements of the susceptibility as well as for all values of $\boldsymbol{q}$ the same amount of iterations once for the Haydock-Lanczos m_{Hay} and for the Neumann algorithm m_{PT}.

Numerical scaling:

Diagonal elements ($\boldsymbol{G} = \boldsymbol{G}'$) and off-diagonal elements ($\boldsymbol{G} \neq \boldsymbol{G}'$) of $\bar{\chi}_{\boldsymbol{G},\boldsymbol{G}'}(\boldsymbol{q},\omega)$:

1. Haydock: $O\left(m_{Hay} \cdot N^2\right)$, with m_{Hay} iterations.

2. Neumann: $O\left(m_{PT} \cdot N^2\right)$, with m_{PT} iterations.

We see that, in principle, both algorithms behave similarly. However, this is only partially correct because if we were interested in many frequencies, then the Haydock-Lanczos would tend to offer an advantage. The reason lies in the representation of the Haydock-Lanczos as a continued fraction (Eq. (5.13)) and the corresponding expansion in the Lanczos basis (Eq. (5.9)). In principle, this representation means that only the scalar values a_n and b_n have to be calculated and stored, and the series then only has to be calculated for each Ω. Since the scalar values generate the numerical scaling, it can generally be assumed that the Haydock-Lanczos algorithm will be more efficient than the Neumann series, which requires a series expansion for each Ω.

However, we do not want to use the Neumann series to calculate the entire spectrum, but only individual values and to better understand the structure of the excitonic effects. Furthermore, scaling is only one part of the numerical efficiency. Both algorithms would, for example, also make it possible to switch to a more efficient basis and thus may reduce the numerical costs or carry out parallelization, which is much easier to achieve for the Neumann series than for the Haydock-Lanczos algorithm. Therefore, a direct comparison of the two algorithms in CPU or GPU time is not part of this work, as it is not about rewriting the algorithms into a high-performance version. However, we want to finally look at the number of iterations necessary to achieve convergence of the Haydock-Lanczos and the Neumann algorithm for the real part of the macroscopic dielectric function for the three selected frequency points $\omega = 0$ eV, $\omega = 5.0$ eV and $\omega = 10$ eV for Si on a 5k-grid and a Lorentzian broadening of $3.67 \cdot 10^{-3}$ Ha. For this purpose, we define $\Delta\epsilon^{mac} :=$ $\mathrm{Re}\left(\epsilon^{mac,\mathrm{diagonalization}} - \epsilon^{mac,\mathrm{algorithm}}\right)$ and the full macroscopic dielectric function, including the antiresonant term:

$$\epsilon(\omega) = 1 + \lim_{\boldsymbol{q}\to\boldsymbol{0}} v_c(\boldsymbol{q})\tilde{\underline{\rho}}^{\dagger}(\boldsymbol{q})\left[\underline{\underline{\Delta E}}^{-1}_{\Omega}\left(\underline{\underline{1}} - \underline{\tilde{h}}\,\underline{\underline{\Delta E}}^{-1}_{\Omega}\right)^{-1} + \underline{\underline{\Delta E}}^{-1}_{-\Omega}\left(\underline{\underline{1}} + \underline{\tilde{h}}\,\underline{\underline{\Delta E}}^{-1}_{-\Omega}\right)^{-1}\right]\tilde{\underline{\rho}}(\boldsymbol{q}), \quad (5.20)$$

where the first term represents the resonant part and the second the antiresonant.

To calculate the values, we define a result as converged when an accuracy of $\Delta\epsilon^{mac} = 10^{-4}$ or less has been achieved, and the subsequent orders will lead to an even more accurate result. The necessary iteration steps can be found in the table.

ω [eV]	Neumann: m_{PT} iterations	Haydock-Lanczos: m_{Hay} iterations
0.0	5	4
5.0	1	250
10.0	1	255

As the table shows us, both algorithms require almost the same number of iterations for $\omega = 0$ eV. However, the Neumann series only requires one iteration for higher frequencies, whereas the Haydock-Lanczos requires around 250 iterations. This illustrates that the Neumann algorithm for Si requires either almost the same number of iteration steps or significantly fewer than the Haydock-Lanczos.

If we summarize the result, we can assume that the Neumann algorithm is advantageous for calculating absorption spectra for materials with relatively weak excitonic effects. Furthermore, the algorithm allows us to conduct a mathematical analysis of the excitonic effects concerning the absorption spectrum. As we will see later (Ch. 5.5, Ch. 5.6.2), this possibility will be invaluable if we are interested not only in the absorption spectrum but also in the eigenvalues and eigenvectors of the excitons.

We also showed that if only individual information, for example, the dielectric function or susceptibility for individual frequencies, is necessary, the Neumann series requires significantly fewer iteration steps than the Haydock-Lanczos algorithm and, therefore, lower numerical scaling until the result can be viewed as converged. The Neumann series is, therefore, not only valuable for analytical purposes but also for practical applications.

From now on, our goal is not to transform this algorithm into a high-performance one but to gain an even deeper insight into the excitonic effects. For this reason, we will deal in detail with the eigenvalues and eigenvectors, both analytically and numerically.

5.2   Exact eigenvector and eigenvalue approach

The current section is the heart of this work and deals with a derivation of an equation for the eigenvectors and eigenvalues beyond the eigenvalue equation. The advantage of such an approach will be that new ways of calculating eigenvalues and eigenvectors become possible. Furthermore, analytic expressions are always a warm welcome for analyses, as we want and will do for the excitonic effects (Ch. 5.3).

Within this section, we will derive three exact eigenvector structures, each with its peculiarity. Based on the equations for the eigenvectors, we will present a specific equation for the eigenvalues. It should be noted that we are not aware that such a path has already been taken in the field of ab inito BSE-calculations at the time of this work. However, attempts have already been made in many-body physics to derive an approximate version of the eigenvector via perturbation theory [22]. Later in this work (Ch. 5.5) , we will show that such an approximation should be treated with caution and is only valid under certain circumstances.

To begin our journey, we want to ask ourselves the question: *Is it possible to derive an equation for the eigenvectors beyond the eigenvalue equation?*

5.2.1 The exact eigenvector approach

Generally, deriving a new exact equation beyond the eigenvalue equation is not trivial. In order to do so, we use a trick. We use applied mathematics, i.e. numerics, to see how the eigenvalue equation is generally solved. As a next step, we do not write down an algorithm in its steps but formulate it as a single mathematical equation. To follow this path, we start by defining a general eigenvalue equation as:

$$\sum_{\mu'} H_{\mu\mu'} A_\lambda^{\mu'} = E_\lambda A_\lambda^\mu, \quad H_{\mu\mu'} := E_{\mu\mu'}^0 \delta_{\mu\mu'} + h_{\mu\mu'}, \tag{5.21}$$

where in the special case of the resonant part of the two-particle Hamiltonian (Eq. (4.12)), $h_{\mu\mu'} = \Xi_{\mu\mu'}^{AGW}$ holds. However, since we want to keep the discussion general, we will only write $h_{\mu\mu'}$ in the following, even if we evaluate the results for the resonant part of the two-particle Hamiltonian. Furthermore, we will use the following notation: For the component $\mu \in \{1, N\}$ of a vector $\underline{A}_\lambda$ associated with the λ eigenvalue E_λ ($\lambda \in \{1, N\}$) we use the notation A_λ^μ. For the μ, μ' component of a general complex valued matrix $\underline{\underline{H}} \in \mathbb{C}^{N \times N}$ we use the notation $H_{\mu\mu'}$.

The key to deriving equations beyond the eigenvalue equation will now be to start with a suitable algorithm which is straightforward in its structure.

The Gaussian algorithm is a standard algorithm that is easy to understand and can be used to solve a system of linear equations, i.e. an eigenvalue equation. For these reasons, the Gaussian algorithm could be an ideal starting point.

To familiarize us with the algorithm, we will start with a brief introduction of the algorithm by explaining the most critical steps and conditions for the Gaussian algorithm:

1. Since $\det(\underline{\underline{H}} - E_\lambda \underline{\underline{1}}) = 0$ holds, the rank of the matrix is reducible by one, which allows a zero row to be generated,

2. The Gaussian algorithm aims to transform the matrix into a triangular form so that the eigenvector entries can be read off directly by inserting from bottom to top.

A more detailed description and illustration of the algorithm can be found in Appendix E. Point (1) and (2) are sufficient for our purposes.

We will now use condition (1) and (2) in order to rewrite our system of equations for the eigenvalue equation (Eq. (5.21)):

$$\sum_{\mu'} \left(\underline{\underline{H}} - E_\lambda \underline{\underline{1}}\right)_{\mu\mu'} A_\lambda^{\mu'} \rightarrow \begin{pmatrix} E_{11}^0 + h_{11} - E_\lambda & h_{12} & \cdots & h_{1N} \\ h_{21} & E_{22}^0 + h_{22} - E_\lambda & \cdots & h_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N1} & h_{N2} & \cdots & E_{NN}^0 + h_{NN} - E_\lambda \end{pmatrix} \begin{pmatrix} A_\lambda^1 \\ A_\lambda^2 \\ \vdots \\ A_\lambda^N \end{pmatrix} = 0,$$

into a single equation. In order to solve the system of equations, we choose, according to condition (1), without loss of generality, the entry $\mu = \lambda$ as our zero line. This, in turn, means we can choose the associated element of the eigenvector as:

$$A_\lambda^\lambda = A_\mu^\mu = 1.$$

In the case of the other vector components ($\mu \neq \lambda$), however, we proceed somewhat differently than in the Gaussian algorithm. Instead of putting the matrix into a triangular form, we resolve each additional row according to their respective diagonal element ($\mu = \mu'$); therefore, the following applies for the remaining eigenvector entries:

$$A_\lambda^\mu = -\frac{\sum_{\mu' \neq \mu} h_{\mu\mu'} A_\lambda^{\mu'}}{E_\mu^0 + h_{\mu\mu} - E_\lambda} = \frac{\sum_{\mu' \neq \mu} h_{\mu\mu'} A_\lambda^{\mu'}}{E_\lambda - (E_\mu^0 + h_{\mu\mu})} = \frac{\sum_{\mu'} (1 - \delta_{\mu\mu'}) h_{\mu\mu'} A_\lambda^{\mu'}}{E_\lambda - (E_\mu^0 + h_{\mu\mu})}.$$

If we put the two equations together and take into account the normalization of the eigenvectors $N_\lambda = \sqrt{|\underline{A}_\lambda|^2}$, then we get the equation for the normalized eigenvectors as:

$$\bar{A}_\lambda^\mu = \frac{A_\lambda^\mu}{N_\lambda} = \frac{\delta_{\mu\lambda}}{N_\lambda} + \frac{(1 - \delta_{\mu\lambda})}{N_\lambda} \frac{\sum_{\mu'} (1 - \delta_{\mu\mu'}) h_{\mu\mu'} A_\lambda^{\mu'}}{E_\lambda - (E_\mu^0 + h_{\mu\mu})}. \tag{5.22}$$

It is crucial to mention that equation (5.22) is universally valid for all linear systems of equations for which the Gaussian algorithm is also applicable.

To get a better feeling for the obtained equation and to test it empirically, we will consider a simple example of a 2x2 non-hermitian matrix:

$$\underline{\underline{H}} = \begin{pmatrix} 1 & 9 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 0.5 & 0 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 0.5 & 9 \\ 2 & 2 \end{pmatrix} = \underline{\underline{H}}^0 + \underline{\underline{h}} = \begin{pmatrix} E_1^0 & 0 \\ 0 & E_2^0 \end{pmatrix} + \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix}$$

with eigenvalues $E_1 = -2$ and $E_2 = 7$. To determine the eigenfunction according to the Gaussian algorithm, we need to solve the linear system of equations:

$(I)\ (1 - E_\lambda) A_\lambda^1 + 9 A_\lambda^2 = 0$

$(II)\ 2 A_\lambda^1 + (4 - E_\lambda) A_\lambda^2 = 0$

To do so for the eigenvalue $E_1 = -2$, we set the first line (I) to a zero line $A_1^1 = 1$ and for the second line (II) we obtaine $A_1^2 = \frac{2}{(E_1 - 4)} = \frac{h_{21} A_1^1}{E_1 - (E_2^0 + h_{22})}$. In total the eigenvector to the eigenvalue $E_1 = -2$ is then given as $\underline{A}_1^T = (1, \frac{h_{21}}{E_1 - (E_2^0 + h_{22})})$. This result aligns with what we would obtain using equation (5.22) without normalization for the calculation of the eigenvectors.

However, since our equation for the eigenvectors (Eq. (5.22)) comes with the disadvantage that the elements of the eigenvectors must already be known to calculate further elements, we cannot work with the equation in the current form. For this reason, we have to reformulate the obtained eigenvector equation into a usable one.

S-Vector approach

The primary disadvantage of the obtained equation (Eq. (5.22)) is that in order to get one element of the eigenvector, we must already know all the others, so our goal must be to get rid of this dependency. For this reason, we will reformulate the equation. To do so, we start by defining the vector element S_λ^μ as:

$$S_\lambda^\mu := \sum_{\mu'} (1 - \delta_{\mu\mu'}) h_{\mu\mu'} A_\lambda^{\mu'},$$

and the denominator as:

$$\left(\Delta \tilde{E}_{\mu\mu'}(E_\lambda) \delta_{\mu\mu'} \right)^{-1} := \frac{(1 - \delta_{\mu\lambda})}{E_\lambda - (E_\mu^0 + h_{\mu\mu})}.$$

With those two definitions the equation for the eigenvector (Eq. (5.22)) without normalization becomes

$$A_\lambda^\mu = \delta_{\mu\lambda} + \left(\Delta\tilde{E}_{\mu\mu}(E_\lambda)\right)^{-1} S_\lambda^\mu. \tag{5.23}$$

Since our target equation is the eigenvector, we want to transform equation (5.23) into a matrix-vector equation. In oder to obtain a vector equation, we use the fact, that $\left(\Delta\tilde{E}_{\mu\mu'}(E_\lambda)\delta_{\mu\mu'}\right)^{-1}$ is a diagonal matrix and $\delta_{\mu\lambda}$ can be defined as $\underline{1}_\lambda$, e.g. a vector that is zero everywhere except in the row where $\mu = \lambda$ applies, therefore we can write the component equation of the eigenvector into a vector-matrix equation:

$$\underline{A}_\lambda = \underline{1}_\lambda + \underline{\underline{\Delta\tilde{E}}}_\lambda^{-1} \underline{S}_\lambda. \tag{5.24}$$

Because the S-vector still depends on the individual components of the eigenvector, we want to determine $\underline{S}_\lambda$ independent of them. Therefore, our aim must be to obtain a determination equation for $\underline{S}_\lambda$, e.g. a Dyson-like structure. For this reason, we multiply equation (5.23) by $\sum_\mu (1 - \delta_{\mu''\mu})h_{\mu''\mu}$, obtaining:

$$S_\lambda^{\mu''} = \sum_\mu \tilde{h}_{\mu''\mu} A_\lambda^\mu = \tilde{h}_\lambda^{\mu''} + \sum_\mu \tilde{h}_{\mu''\mu} \left(\Delta\tilde{E}_{\mu\mu}(E_\lambda)\right)^{-1} S_\lambda^\mu, \quad \tilde{h}_{\mu\mu'} := (1 - \delta_{\mu\mu'})h_{\mu\mu'}. \tag{5.25}$$

The chosen notation $\tilde{h}_\lambda^{\mu''}$ should not lead to confusion; this notation is identical to $\tilde{h}_\lambda^{\mu''} = \tilde{h}_{\mu''\lambda}$ and should only indicate that this quantity becomes a vector in the further course. Equation (5.25) can now be rewritten into a vector-matrix equation

$$\underline{S}_\lambda = \underline{\tilde{h}}_\lambda + \underline{\underline{\tilde{h}}} \, \underline{\underline{\Delta\tilde{E}}}_\lambda^{-1} \underline{S}_\lambda,$$

which allows us, since it shows a Dyson-like character, to obtain $\underline{S}_\lambda$:

$$\underline{S}_\lambda = \left(\underline{\underline{1}} - \underline{\underline{\tilde{h}}} \, \underline{\underline{\Delta\tilde{E}}}_\lambda^{-1}\right)^{-1} \underline{\tilde{h}}_\lambda = \underline{\underline{\Delta\tilde{E}}}_\lambda \left(\underline{\underline{\Delta\tilde{E}}}_\lambda - \underline{\underline{\tilde{h}}}\right)^{-1} \underline{\tilde{h}}_\lambda. \tag{5.26}$$

Finally, we put equation (5.26) into equation (5.24) and get

$$\underline{A}_\lambda = \underline{1}_\lambda + \underline{\underline{\Delta\tilde{E}}}_\lambda^{-1} \left(\underline{\underline{1}} - \underline{\underline{\tilde{h}}} \, \underline{\underline{\Delta\tilde{E}}}_\lambda^{-1}\right)^{-1} \underline{\tilde{h}}_\lambda = \underline{1}_\lambda + \left(\underline{\underline{\Delta\tilde{E}}}_\lambda - \underline{\underline{\tilde{h}}}\right)^{-1} \underline{\tilde{h}}_\lambda, \tag{5.27}$$

or in component notation:

$$A_\lambda^\mu = \delta_{\mu\lambda} + \sum_{\mu'} \left(\underline{\underline{\Delta\tilde{E}}}_\lambda - \underline{\underline{\tilde{h}}}\right)_{\mu\mu'}^{-1} \tilde{h}_{\mu'\lambda} = \delta_{\mu\lambda} + \frac{(1 - \delta_{\mu\lambda})}{E_\lambda - E_\mu^0 - h_{\mu\mu}} \sum_{\mu'} \left(\underline{\underline{1}} - \underline{\underline{\tilde{h}}} \, \underline{\underline{\Delta\tilde{E}}}_\lambda^{-1}\right)_{\mu\mu'}^{-1} (1 - \delta_{\mu'\lambda})h_{\mu'\lambda}. \tag{5.28}$$

With equation (5.27), we have shown that it is possible to get an exact equation for the eigenvector, which does not depend on other components of the eigenvector but on elements that we can take from the Hamiltonian-matrix itself and the exact eigenvalue. Worth mentioning is that the reason why the eigenvector equation depends on the eigenvalue is because we started from the Gaussian algorithm, which requires the knowledge of the eigenvalues to obtain the eigenvectors.

Without going into more detail at this point, the structure of the eigenvector (Eq. (5.27), (5.28)) tells us that it will be an exciting object for analysis, because we can use it to analyze numerically, as

well as analytically (Ch. 5.3.1), how different degrees of excitonic effects will affect the eigenvectors. As well as developing new exact (Ch. 5.2.2, 5.6.2) and approximate approaches (Ch. 5.4, 5.6.1) for the calculation of eigenvalues and eigenvectors.

However, before we will turn to these chapters, we will take a look at which structures we can rewrite the eigenvectors into and what special features these structures entail. The results will be an important aid for later analyses and for a better understanding of the eigenvector structure itself.

Beta-matrix approach

The beta-matrix approach is another formal, exact way to obtain an equation for the eigenvectors. However, different than in the S-vector approach, in the beta-matrix approach, we only work with scalar quantities in the derivation of an expression for the eigenvectors.

The starting point for the derivation is the same as for the S-vector equation (Eq. (5.23)), but the turning point where the two paths differ is equation (5.25). In order to rewrite equation (5.25) into a suitible form, we use the property $S_\lambda^\mu \cdot (S_\lambda^\mu)^{-1} = 1$, which is only well-defined as a scalar property. The property allows us to find $S_\lambda^{\mu''}$ in the scalar form as:

$$S_\lambda^{\mu''} = \sum_\mu \tilde{h}_{\mu''\mu} A_\lambda^\mu = \tilde{h}_\lambda^{\mu''} + \sum_\mu \tilde{h}_{\mu''\mu} \left(\Delta\tilde{E}_{\mu\mu}(E_\lambda)\right)^{-1} S_\lambda^\mu \left(S_\lambda^{\mu''}\right)^{-1} S_\lambda^{\mu''} =: \tilde{h}_\lambda^{\mu''} + \beta_{\mu''\mu''}(E_\lambda) S_\lambda^{\mu''}.$$

As before, the structure of the determination equation shows a Dyson-like arrangement, which allows us to solve it for $S_\lambda^{\mu''}$, getting:

$$S_\lambda^{\mu''} = \frac{\tilde{h}_\lambda^{\mu''}}{1 - \beta_{\mu''\mu''}(E_\lambda)}, \quad \beta_{\mu''\mu''}(E_\lambda) = \sum_\mu \tilde{h}_{\mu''\mu} \left(\Delta\tilde{E}_{\mu\mu}(E_\lambda)\right)^{-1} S_\lambda^\mu \left(S_\lambda^{\mu''}\right)^{-1}. \tag{5.29}$$

Thus, we have a scalar equation for $S_\lambda^{\mu''}$, which can be inserted into equation (5.23) in order to get the scalar equation for the eigenvector components:

$$A_\lambda^\mu = \delta_{\mu\lambda} + \left(\Delta\tilde{E}_{\mu\mu}(E_\lambda)\right)^{-1} \frac{\tilde{h}_\lambda^\mu}{1 - \beta_{\mu\mu}(E_\lambda)}. \tag{5.30}$$

Even if this equation is already usable, we want to transform it into a vector-matrix equation. For this we define an effective diagonal-matrix equation for $\underline{\underline{\beta}}_\lambda$ as:

$$\underline{\underline{\beta}}_\lambda := \text{Diag}\left(\underline{\underline{S}}_\lambda^{-1}\underline{\underline{\tilde{h}}}\,\underline{\underline{\Delta\tilde{E}}}_\lambda^{-1}\underline{\underline{S}}_\lambda\right) = \text{diag}\left(\beta_\lambda^1, \beta_\lambda^2, \ldots, \beta_\lambda^N\right), \tag{5.31}$$

where we defined $\underline{\underline{S}}_\lambda^{-1} := \text{diag}\left(\frac{1}{S_\lambda^1}, \frac{1}{S_\lambda^2}, \ldots, \frac{1}{S_\lambda^N}\right)$ with $S_\lambda^\mu = \left[\left(\underline{\underline{1}} - \underline{\underline{\tilde{h}}}\,\underline{\underline{\Delta\tilde{E}}}_\lambda^{-1}\right)^{-1}\underline{\tilde{h}}_\lambda\right]^\mu$.

The definition of the diagonal beta-matrix (Eq. (5.31)) allows us to rewrite the scalar equation for the eigenvectors into an effective vector-matrix equation

$$\underline{A}_\lambda = \underline{1}_\lambda + \underline{\underline{\Delta\tilde{E}}}_\lambda^{-1} \frac{1}{\underline{\underline{1}} - \underline{\underline{\beta}}_\lambda} \underline{\tilde{h}}_\lambda. \tag{5.32}$$

The fascinating point about this equation is that although the Hamiltonian associated with the eigenvectors is not a diagonal matrix, the determination equation for the eigenvectors can be rewritten so that all elements within it represent either vectors or diagonal matrices. This in turn means that if the individual components of the eigenvector were known, a single eigenvector could

be calculated with a numerical scaling of $\sim O(N)$. Furthermore, this equation could prove to be particularly useful for numerical applications because, as we will see (Ch. 5.5), the most critical part of the whole equation is the matrix inversion $\left(\underline{\underline{1}} - \underline{\underline{\tilde{h}}}\,\underline{\underline{\Delta \tilde{E}}}_\lambda^{-1}\right)^{-1}$. However, this part appears twice in β_λ in $\underline{S}_\lambda$ and $\underline{S}_\lambda^{-1}$, which could result in a partial cancellation of incomplete information and thus actually turn weak approximations into useful approximations, as is often the case in many-body physics, for example in the choice of $W = W^{RPA}$ (Ch. 3.4.3), in the case of the nano-quantum kernel in time-dependent density functional theory [21] or in the case of the connector approach [117, 118].

After briefly discussing the advantages of the eigenvector in the beta-matrix representation , we will move on to the final structure of the eigenvector, which will be presented in this work. For the last structure, we want to proceed a little differently than before and ask ourselves the question: *Can we postulate a structure of the eigenvector and derive the missing elements from the eigenvalue equation?*

G-matrix approach

The last approach stands out from the other two because we do not obtain something by reformulating it but by postulating a structure. The idea is as follows: We have seen that the exact eigenvector comprises two components—firstly, a delta function and, secondly, a matrix-vector multiplication. Since we only have one equation, namely the eigenvalue equation (Eq. (5.21)), we can only leave one quantity unknown in our postulated structure, which we then determine from the eigenvalues equation. As the complicated structure is the matrix-inversion, we leave it open and assume a general approach for the eigenvector based on the other two previous ones (Eq. (5.27), (5.32))

$$\underline{A}_\lambda = \underline{1}_\lambda + \underline{\underline{G}}_\lambda \underline{\tilde{h}}_\lambda. \tag{5.33}$$

Based on the postulated structure the target is now to determine $\underline{\underline{G}}_\lambda$. For this reason, we use the eigenvalue equation in order to get $\underline{\underline{G}}_\lambda$, which means we operate with the matrix $(E_\lambda \underline{\underline{1}} - \underline{\underline{H}})$ onto the postulated eigenvector (Eq. (5.33)), obtaining:

$$(E_\lambda \underline{\underline{1}} - \underline{\underline{H}})\underline{A}_\lambda = (E_\lambda \underline{1}_\lambda - \underline{H}_\lambda) + (E_\lambda \underline{\underline{1}} - \underline{\underline{H}})\underline{\underline{G}}_\lambda \underline{\tilde{h}}_\lambda = 0 \tag{5.34}$$

Here $\underline{H}_\lambda$ is the vector which still contains the element $\mu = \lambda$ in opposite to $\underline{\tilde{h}}_\lambda$. Since we are looking for the expression for $\underline{\underline{G}}_\lambda$ or $\underline{\underline{G}}_\lambda \underline{\tilde{h}}_\lambda$, the next step is to move the first term from equation (5.34) to the right-hand side and then perform a matrix inversion, which leads us to the expression

$$\underline{\underline{G}}_\lambda \underline{\tilde{h}}_\lambda = \left(E_\lambda \underline{\underline{1}} - \underline{\underline{H}}\right)^{-1} \left(\underline{H}_\lambda - E_\lambda \underline{1}_\lambda\right) = \left(E_\lambda \underline{\underline{1}} - \underline{\underline{H}}\right)^{-1} \underline{\tilde{H}}_\lambda, \tag{5.35}$$

with $\underline{\tilde{H}}_\lambda := (\underline{H}_\lambda - E_\lambda \underline{1}_\lambda)$. Before we insert equation (5.35) into the equation for the eigenvector (Eq. (5.33)), we want to be aware of one characteristic that the obtained equation (5.35) has, which becomes obvious when we represent the matrix inversion in spectral representation:

$$(E_\lambda \underline{\underline{1}} - \underline{\underline{H}})^{-1} = \sum_{\lambda'} \frac{\bar{A}_{\lambda'} \bar{A}_{\lambda'}^\dagger}{E_\lambda - E_{\lambda'}}. \tag{5.36}$$

We observe that the obtained equation has a divergence, namely $E_\lambda = E_{\lambda'}$, which, of course, presents us with a numerical problem. In order to avoid such a problem, we introduce a broadening

parameter into the equation:

$$\lim_{\gamma \to 0} ((E_\lambda + i\gamma)\underline{\underline{1}} - \underline{\underline{H}})^{-1} = \lim_{\gamma \to 0} \sum_{\lambda'} \frac{\bar{A}_{\lambda'} \bar{A}_{\lambda'}^\dagger}{(E_\lambda + i\gamma) - E_{\lambda'}}, \tag{5.37}$$

as it is done in the case of the correlation function (Ch. 3.3.1). Since this is a non-trivial procedure, it is worth briefly explaining in more detail why such a procedure is necessary. To do so, we first insert equation (5.37) into equation (5.35), obtaining:

$$\underline{\underline{G}}_\lambda \tilde{\underline{h}}_\lambda = \lim_{\gamma \to 0} ((E_\lambda + i\gamma)\underline{\underline{1}} - \underline{\underline{H}})^{-1} \tilde{\underline{H}}_\lambda \tag{5.38}$$

Next, we want to assume the case in which it is immediately apparent that we will have a numerical problem, namely the case in which our Hamiltonian is a diagonal matrix with the entries $H_{\mu\mu}$. In this case, the eigenvalues are trivially given by $E_\lambda = H_{\lambda\lambda}$. If we now write equation (5.38) in components for the diagonal Hamiltonian case, we obtain

$$\lim_{\gamma \to 0} \frac{1}{(H_{\lambda\lambda} + i\gamma) - H_{\mu\mu}} (H_{\mu\lambda} - H_{\lambda\lambda}).$$

As long as $\mu \neq \lambda$ holds, there is no problem and the limit $\gamma \to 0$ can be applied. If the case $\mu = \lambda$ occures, we get a numerical problem, because the matrix inversion would diverge $\frac{1}{H_{\lambda\lambda} - H_{\lambda\lambda}}$ while the vector-component $H_{\lambda\lambda} - H_{\lambda\lambda}$ disappears. Whereas, if we insert the γ parameter, we can first set $\mu = \lambda$ and then $\gamma \to 0$, which means that the element $\mu = \lambda$ disappears

$$\lim_{\gamma \to 0} \frac{1}{(H_{\lambda\lambda} + i\gamma) - H_{\lambda\lambda}} (H_{\lambda\lambda} - H_{\lambda\lambda}) \to 0,$$

which is exactly what we observe in the other cases of eigenvectors (Eq. (5.27), (5.32)). This illustrative example should make it clear to us why we need the factor γ for numerical applications. If we finally insert equation (5.38) into the postulated starting equation of the eigenvector (Eq. (5.33)), we obtain our third and final equation for the eigenvector:

$$\underline{A}_\lambda = \underline{1}_\lambda + \lim_{\gamma \to 0} \left((E_\lambda + i\gamma)\underline{\underline{1}} - \underline{\underline{H}} \right)^{-1} \tilde{\underline{H}}_\lambda = \underline{1}_\lambda + \lim_{\gamma \to 0} \underline{\underline{\Delta E}}_\lambda^{-1}(\gamma) \left(\underline{\underline{1}} - \tilde{\underline{h}} \, \underline{\underline{\Delta E}}_\lambda^{-1}(\gamma) \right)^{-1} \tilde{\underline{H}}_\lambda, \tag{5.39}$$

with the definition $\underline{\underline{\Delta E}}_\lambda^{-1}(\gamma) := \left((E_\lambda + i\gamma)\underline{\underline{1}} - \mathrm{diag}\left(\underline{\underline{H}} \right) \right)^{-1}$.

Having reached this point, at which we have three exact equations for the eigenvector, it is time to compare them with each other on an analytically basis. For the analytical comparison, we will write down the three equations one below the other:

S-vector:

$$\underline{A}_\lambda = \underline{1}_\lambda + \underline{\underline{\Delta \tilde{E}}}_\lambda^{-1} \left(\underline{\underline{1}} - \tilde{\underline{h}} \, \underline{\underline{\Delta \tilde{E}}}_\lambda^{-1} \right)^{-1} \tilde{\underline{h}}_\lambda,$$

beta-matrix:

$$\underline{A}_\lambda = \underline{1}_\lambda + \underline{\underline{\Delta \tilde{E}}}_\lambda^{-1} \frac{1}{\underline{\underline{1}} - \underline{\underline{\beta}}_\lambda} \tilde{\underline{h}}_\lambda,$$

with $\underline{\underline{\beta}}_\lambda := \mathrm{Diag} \left(\underline{\underline{S}}_\lambda^{-1} \tilde{\underline{h}} \, \underline{\underline{\Delta \tilde{E}}}_\lambda^{-1} \underline{\underline{S}}_\lambda \right)$ and $S_\lambda^\mu = \left[\left(\underline{\underline{1}} - \tilde{\underline{h}} \, \underline{\underline{\Delta \tilde{E}}}_\lambda^{-1} \right)^{-1} \tilde{\underline{h}}_\lambda \right]^\mu$.

G-matrix:

$$\underline{A}_\lambda = \underline{1}_\lambda + \lim_{\gamma \to 0} \underline{\underline{\Delta E}}_\lambda^{-1}(\gamma) \left(\underline{\underline{1}} - \tilde{\underline{h}} \, \underline{\underline{\Delta E}}_\lambda^{-1}(\gamma) \right)^{-1} \tilde{\underline{H}}_\lambda.$$

The first observation we want to note is that the intrinsic structure of all three equations is almost identical. All equations have an independent part $\underline{1}_\lambda$, which is only one for the element $\mu = \lambda$

and zero for the rest. Furthermore, all eigenvectors have a matrix in their second part ($\underline{\underline{\Delta \tilde{E}}}_\lambda^{-1}$, $\underline{\underline{\Delta E}}_\lambda^{-1}(\gamma)$) which represents an inverse energy difference between the exact eigenvalue E_λ and the diagonal elements of the Hamiltonian $H_{\mu\mu}$. Since, as we discussed in the case of the G-matrix approach, this component can in principle exhibit divergent behavior and will give therefore a decisive contribution to the structure of the eigenvector, as we will see in section 5.3 and already see in figure 5.8. The crucial difference between the structures is that in the case of the S-vector and the G-matrix approach, the inversion $\left(\underline{\underline{1}} - \underline{\underline{\tilde{h}}}\,\underline{\underline{\Delta \tilde{E}}}_\lambda^{-1}\right)^{-1}$ resp. $\left(\underline{\underline{1}} - \underline{\underline{\tilde{h}}}\,\underline{\underline{\Delta E}}_\lambda^{-1}(\gamma)\right)^{-1}$ is executed directly in the structure of the eigenvector and therefore leads to a multiplication between a non-diagonal matrix and a vector $\underline{\tilde{h}}_\lambda$ resp. $\underline{\tilde{H}}_\lambda$, but in the case of the beta-matrix approach it is a diagonal matrix, which carries the information of the inversion in itself, but is by definition a diagonal matrix. For this reason, we have the same multiplication between a matrix and a vector as in the other two cases, but the matrix is diagonal. This is an impressive result, because the diagonal matrix must contain all the information contained in the non-diagonal matrix and pass it on to the vector $\underline{\tilde{h}}_\lambda$ so that the $\left(\underline{\underline{1}} - \underline{\underline{\tilde{h}}}\,\underline{\underline{\Delta \tilde{E}}}_\lambda^{-1}\right)^{-1}\underline{\tilde{h}}_\lambda = \frac{1}{\underline{\underline{1}}-\underline{\underline{\beta}}_\lambda}\underline{\tilde{h}}_\lambda$ applies, which is very exciting from a mathematical point of view. Another important difference is that, the G-matrix approach contains a parameter γ, which must approach zero in the exact limit, but as we will see later (Ch. 5.6.1) will play an important role for approximations.

Now that we have dealt with the similarities and differences, we want to make sure that all three approaches actually give the same result for the eigenvectors before we move on to the eigenvalues.
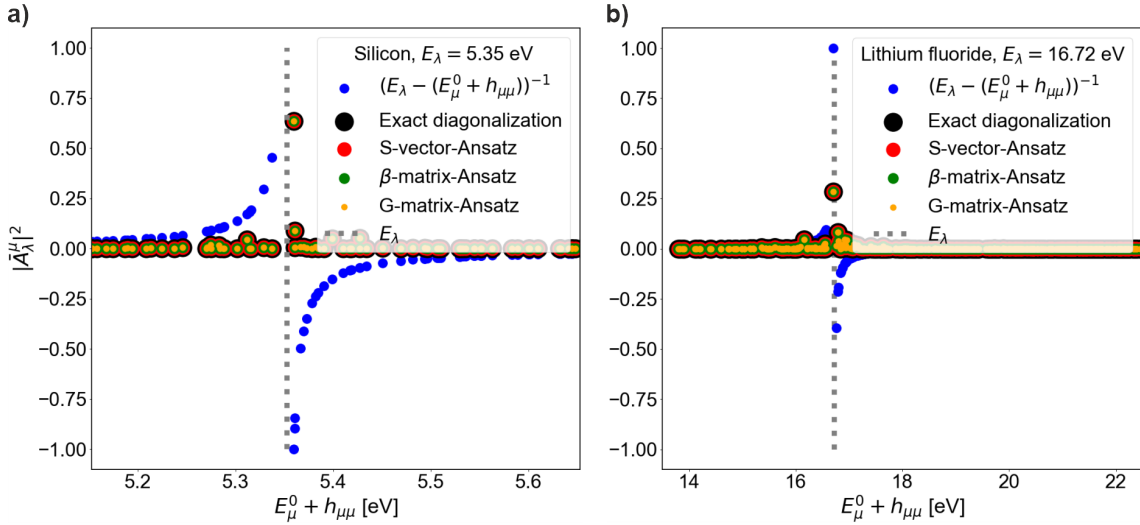

Figure 5.8: Exact normalized eigenvector $\underline{\bar{A}}_\lambda$, belonging to the eigenvalue E_λ, in the S-vector (Eq. (5.27)) (red), beta-matrix (Eq. (5.32)) (green) and G-matrix approach (Eq. (5.39), $\gamma = 10^{-6}$ Ha) (orange) compared to the exact diagonalization eigenvector (black). Plotting the norm of the normalized eigenvector-components, $|\bar{A}_\lambda^\mu|^2$, against the diagonal elements $E_\mu^0 + h_{\mu\mu}$ of the Hamiltonian $\underline{\underline{H}}^{2p,reso}$ (Eq. (4.13)) for **a)** silicon and **b)** lithium fluoride. All three approaches reproduce the exact result. The eigenvectors are formed around the divergent part of $(E_\lambda - (E_\mu^0 + h_{\mu\mu}))^{-1}$ (blue). Numerical values given in the Appendix A calculations were performed on an 3k grid (108 k-points) using a Kohn-Sham band structure, with 6 transition bands for the optical calculation.

As expected we observe, that all approaches yield the same result as the exact diagonalized one
(Fig. 14, black) on the visible scale. Additionally, we see that the prominent peak of the eigenvector
is around $E_\mu^0 + h_{\mu\mu} \approx E_\lambda$, and most eigenvector entries are located close to zero. In order to better
understand this behavior, we need a detailed analysis of the Hamiltonian and the eigenvector
structure, which we will postpone to section 5.3.

Summarizing this section, we have shown three approaches for the eigenvector. Each of them
comes with its special features. One comes directly from a diagonal matrix, and one includes
broadening. However, all possess the same problem: They depend on the exact eigenvalue. So,
we have to know the eigenvalue to determine the eigenvector. At the moment, it seems like the
chicken-and-egg problem, but that is only apparently the case. For this reason, we will turn our
attention to developing a method that allows us to obtain the eigenvalues from the structure of
the eigenvectors.

5.2.2   The exact eigenvalue approach

The succeeding section will deal with developing an equation for the eigenvalues beyond the eigen-
value equation. The target on a long term will be to get the eigenvalues numerically cheaply, both
for their own sake and for calculating the eigenvectors. Furthermore, developing an equation for
the eigenvalues will allow us to understand better the differences between the individual exciton
types (continuum and bound, Ch. 4.6). Since we have exact equations for the eigenvectors, which
allow us to set up an exact equation for the eigenvalues, we can be in good spirits to make exciting
discoveries.

For the task of formulating an equation for the eigenvalues, we will simplify the notation a little
and write the eigenvector equation (Eq. (5.27), (5.32), (5.39)) in braket notation

$$|\bar{A}_\lambda\rangle = \frac{|1_\lambda\rangle + |L_\lambda\rangle}{N_\lambda}, \ N_\lambda = \sqrt{\langle A_\lambda | A_\lambda \rangle}.$$

Since our goal is to calculate eigenvalues, we use the eigenvalue equation:

$$E_\lambda = \langle \bar{A}_\lambda | \hat{H} | \bar{A}_\lambda \rangle = \frac{1}{N_\lambda^2} \langle A_\lambda | \hat{H} | A_\lambda \rangle,$$

in order to obtain

$$N_\lambda^2 E_\lambda = \langle 1_\lambda | \hat{H} | 1_\lambda \rangle + \langle L_\lambda | \hat{H} | L_\lambda \rangle + 2\Re(\langle L_\lambda | \hat{H} | 1_\lambda \rangle),$$

after substituting the expression for the eigenvectors into the eigenvalue equation. Using $\langle 1_\lambda | \hat{H} | 1_\lambda \rangle = E_\lambda^0 + h_{\lambda\lambda}$, together with

$$N_\lambda^2 = \langle A_\lambda | A_\lambda \rangle = \langle 1_\lambda | 1_\lambda \rangle + 2\Re(\langle L_\lambda | 1_\lambda \rangle) + \langle L_\lambda | L_\lambda \rangle,$$

we get the equation for the eigenvalues depending on the components of the eigenvector

$$E_\lambda = \frac{1}{\langle 1_\lambda | 1_\lambda \rangle + 2\Re(\langle L_\lambda | 1_\lambda \rangle) + \langle L_\lambda | L_\lambda \rangle} \big[E_\lambda^0 + h_{\lambda\lambda} + \langle L_\lambda | \hat{H} | L_\lambda \rangle + 2\Re(\langle L_\lambda | \hat{H} | 1_\lambda \rangle) \big].$$

To convert this equation into a structure as simple as possible, we define:

$$a_\lambda := 2\Re(\langle L_\lambda | 1_\lambda \rangle), \ b_\lambda := \langle L_\lambda | \hat{H} | L_\lambda \rangle, \ \bar{b}_\lambda := 2\Re(\langle L_\lambda | \hat{H} | 1_\lambda \rangle), \ c_\lambda := \langle L_\lambda | L_\lambda \rangle,$$

$$d_\lambda := \frac{b_\lambda/c_\lambda}{E_\lambda^0 + h_{\lambda\lambda}}, \, \bar{d}_\lambda := \frac{\bar{b}_\lambda/a_\lambda}{E_\lambda^0 + h_{\lambda\lambda}},$$

in order to get the final expression for the eigenvalues

$$E_\lambda = \frac{1 + \bar{d}_\lambda a_\lambda + d_\lambda c_\lambda}{1 + a_\lambda + c_\lambda} \left[E_\lambda^0 + h_{\lambda\lambda} \right] =: e_\lambda \left[E_\lambda^0 + h_{\lambda\lambda} \right]. \tag{5.40}$$

The obtained equation for the eigenvalues displays the eigenvalues in terms of two factors, namely the diagonal elements of the Hamiltonian $E_\lambda^0 + h_{\lambda\lambda}$ and a prefactor e_λ which changes for every eigenvalue and will be called in the following the matrix structure factor, as it transfers the information beyond the diagonal elements from the hamiltonian into the eigenvalues. Even if equation (5.40) may look very simple at first, we will see that it will be invaluable for analytical purposes, as it will give us an insight into how the eigenvalues are obtained and thus help us to derive and understand important approximations (Ch. 5.3.1, 5.4).

Before we take a look at the matrix structure factor in practice, i.e. what values it assumes in materials such as Si and LiF, we will take a look at a few properties of the matrix structure factor related to our hermitian matrix given by the resonant part of the two-particle Hamiltonian. Since the Hamiltonian is a hermitian matrix, the eigenvalues are real valued $E_\lambda \in \mathbb{R}$ and positive $\forall_\lambda E_\lambda > 0$. Furthermore, since in the thermodynamic limit ($N_k \to \infty$) the following applies: $\lim_{N_k \to \infty} \left[E_\lambda^0 + h_{\lambda\lambda} \right] = \lim_{N_k \to \infty} \left[E_\lambda^0 + \frac{1}{N_k} \bar{h}_{\lambda\lambda} \right] \to E_\lambda^0$, it holds that $|E_\lambda^0| \gg |h_{\lambda\lambda}|$. Together with the fact that for the independent-particle transition energies $E_\lambda^0 \in \mathbb{R}_{>0}$ holds, it follows that the matrix structure factor is also real valued and positive $e_\lambda \in \mathbb{R}_{>0}$.

Now that we have found a lower bound for the matrix structure factor, the question arises whether there is also an upper bound. Mathematically, it may not be easy to derive the upper bound in general for our two-particle Hamiltonian, but we can derive it argumentatively with physical intuition. When we speak of excitons, we generally mean bound electron-hole pairs, which are created by an attractive interaction between electrons and holes. Thus it follows from intuition that the energies E_λ of the excitons can be at most less than or equal to the energies of the unbound electron-hole system E_λ^0. In order to understand what this means for the matrix structure factor we now assume that we are in the thermodynamic limit, so that $|E_\lambda^0| \gg |h_{\lambda\lambda}|$ holds and thus $E_\lambda \approx e_\lambda E_\lambda^0$. If we now take into account that $E_\lambda \leqslant E_\lambda^0$, the matrix structure factor can have a maximum value of one ($e_\lambda \leqslant 1$). It should be pointed out once again that this conclusion is based on physical intuition and is therefore not a mathematical proof. Nevertheless, this conclusion gives us a direction of what to expect for the matrix structure factor, namly that the value will vary between zero and one.

Before we go over to the value that the matrix structure factor will assume in Si and LiF, we will briefly clarify the following two points:

1. Non-interacting system: If the Hamiltonian $\underline{\underline{H}}$ is a diagonal matrix, then $e_\lambda = 1$ immediately follows.

2. Interacting system: If the Hamiltonian $\underline{\underline{H}}$ is less diagonal, then $e_\lambda \neq 1$ is possible, but not mandatory.

Point 1. is trivial, because without interaction the eigenvalues immediately change to those of the independent-particle transition energies and thus $e_\lambda = 1$ must apply.

Point 2. is less trivial, because of course we expect that if more information is available in the Hamiltonian, it will also pass into the eigenvalues in some form and thus $e_\lambda \neq 1$ will apply, but this depends strongly on the structure of the Hamiltonian and thus on its eigenvectors. It is quite possible that only certain regions of the Hamiltonian contribute significantly to certain eigenvalues, so that a value very different from one is necessary for the matrix structure factor, but other eigenvalues hardly differ from their diagonal elements, since the information that contributes to these eigenvalues comes rather little from the off-diagonal elements. Thus, even for Hamiltonians with non-negligible off-diagonal elements, regions can occur in which the matrix structure factor is close to one. Again, however, this is a more argumentative explanation of why we might get what we get. For a deeper analysis, we refer to the following sections 5.3.1 and 5.4.
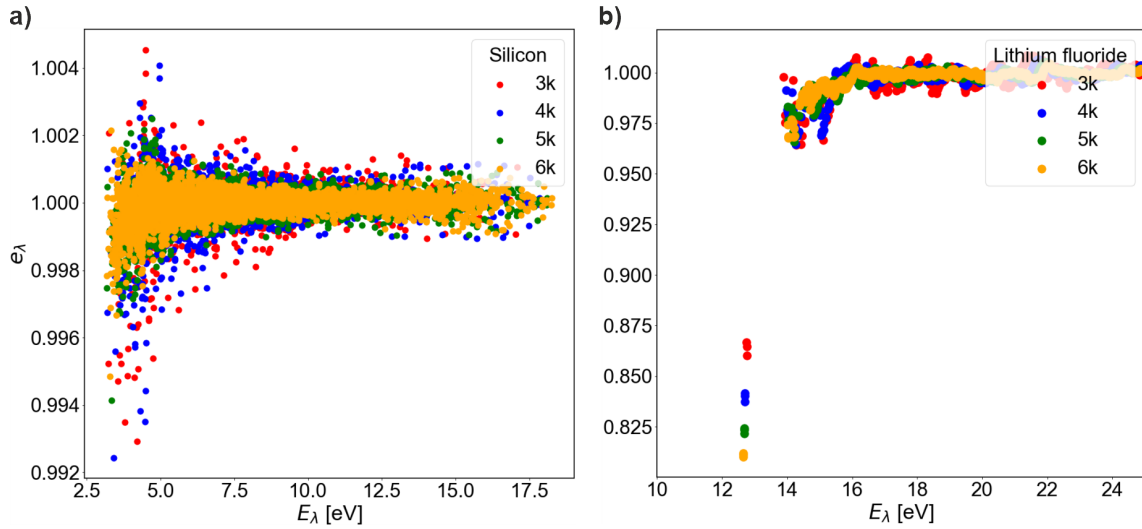
Figure 5.9: Thermodynamic convergence of the matrix structure factor e_λ (Eq. (5.40)) for the Hamiltonian $\underline{\underline{H}}^{2p,reso}$ (Eq. (4.13)) of **a)** silicon and **b)** lithium fluoride. Results shown e_λ plotted against the eigenvalue E_λ for increasingly dense k-grids of 3k (red), 4k (blue), 5k (green) and 6k (orange). It is observable that for silicon (a)) the matrix structure factor is nearly one everywhere, whereas for (b)) lithium fluoride it is one in the continuum but different from one for the first three eigenvalues inside the optical bandgap. Numerical values are given in Appendix A; calculations were performed with 6 transition bands for the optical calculation; using a Kohn-Sham band structure.

Looking at figure 5.9, for Si (Fig. 5.9 a)), we observe three properties: firstly, with increasing energy (E_λ), e_λ converges faster towards 1. Secondly, the denser the k-grid, the faster the values converge towards 1 (thermodynamic limit), and thirdly, all values seem to be converging towards 1. If we look at LiF (Fig. 5.9 b)), we see that all three points apply to the continuum (Si: $E_\lambda \geqslant 3.27$ eV, LiF: $E_\lambda \geqslant 14.92$ eV), but they do not to the eigenvalues responsible for the strong bound excitons. This can be seen particularly for the first three eigenvalues at around 12.7 eV.
However, regardless of the strongly bound excitons, it is astonishing that it appears that for all other excitonic energies, both in Si and in LiF, $e_\lambda \to 1$ holds. This is not something one would trivially expect, especially not for LiF, a material with, as we have seen, very strong excitonic effects. One would perhaps expect such behavior for materials with very weak excitonic effects

($E_\lambda \approx E_\lambda^0$), but certainly not for one with strong excitonic effects.

For the moment, we want to consider the observed difference between Si and LiF and especially between the continuum and the strong bound excitons as an observation. We will postpone the discussion about the behavior of the matrix structure factor e_λ to the next section because we still need some information for a fully comprehensive discussion.

However, the special nature of this discovery should be emphasized once again. If this behavior turns out not to be a coincidence, but an explainable mathematical or physical pattern, then this would mean, conversely, that except for the eigenenergies of strongly bound excitons, all eigenenergies are known and given by the equation $E_\lambda = E_\lambda^0 + h_{\lambda\lambda}$, without any additional computational effort other than setting up the Hamiltonian.

Inspired by this discovery, we now want to examine whether the observed behavior is coincidence or whether there is a certain pattern behind it. To find out, the equations for eigenvectors and eigenvalues developed in this section will be invaluable, as they will prevent us from having to analyze the Hamiltonian itself and allow us to focus more on the physics.

5.3 Eigenvector structure analysis: From weak to strong excitonic effects

Our focus in the following will be to enhance the comprehension of how the strength of excitonic effects and the formation of the eigenvectors interact. Understanding this interplay is of fundamental importance for developing approximations for both the eigenenergies and the eigenvectors, as it will help us recognize and understand patterns and relationships as we have seen them in the last section. Even if we can learn a lot from analyzing materials such as Si and LiF, it is not easy to understand why this or that effect forms the way it does, as much information contributes. For this reason, we will concentrate on developing an understanding based on two models, which allow us to scale the excitonic effects via a single parameter. Furthermore. one of the models, which we will call the 2×2 model, will allow us to analyze the interplay between the strength of the excitonic effects and the formation of the eigenvectors in an analytical way. In the further course we will subsequently apply the acquired knowledge to interpret observations in real materials, specifically examining the tendency of the matrix structure factor in the continuum to converge towards the value one.

Before we move on to the models, we will briefly recall the structure of the eigenvectors. For the following analyses, it will be advantageous to use the S-vector derivation (Eq. (5.27)), as the eigenvector structure is derived directly from the Gaussian algorithm and will thus facilitate analytical analyses, in a much more insightful way than the beta- or G-matrix approach could do. In component representation, the eigenvector in the S-vector approach can be represented as:

$$A_\lambda^\mu = \delta_{\mu\lambda} + \frac{(1 - \delta_{\mu\lambda})}{E_\lambda - (E_\mu^0 + h_{\mu\mu})} \sum_{\mu'} \left(\underline{\underline{1}} - \underline{\underline{\tilde{h}}} \, \underline{\underline{\Delta \tilde{E}}}_\lambda^{-1} \right)_{\mu\mu'}^{-1} \tilde{h}_{\mu'\lambda}. \tag{5.41}$$

5.3.1   The c-models

Since the aim of this section is to gain a better understanding of the connection between excitonic effects and the structure of the eigenvectors and eigenvalues, we want to concentrate in the following on two models, one of which will be given by a general 2×2 Hamiltonian and thus will give us an analytical insight and one which is more realistic and will allow us to visualize what to expect in real materials. Both models will be based on the following equation

$$\underline{\underline{H}}^{2p,reso} = \underline{\underline{H}}^0 + c \cdot \underline{\underline{\Xi}}^{AGW} = \underline{\underline{H}}^0 + c \cdot \underline{\underline{h}}, \tag{5.42}$$

in which the parameter c will allow us to scale the excitonic effects. In the case of the simple 2×2 model, we will use the Hamiltonian (Eq. (5.42)) in the form

$$\underline{\underline{H}}^{2p,reso} = \underline{\underline{H}}^0 + c \cdot \underline{\underline{h}} = \begin{pmatrix} E_1^0 & 0 \\ 0 & E_2^0 \end{pmatrix} + c \cdot \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix}. \tag{5.43}$$

For the more realistic model, which we want to use for visualization purposes and to prove the statements from the analytical model, we want to use the resonant part of the two-particle Hamiltonian for Si and scale the interaction part with a constant c. Thus, with this more realistic model, we can simulate weaker excitonic effects than for Si ($c < 1$) and stronger ones than for Si ($c > 1$). However, before we move on to the visualization, we want to build up an analytical knowledge of what to expect and why, so we start with the 2×2 model. In order not to complicate the analysis unnecessarily, we will limit ourselves to the eigenvector corresponding to the eigenenergy $E_{\lambda=1}$, which is given by the equation

$$\underline{A}_{\lambda=1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{c \cdot h_{21}}{E_{\lambda=1} - (E_2^0 + c \cdot h_{22})} \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{5.44}$$

With equation (5.44), we can perform various analyses for different excitonic strengths c. For the analysis, we will turn on the strength of the excitonic effects c step by step, from negligible ($c = 0$) to weak ($c \approx 0$), over to strong excitonic effects ($c \gg 0$).

The first case is the simplest, namely, where we neglect the excitonic effects, $c = 0$. In this scenario, as one would expect, the eigenvector is given by:

$$\underline{A}_{\lambda=1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{5.45}$$

Therefore, if we plot the eigenvector components, it is clear that we observe one pronounced peak. If we go one step further and assume that the excitonic effects are weak but do not disappear, $c \approx 0$, then we can set up a Taylor series for the c-dependent part of the eigenvector, which gives us in the first-order, the following equation:

$$\underline{A}_{\lambda=1} \approx \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \partial_c \left[\frac{c \cdot h_{21}}{E_{\lambda=1} - (E_2^0 + c \cdot h_{22})} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right]_{c=0} \cdot c =$$
$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{c \cdot h_{21}}{(E_{\lambda=1})_{c=0} - E_2^0} \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{5.46}$$

The obtained structure for the eigenvector is already interesting since the denominator $(E_{\lambda=1})_{c=0} - E_2^0$ can diverge. In order to make the behavior more visible, we also consider for the eigenvalues

that $c \approx 0$, then according to equation (5.44), the eigenvalues are given in leading order by the independent-particle transition energies $E_\lambda \approx E_\lambda^0$. Under this assumption, the eigenvector is given by:

$$\underline{A}_{\lambda=1} \approx \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{c \cdot h_{21}}{(E_{\lambda=1})_{c=0} - E_2^0} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \approx \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{c \cdot h_{21}}{E_1^0 - E_2^0} \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{5.47}$$

To clarify why the denominator can diverge, we assume that our simple 2×2 model represents a section of a realistic Hamiltonian for Si, for example. Furthermore, we assume that the two energies E_1^0 and E_2^0 have the same band indexes but differ in the k-point $E_1^0 = E_{(v_1, c_1, k_1)}^0$, $E_2^0 = E_{(v_1, c_1, k_2)}^0$. In that case, we realize that if we proceed to the thermodynamic limit $(N_k \to \infty)$, i.e. the k-points are very close together, that for the independent-particle transition energies $E_2^0 = \lim_{\nu \to 0} \left[E_1^0 + \nu \right]$ applies. However, this would mean according to our equation for the eigenvector (5.47), that we will observe that depending on the size of h_{21}, which scales with the number of k-points $h_{21} \sim \frac{1}{N_k}$ the second part of our eigenvector will diverge, since $E_1^0 \approx E_2^0$ holds. So, if we normalize the vector

$$\underline{A}_{\lambda=1} \approx \frac{1}{\left(1 + \left(\frac{c \cdot h_{21}}{E_1^0 - E_2^0} \right)^2 \right)^{1/2}} \left[\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{c \cdot h_{21}}{E_1^0 - E_2^0} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right], \tag{5.48}$$

and assume that applies $1 \ll \left(\frac{c \cdot h_{21}}{E_1^0 - E_2^0} \right)^2 \approx \lim_{\nu \to 0} \left(\frac{c \cdot h_{21}}{E_1^0 - E_1^0 + \nu} \right)^2$, than we obtain that the eigenvector is approximated by a single entry

$$\underline{A}_{\lambda=1} \approx \frac{E_1^0 - E_2^0}{c \cdot h_{21}} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \approx \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{5.49}$$

which lies in the entry of the eigenvector that diverges the most and therefore not in the same entry as in the non-interacting case $c = 0$ (Eq. (5.45)). The result shows us that we should expect to observe eigenvectors with one pronounced peak in the non-interactive case and materials with weak excitonic effects.

After we analyzed the weak excitonic interactions, we want to venture into the regime of strong excitonic effects $(c \gg 0)$, particularly the strongly bound excitons. In this case, the eigenvalue E_λ will deviate strongly from the diagonal elements, as we saw in the last chapter (Fig. 5.9), so we can simplify the eigenvector equation by replacing $E_{\lambda=1} - (E_2^0 + c \cdot h_{22})$ with a c dependent constant $|b(c)| \gg 0$, which leads to the following form of the eigenvector:

$$\underline{A}_{\lambda=1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{c \cdot h_{21}}{b(c)} \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{5.50}$$

Since by definition in the case of strongly bound excitons $(E_\lambda < \min(E_\mu^0))$, both $|b| \gg 0$ and $|c| \gg 0$ apply, there is no reason for us to assume that in this case, we have to observe a form of divergence. Thus, depending on the magnitude of c and b, the vector takes a form different from the pronounced peak structure, e.g. a broadened structure.

Summarizing the three results, we expect that from negligible excitonic effects to strong excitonic effects, the eigenvector will change from a structure with a pronounced peak into a broadened structure.

However, the provided analysis cannot explain why, in the last chapter (Fig. 5.9), we observed that the matrix structure factor in the continuum converges to one in both Si and LiF. Because so far

this should not be the case according to our analysis, since in turn it means that the eigenvalues are given by the diagonal elements of the Hamiltonian $E_\lambda = e_\lambda \left[E_\lambda^0 + h_{\lambda\lambda}\right] \to E_\lambda^0 + h_{\lambda\lambda}$ and thus the corresponding eigenvectors should have a structure with one pronounced peak also for materials with strong excitonic effects.

Due to this observation, we can claim that there must be a similarity between the continuum of different materials, which explains why for the eigenenergies E_λ the strength of the excitonic effects plays a minor role in the continuum than within the optical bandgap and thus causes the matrix structure factor to approach one ($e_\lambda \to 1$).

To understand the reason behind this behavior, we want to proceed as follows: The target is to understand why e_λ tends to be one in the continuum but not inside the optical bandgap. For this, we want to rewrite the equation for the matrix structure factor (Eq. (5.40)) for the 2×2 model into:

$$e_{\lambda=1} = \frac{E_{\lambda=1}}{E_1^0 + h_{11}}. \tag{5.51}$$

In order to make our life easier, we will assume that the excitonic interactions are identical for all bands and k-points, and therefore, $h_{\mu\mu'} = \tilde{c}$ applies with $|\tilde{c}| < \min(E_\mu^0)$. The assumption is drastic and certainly not justified in real materials. However, if we neglect for a moment the electron-hole exchange part in the interaction part and consider only the screened Coulomb part of the interaction (Eq. (4.14)), then we can figure out under which circumstances such an assumption of a constant interaction might be justified. In order not to complicate matters, we assume that only one valence band and one conduction band are sufficient. In this case, the following results for the interaction part:

$$h_{(vc\boldsymbol{k})(vc\boldsymbol{k}')} = \Xi^{AGW}_{(vc\boldsymbol{k})(vc\boldsymbol{k}')} \approx -\int d\boldsymbol{r}d\boldsymbol{r}'\, \Psi_{c,\boldsymbol{k}}^*(\boldsymbol{r})\Psi_{c,\boldsymbol{k}'}(\boldsymbol{r})W(\boldsymbol{r},\boldsymbol{r}')\Psi_{v,\boldsymbol{k}}(\boldsymbol{r}')\Psi_{v,\boldsymbol{k}'}^*(\boldsymbol{r}'). \tag{5.52}$$

Since the equation is still not easy to analyze in this form, we want to take a tight-binding approach for the Bloch functions, i.e. assume that the electrons are strongly localized:

$$\Psi_{jk}(r) = \frac{1}{\sqrt{N}}\sum_{\boldsymbol{R}} e^{i\boldsymbol{k}\boldsymbol{R}}\phi_j(\boldsymbol{r} - \boldsymbol{R}),$$

where $\boldsymbol{R}$ is a lattice vector, ϕ_j is a normalized atomic orbital function belonging to a band index j and centered on the atom at the point $\boldsymbol{R}$ and N is the number of atoms. It can be shown [119] that the screened Coulomb part (Eq. (5.52)) can be approximated by assuming that the atomic orbitals are strongly localized to

$$\Xi^{AGW}_{(vc\boldsymbol{k})(vc\boldsymbol{k}')} \approx -\frac{1}{N^2}\sum_{\boldsymbol{R}_1,\boldsymbol{R}_2} e^{i(\boldsymbol{k}-\boldsymbol{k}')(\boldsymbol{R}_1-\boldsymbol{R}_2)}W(\boldsymbol{R}_1 - \boldsymbol{R}_2). \tag{5.53}$$

Thus, the $\boldsymbol{k}, \boldsymbol{k}'$ dependency disappears exactly when $\boldsymbol{R}_1 = \boldsymbol{R}_2$ applies and thus the screened Coulomb interaction W is short-range. After this brief analysis, we see that our constant interaction model may be passably suitable for materials that exhibit a strong short-range interaction.

However, in order now to be able to analyze equation (5.51) for e_λ, we calculate the eigenvalues

$$E_\lambda = \frac{E_1^0 + E_2^0 + 2\tilde{c} \pm \sqrt{(E_1^0 - E_2^0)^2 + (2\tilde{c})^2}}{2} \tag{5.54}$$

for the 2×2 model with a constant interaction part. For the further course, we assume, as above, that we are on a very dense k-grid and that the relevant information for the eigenvalue stems

from regions in the Hamiltonian with the same band index but different k-points. Under this assumption, the independent-particle transition energies are very close together, so that $E_1^0 \approx E_2^0$ holds and therefore, the equation for the eigenvalues (Eq. (5.54)) becomes:

$$E_\lambda \approx E_1^0 + \tilde{c} \pm (\tilde{c}). \tag{5.55}$$

If we insert the obtained equation for the eigenvalues into equation (Eq. (5.51)) and assuming $h_{11} = \tilde{c}$, we get:

$$e_{\lambda=1} = \frac{E_1^0 + \tilde{c} \pm \tilde{c}}{E_1^0 + \tilde{c}} = \frac{E_1^0 + \tilde{c}}{E_1^0 + \tilde{c}} \pm \frac{\tilde{c}}{E_1^0 + \tilde{c}} = 1 \pm \frac{\tilde{c}}{E_1^0 + \tilde{c}}. \tag{5.56}$$

At this point it is almost irrelevant which case ($\pm$) we are looking at, since both cases will hold the same result if we go deep into the continuum: The matrix structure factor e_λ converges towards one, with increasing E_1^0

$$e_{\lambda=1} = \lim_{E_1^0 \to \infty} 1 \pm \frac{\tilde{c}}{E_1^0 + \tilde{c}} \to 1. \tag{5.57}$$

However, the smaller the energy E_1^0 becomes, the more the interaction $\tilde{c}$ will contribute and the more the matrix structure factor will deviate from one ($e_\lambda \neq 1$). Now that we have understood the behavior of the matrix structure factor, we want to understand how the eigenvectors behave when we move from the optical bandgap ($E_\lambda < \min(M^{IP}(E_\lambda))$) to the deep continuum ($E_\lambda \gg \min(M^{IP}(E_\lambda))$). To do this, we write our eigenvector in the 2×2 $\tilde{c}$-model

$$\underline{A}_{\lambda=1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{\tilde{c}}{e_1 \left[E_1^0 + \tilde{c}\right] - \left[E_2^0 + \tilde{c}\right]} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{5.58}$$

and insert the equation for the matrix structure factor (Eq. (5.57)) for the general representation:

$$\underline{A}_{\lambda=1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{\tilde{c}}{\left[E_1^0 - E_2^0\right] \pm \tilde{c}} \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{5.59}$$

Without having to think long, we immediately see that, depending on the strength of the excitonic effects $\tilde{c}$ in the general case, the $\mu = 2$ entry of the eigenvector either diverges when $E_1^0 \approx E_2^0$ and $\tilde{c} \approx 0$ applies or the eigenvector component goes against a constant ($\tilde{c} \neq 0$). The obtained results confirm the difference we can expect depending on the strength of the excitonic effects, either a localized structure of the eigenvector ($\tilde{c} \approx 0$) or a broadened structure $\tilde{c} \neq 0$. Here and in the following we mean by localized that most of the information is located around the divergence in the eigenvector, while other information is suppressed by the divergence. However, if we take into account from the beginning that we are in the deep continuum ($E_1^0 \gg \min(E_\mu^0)$), i.e. $e_\lambda \to 1$, then we find an eigenvector structure which is more localized:

$$\underline{A}_{\lambda=1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{\tilde{c}}{\left[E_1^0 - E_2^0\right]} \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{5.60}$$

Summarizing the observation we can claim the following behavior for the eigenvectors according to our $\tilde{c}$-model:

1. The weaker the excitonic effects are, the more localized the eigenvector is up to a single pronounced peak.

2. The stronger the excitonic effects are, the broader the structure of the eigenvector will be and the weaker the magnitude.

3. Depending on the strength of the excitonic effects, the eigenvector moves towards a localized structure in the deep continuum.

It is important to note that, according to our $\tilde{c}$-models, the matrix structure factor (Eq. (5.57)) runs continuously towards 1.

Although this model is straightforward and indeed not the last word in wisdom, the results are very revealing. However, in order to test the analytical results, we will turn to a more realistic model in the following, in which we scale the interaction Hamiltonian of Si with a constant c to mimic weak and strong excitonic effects.
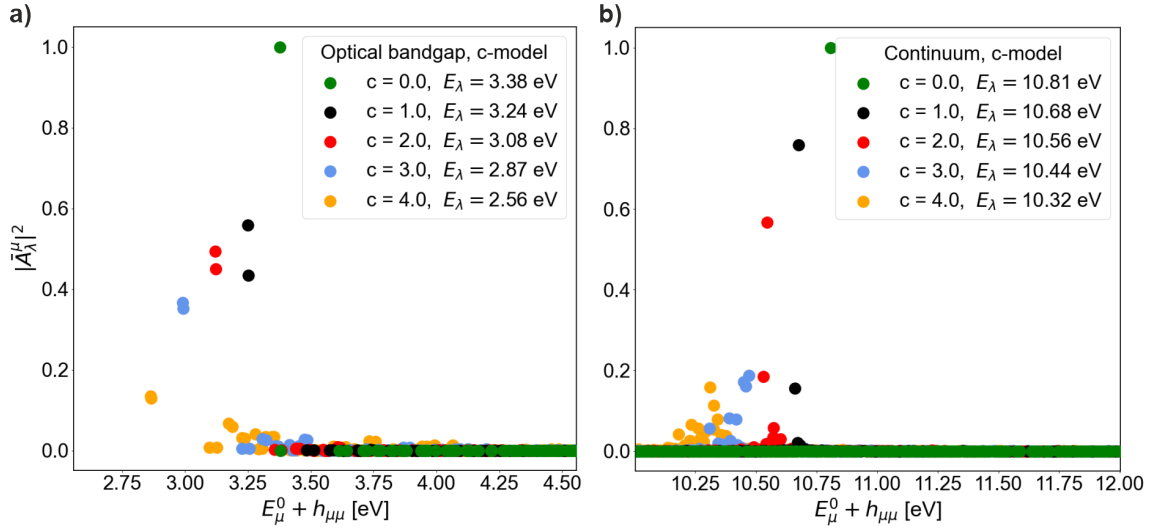
Figure 5.10: Shape of the normalized eigenvectors $\underline{\bar{A}}_\lambda$ (Eq. (5.27)), belonging to the eigenvalue E_λ, for the silicon c-model **a)** at the optical band edge ($E_\lambda = \min(M^{IP}(E_\lambda))$) and **b)** in the continuum ($E_\lambda \geqslant \min(M^{IP}(E_\lambda))$). Plotting the norm of the normalized eigenvector-components, $|\bar{A}^\mu_\lambda|^2$, against the diagonal elements $E^0_\mu + h_{\mu\mu}$ of the Hamiltonian $\underline{\underline{H}}^{2p,reso}$ (Eq. (4.13)). The figures show that with increasing exciton strength c, the eigenvector changes from a pronounced peak structure (green, black) systematically into a broader structure (light blue, orange). Furthermore, we see a shift towards smaller energies with increasing c-value. Numerical values given in Appendix A calculations were performed on a 3k grid (108 k-points) using a Kohn-Sham band structure, with 6 transition bands for the optical calculation.

If we compare the results from figure 5.10 with our predicted results from the 2×2 model, we can observe what we expected. The weaker the excitonic effects, the more the eigenvectors show a localized structure (green, black, red)(Eq. (5.46)), within the optical bandgap (Fig. 5.10 a)) and in the continuum (Fig. 5.10 b)). However, as soon as the excitonic interaction increases, the pronounced peak structure changes into a broader structure (light blue, orange) (Eq. (5.50)), but as we can see, the eigenvectors in the continuum still suggest a stronger localization then in the optical bandgap.

What is not clear at this stage, however, is how the structure of the eigenvectors changes with the

transition from the optical bandgap to the continuum to the deep continuum. Since the models can only give us an insight, but do not provide a complete picture, we want to make this observation using the real materials Si and LiF, and additionaly check whether our previous assumptions are valid for real materials.

5.3.2   Eigenvector structure of Silicon and Lithium fluoride

Before we use the acquired knowledge for approximations, we want to make sure that the behavior of the eigenvectors we derive from our models is also the one we observe in real materials. Furthermore, we want to find out how the eigenvectors change from the optical bandgap to the deep continuum. For this reason, in this section, we will consider the eigenvectors for both bound and continuum excitons for Si and LiF.
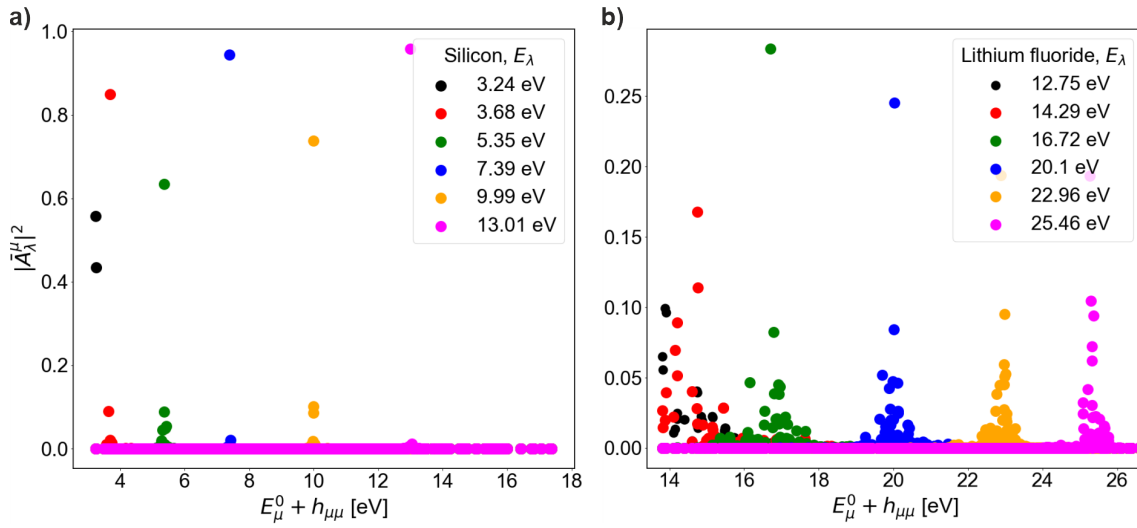
Figure 5.11: Shape of the normalized eigenvectors $\underline{\bar{A}}_\lambda$ (Eq. (5.27)), belonging to the eigenvalue E_λ, for **a)** silicon and **b)** lithium fluoride. Plotting the norm of the normalized eigenvector-components, $|\bar{A}_\lambda^\mu|^2$, against the diagonal elements $E_\mu^0 + h_{\mu\mu}$ of the Hamiltonian $\underline{\underline{H}}^{2p,reso}$ (Eq. (4.13)). Figure a) shows that for silicon (optical band edge $E_\lambda = 3.24$ eV), the eigenvectors show a pronounced peak as well as a highly localized structure both in the continuum (red, green, blue, orangen, magenta) and inside the optical bandgap (black), whereas for lithium fluoride (optical band edge $E_\lambda = 12.75$ eV), figure b), the eigenvectors show a broadened structure within the optical bandgap (black) to eigenvectors, which show a more localized structure the deeper we go into the continuum (green, blue, orange, magenta). Numerical values given in Appendix A calculations were performed on a 3k grid (108 k-points) using a Kohn-Sham band structure, with 6 transition bands for the optical calculation.

The results for the eigenvectors as shown in figure 5.11 are in agreement with our predictions: The eigenvectors for Si (Fig. 5.11 a)), a material with relatively weak excitonic interactions, show an apparent pronounced peak and highly localized structure. In the case of LiF (Fig. 5.11 b)), a material with strong excitonic effects, we see that the eigenvectors are generally much broader than for Si.

However, we also see clearly in the case of LiF and only weakly for Si is that the further we go into

Chapter 5: Analysis and perturbative approach to excitonic effects for absorption spectra,
eigenvectors and eigenvalues
102

the continuum, the more localized the structure of the eigenvectors become. This observation is consistent with what we were able to derive from our $\tilde{c}$-model for the matrix structure factor (Eq. (5.57)).

Thus, we have now not only understood how the eigenvectors behave depending on the strength of the excitonic effect, but have also gained the knowledge with which we can explain our observation regarding the matrix structure factor in Si and LiF (Fig. 5.9). Depending on the strength of the excitonic effects, the matrix structure factor converges systematically towards the value one for eigenvectors lying deep in the continuum (Eq. (5.57)). This in turn leads to the fact that the corresponding eigenvectors show a stronger localization than those within the optical bandgap (Eq. (5.60)). Like the matrix structure factor, the localization increases systematically the deeper the associated eigenenergy to the eigenvector lies in the continuum.

For the rest of the work we want to use the acquired knowledge to make accurate approximations for the eigenvalues and eigenvectors. The first and most straigthforward approximation based on our gained knowledge will be in the next section the so-called continuum approximation of the eigenenergies, which we have already derived for the $\tilde{c}$-model, namely the case $e_\lambda \to 1$. In the following sections, however, we want to proceed a little differently than before.

5.4 The continuum eigenenergy approximation

In the last section, we dealt with the structure of the eigenvectors and, in particular, their connection to the strength of the excitonic effects. Within our 2×2 model, we have shown (Eq. (5.56)) that if we are in the continuum of the eigenenergies, the matrix structure factor will tend to the value one ($e_\lambda \to 1$) and therefore the eigenenergies are given by $E_\lambda = E_\lambda^0 + h_{\lambda\lambda}$, which is what we observed for Si and LiF (Fig. 5.9). Since our 2×2 model is straightforward, we want to go beyond the model in this section. To do this, we first want to derive a formally exact equation for the eigenenergies, with the help of which we can see how the different components contribute to the eigenenergy and, thus, to the matrix structure factor. In the further course we want to test the individual components depending on the strength of the excitonic effects using real materials. In order to derive a formally exact equation for the eigenenergies, we start with the eigenvalue equation in braket notation

$$\hat{H} \left| A_\lambda \right\rangle = E_\lambda \left| A_\lambda \right\rangle.$$

The next step is to multiply the eigenvalue equation by the one vector or the λ eigenvector belonging to the diagonal Hamiltonian H^0 defined as $\left| A_\lambda^0 \right\rangle = \left| 1_\lambda \right\rangle$ from the left and use that we can write the eigenvectors according to the S-vector approach (Eq. (5.27)) as $\left| A_\lambda \right\rangle = \left| 1_\lambda \right\rangle + \left| L_\lambda \right\rangle$, where we have formally defined $\left| L_\lambda \right\rangle := \Delta \hat{\tilde{E}}_\lambda^{-1} \left(1 - \hat{\tilde{h}} \Delta \hat{\tilde{E}}_\lambda^{-1} \right)^{-1} \left| \tilde{h}_\lambda \right\rangle$. Under these conditions, we get the following expression:

$$\left\langle 1_\lambda | \hat{H} | 1_\lambda \right\rangle + \left\langle 1_\lambda | \hat{H} | L_\lambda \right\rangle = E_\lambda \left(\left\langle 1_\lambda | 1_\lambda \right\rangle + \left\langle 1_\lambda | L_\lambda \right\rangle \right).$$

Since by construction of the eigenvectors $\left\langle 1_\lambda | L_\lambda \right\rangle = 0$ holds and further the expectation value $\left\langle 1_\lambda | \hat{H} | 1_\lambda \right\rangle = E_\lambda^0 + h_{\lambda\lambda}$ is defined by the diagonal part of the Hamiltonian, we can further rewrite

the equation into:

$$E_\lambda = E_\lambda^0 + h_{\lambda\lambda} + \langle 1_\lambda | \hat{H} | L_\lambda \rangle. \tag{5.61}$$

If we now finally write out $\langle 1_\lambda | \hat{H} | L_\lambda \rangle$, using $\sum_{\mu'} H_{\lambda\mu'} \Delta \tilde{E}_{\mu'\mu'}^{-1}(E_\lambda) = \sum_{\mu'} \tilde{h}_{\lambda\mu'} \Delta \tilde{E}_{\mu'\mu'}^{-1}(E_\lambda)$, which follows from the fact that $\tilde{E}_{\mu'\mu'}^{-1}(E_\lambda) := (1 - \delta_{\mu'\lambda}) \frac{1}{E_\lambda - (E_{\mu'}^0 + h_{\mu'\mu'})}$ applies, we obtain the following equation for the eigenenergies:

$$E_\lambda = E_\lambda^0 + h_{\lambda\lambda} + \sum_{\mu\mu'} (1 - \delta_{\mu\lambda}) \frac{\tilde{h}_{\lambda\mu} \left(\underline{\underline{1}} - \underline{\underline{\tilde{h}}} \, \underline{\underline{\Delta\tilde{E}}}_\lambda^{-1} \right)_{\mu\mu'}^{-1} \tilde{h}_{\mu'\lambda}}{E_\lambda - (E_\mu^0 + h_{\mu\mu})}. \tag{5.62}$$

Equation (5.62) shows us that the eigenenergy is defined in lowest order by the diagonal part of the Hamiltonian $E_\lambda^0 + h_{\lambda\lambda}$ and all higher corrections of the energy depend on the eigenenergy itself. Before we delve deeper into the equation, we first want to derive the equation for the matrix structure factor. To do this, we insert equation (5.62) into the matrix structure factor dependent equation for the eigenenergies (Eq. (5.40)) and obtain

$$e_\lambda = 1 + \frac{1}{E_\lambda^0 + h_{\lambda\lambda}} \sum_{\mu\mu'} (1 - \delta_{\mu\lambda}) \frac{\tilde{h}_{\lambda\mu} \left(\underline{\underline{1}} - \underline{\underline{\tilde{h}}} \, \underline{\underline{\Delta\tilde{E}}}_\lambda^{-1} \right)_{\mu\mu'}^{-1} \tilde{h}_{\mu'\lambda}}{E_\lambda - (E_\mu^0 + h_{\mu\mu})} =: 1 + \frac{\bar{b}_\lambda/2}{E_\lambda^0 + h_{\lambda\lambda}}. \tag{5.63}$$

In the last step of equation (5.63) we have chosen the same notation as for the matrix structure factor $\bar{b}_\lambda = 2 \langle 1_\lambda | \hat{H} | L_\lambda \rangle$.

As the reformulated equation of the matrix structure factor (Eq. (5.63)) shows us, the fundamental structure is very similar to the one obtained from the 2×2 model (Eq. (5.56)). The matrix structure factor is built up from an offset with the value one and is corrected by the denominator, which depends on the independent-particle transition energy and the diagonal element of the electron-hole interaction part of the Hamiltonian associated with the index λ. Further, the numerator contains a λ dependent quantity. Thus, by substituting $h_{\lambda\lambda} = \bar{b}_\lambda/2 = \tilde{c}$ we immediately obtain the result for the 2×2 model (Eq. (5.56)). However, as we see, the reality is much more complex than the simple 2×2 model because the factor $\bar{b}_\lambda$ contains a large part of the information that can be found in the eigenvector itself, so this factor contains much more information than a simple constant for all λ can represent. The exciting thing, however, is that the equation for the matrix structure factor allows us to validate the statement from the 2×2 model that if we are in the continuum, the matrix structure factor approaches the value one ($e_\lambda \to 1$). Furthermore, the equation gives us the opportunity to check what we can expect in the thermodynamic limit ($N_k \to \infty$). For this, we want to assume that we can expand the inversion into a Neumann series and consider that $h_{\mu\mu'} = \frac{1}{N_k} \bar{h}_{\mu\mu'}$ (Eq. (5.4)) applies. Assuming that the Neumann series is valid, we then obtain the equation for the matrix structure factor:

$$e_\lambda = 1 + \frac{1}{E_\lambda^0 + \frac{1}{N_k} \bar{h}_{\lambda\lambda}} \frac{1}{N_k^2} \sum_{\mu\mu'} (1 - \delta_{\mu\lambda}) \frac{\tilde{\bar{h}}_{\lambda\mu} \sum_{n=0} \frac{1}{N_k^n} \left(\underline{\underline{\tilde{\bar{h}}}} \, \underline{\underline{\Delta\tilde{E}}}_\lambda^{-1} \right)_{\mu\mu'}^n \tilde{\bar{h}}_{\mu'\lambda}}{E_\lambda - (E_\mu^0 + \frac{1}{N_k} \bar{h}_{\mu\mu})}. \tag{5.64}$$

What we can conclude from equation (5.64), if we write out the first orders of the series expansion, is that the non-constant part of the matrix structure factor always has a pre-factor $\frac{1}{N_k}$ which is not compensated by a sum $\sum_\mu$ over N_k values. However, this means that if we enter the thermodynamic limit $N_k \to \infty$, the non-constant part will disappear and therefore in the thermodynamic limit $\lim_{N_k \to \infty} e_\lambda \to 1$ holds.

Nevertheless, this can only be part of the truth, because if this were the case, bound excitons would have to behave exactly like the continuum and, in particular, move closer to the independent-particle transition energies with increasing k-points. But as we have already observed for the matrix structure factor (Fig. 5.9), the factor becomes smaller for the strongly bound excitons with increasing k-points and thus moves away from the value one and consequently from the independent-particle transition energies.

In order to simulate the thermodynamic behavior and to check whether the non-constant part of the matrix structure factor actually disappears in the continuum, we want to consider the non-constant part for the two materials Si and LiF with increasing k-grid points.
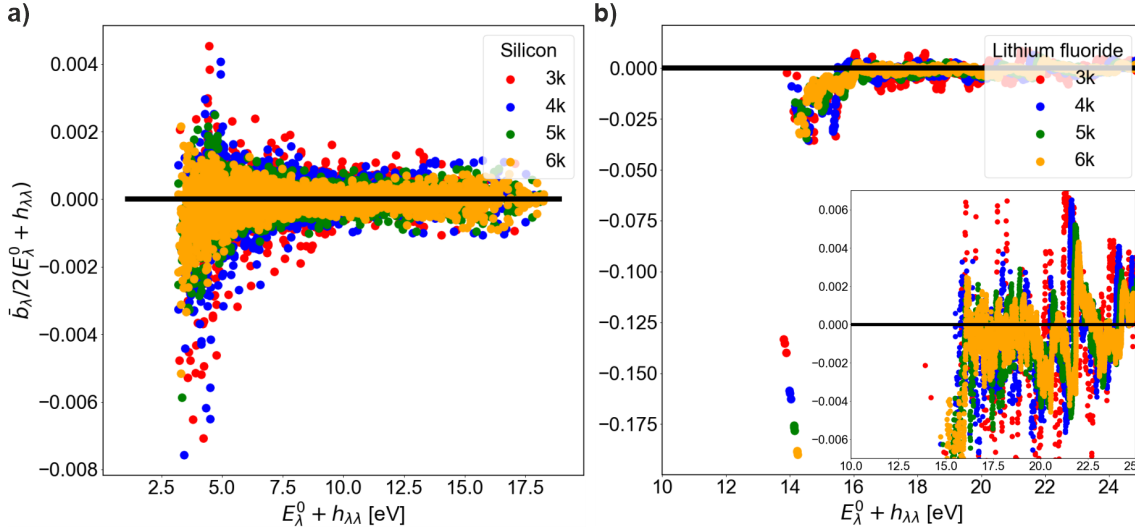
Figure 5.12: Thermodynamic approach of the non-constant component of the matrix structure factor $\frac{\bar{b}_\lambda/2}{E_\lambda^0 + h_{\lambda\lambda}}$ (Eq. (5.63)) against the diagonal elements $E_\lambda^0 + h_{\lambda\lambda}$ of the Hamiltonian $\underline{\underline{H}}^{2p,reso}$ (Eq. (4.13)). Results are shown for increasingly dense k-grids of 3k (red), 4k (blue), 5k (green) and 6k (orange), for **a)** silicon ($\min(M^{IP}(E_\mu^0)) = 3.27$ eV, 6k) and **b)** lithium fluoride ($\min(M^{IP}(E_\mu^0)) = 14.92$ eV, 6k). An approach towards the value zero in the continuum can be observed for both materials with increasing k-grid density. Numerical values are given in Appendix A; calculations were performed with 6 transition bands for the optical calculation using a Kohn-Sham band structure.

As we expected, according to the discussion in the 2×2 model case, the part $\frac{\bar{b}_\lambda/2}{E_\lambda^0 + h_{\lambda\lambda}}$ of the matrix structure factor (Eq. (5.63)) in the continuum approaches towards the value zero and thus $e_\lambda \to 1$. This can be seen particularly well for Si. However, it is also important to mention that the value for the bound excitons in LiF behaves the opposite way, becoming smaller and smaller instead of approaching zero.

At this point in time, however, it is physically and mathematically completely unclear why these excitons behave completely differently than the continuum. Apart from the fact that they are, by definition, not part of the continuum, we cannot give any argument why they do not approach it. But that is not a drawback; it is an interesting observation that these excitons behave completely differently. Nevertheless, we want to look further into the continuum and postulate, based on our observations and discussions, that in the continuum, the matrix structure factor converges to one,

and thus, the eigenenergies of the continuum are approximated by the continuum approximation

$$E_\lambda = E_\lambda^0 + h_{\lambda\lambda}. \tag{5.65}$$

To emphasize this, we will test the continuum approximation on the following four materials: Ge, Si, ZnS and LiF, which show increasing excitonic effects in ascending order.
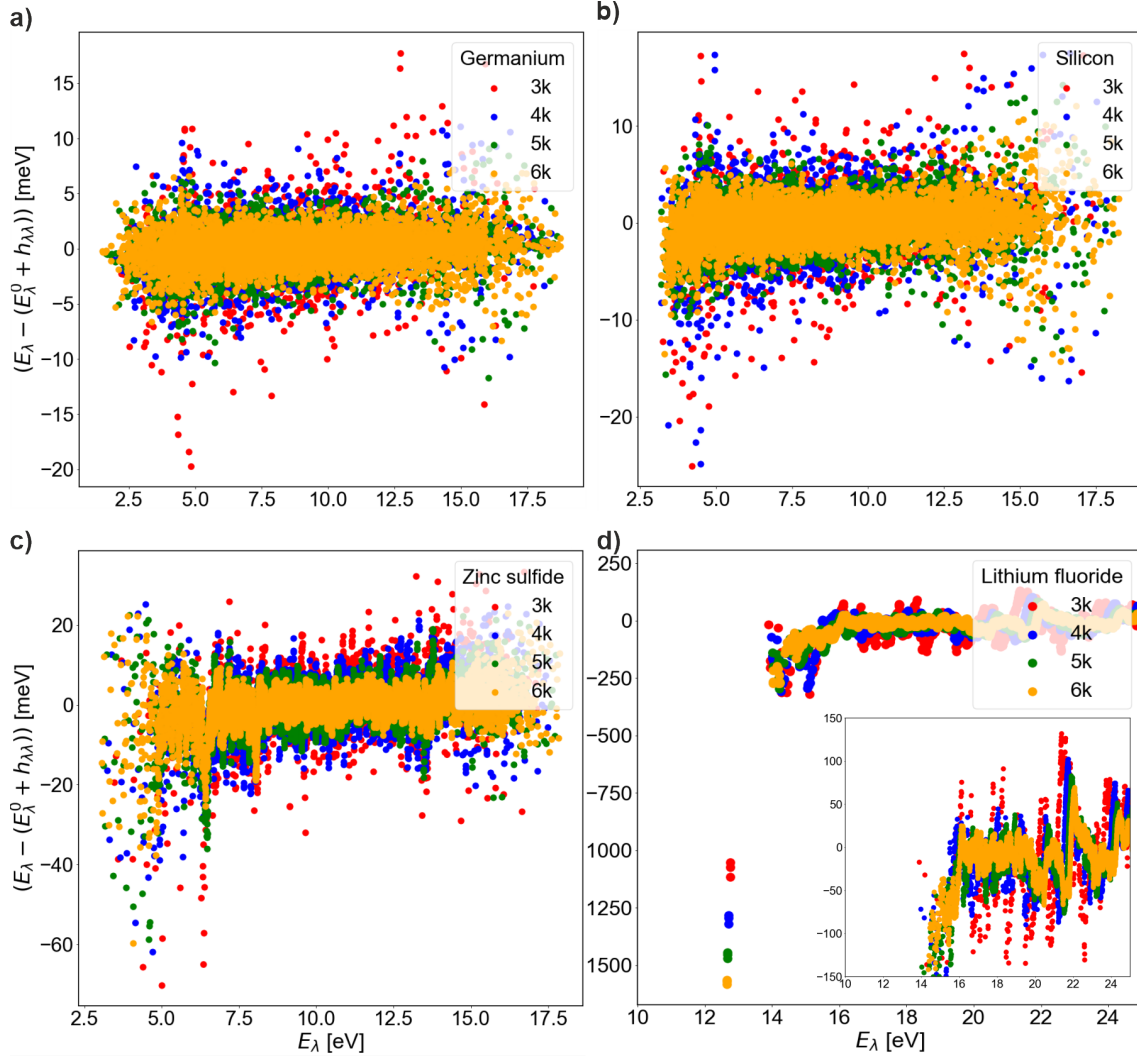

Figure 5.13: Thermodynamic approach of the continuum approximation of the eigenvalues (Eq. (5.65)) against the exact eigenvalues E_λ of the Hamiltonian $\underline{\underline{H}}^{2p,reso}$ (Eq. (4.13)). Results are shown for increasingly dense k-grids of 3k (red), 4k (blue), 5k (green) and 6k (orange), for **a)** germanium ($\min(M^{IP}(E_\mu^0)) = 1.53$ eV, 6k), **b)** silicon ($\min(M^{IP}(E_\mu^0)) = 3.27$ eV, 6k), **c)** zinc sulfide ($\min(M^{IP}(E_\mu^0)) = 3.26$ eV, 6k) and **d)** lithium fluoride ($\min(M^{IP}(E_\mu^0)) = 14.92$ eV, 6k). It is observable that for all four materials, a convergence from E_λ to $E_\lambda^0 + h_{\lambda\lambda}$ with increasing k-grid density takes place in the continuum. Numerical values are given in Appendix A; calculations were performed with 6 transition bands for the optical calculation using a Kohn-Sham band structure.

As predicted, the continuum converges to the continuum approximation of the eigenvalues in all four materials. What we also see, however, is that depending on the strength of the excitonic effects, convergence takes place much more slowly. This is consistent with the expectation from

the 2×2 model that the matrix structure factor depends on the strength of the excitonic effects
(Eq. (5.56)), even in the continuum. However, this dependence seems to be relatively small, which
is why the continuum approximation of the eigenenergies (Eq. (5.65)) provides an approximation
of the eigenenergies with an accuracy of a few to a few tenths of meV in the continuum. The
reason for this could be that if E_λ is in the continuum, the sum over μ in the equation (5.63)
includes partial cancellation between elements with positive and negative signs and thus only a
small proportion of the summation is finally involved in the value for the matrix structure factor.
If we compare the results of the continuum approximation with those of the independent-particle
transition energies (Fig. 4.3), it follows that the continuum approximation approximates the
eigenenergies of the continuum relatively accurately even for very small k-grids. In the case of
the pure independent-particle transition energy approximation, the deviations were in the order of
hundred meV even for Si and in the eV range for LiF. Furthermore, we see that even in LiF (Inset
Fig. 5.13 d)), a material with very strong excitonic effects, the deviation in the continuum between
approximation and exact values is in the range of a few tenths of a meV and becomes steadily
smaller with increasing k-points. However, it is important to note that in the thermodynamic
limit, i.e., the number of k-points approaches infinity, the continuum approximation changes to
the approximation using the independent-particle transition energies. Nevertheless, reaching this
limit in practice is almost impossible, because as we see in figure 5.14, even for non-vanishingly
small lattices of 865 k-points (6k), the diagonal part of the interaction matrix $h_{\lambda\lambda}$ for Si (Fig. 5.14
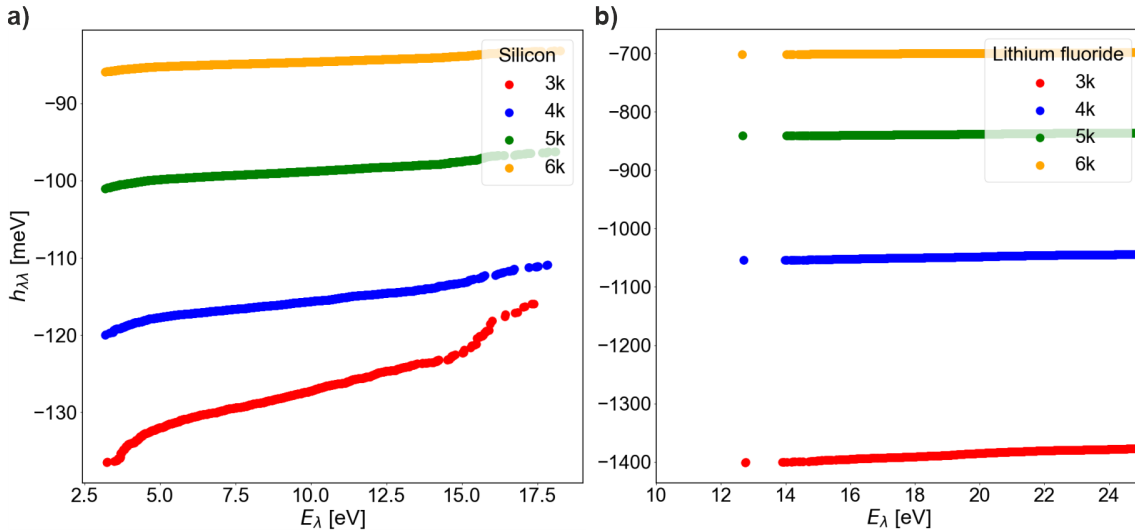a)) is in the order of -85 meV and for LiF (Fig. 5.14 b)) in the range of -700 meV.

Figure 5.14: Thermodynamic approach of the diagonal elements of the interaction matrix $h_{\lambda\lambda}$
against the exact eigenvalues E_λ of the Hamiltonian $\underline{\underline{H}}^{2p,reso}$ (Eq. (4.13)). Results are shown
for increasingly dense k-grids of 3k (red), 4k (blue), 5k (green) and 6k (orange), for **a)** silicon
$(\min(M^{IP}(E_\mu^0)) = 3.27$ eV, 6k) and **b)** lithium fluoride $(\min(M^{IP}(E_\mu^0)) = 14.92$ eV, 6k). It is
observable that a slow but systematic convergence of the diagonal elements of the interactions
matrix towards zero takes place. Numerical values are given in Appendix A; calculations were
performed with 6 transition bands for the optical calculation using a Kohn-Sham band structure.

Now that we have made the important observation that the continuum approximation of the

eigenenergies is much more practical than that of the independent-particle transition energies, although both are identical in the thermodynamic limit, we will look at their difference in practical use. For this purpose, we will investigate what difference the continuum approximation makes for the calculation of absorption spectra compared to the approximation of the independent-particle transition energies and the exact eigenenergies. This should clarify how important convergence in the small meV range for the eigenvalues is for calculating absorption spectra and, therefore, for practical applications.
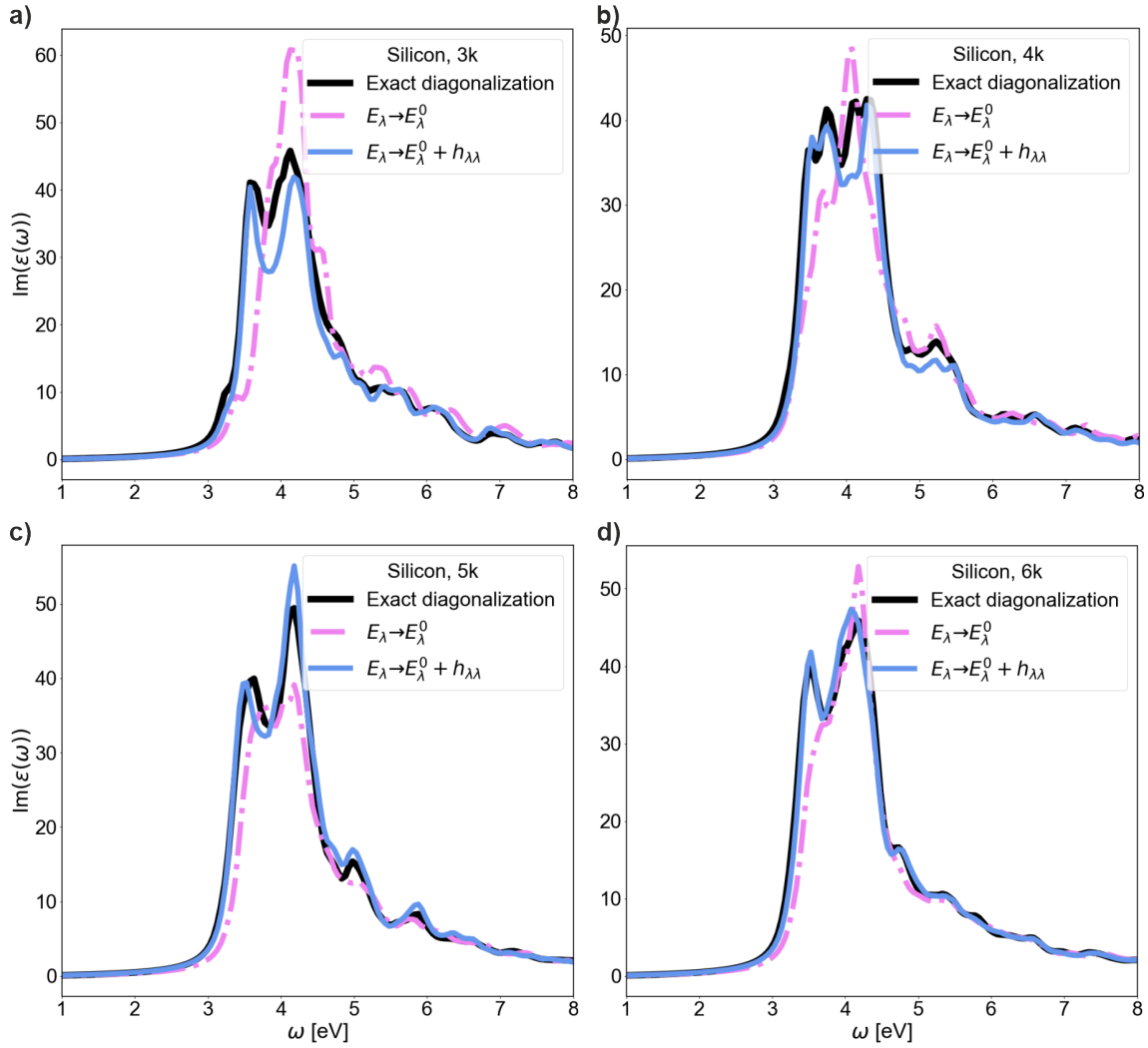
Figure 5.15: Thermodynamic approach of the absorption spectrum $(\mathrm{Im}(\epsilon(\omega))$ (Eq. (4.38), Eq. (4.3)) for silicon in static GW-BSE. Shown are the continuum approximation of the eigenenergies $E_\lambda \approx E_\lambda^0 + h_{\lambda\lambda}$ (Eq. (5.65)) (light blue) and the independent-particle transition energies approximation $E_\lambda \approx E_\lambda^0$ (magenta) in the exact eigenvector equation (Eq. (5.27)). For thermodynamic convergence, the exact spectrum (black) is compared to the approximations on a **a)** 3k, **b)** 4k, **c)** 5k and **d)** 6k k-grid. Results show that the continuum approximation converges relatively quickly to the exact result, unlike the independent-particle transition energies approximation results. Numerical values are given in Appendix A; calculations were performed with 6 transition bands for the optical calculation using a Kohn-Sham band structure and a Lorentzian broadening of $3.67 \cdot 10^{-3}$ Ha.

It is evident from Fig. 5.15 that as the k-grid density increases, the spectrum approaches the exact spectrum more closely for $E_\lambda \approx E_\lambda^0 + h_{\lambda\lambda}$ and exhibits only gradual changes for $E_\lambda \approx E_\lambda^0$. This observation confirms that the continuum approximation of the eigenenergies is well-suited for calculating observable quantities like the absorption spectrum. Especially for materials with weak excitonic effects. Furthermore, the result shows that even though in the thermodynamic limit $\lim_{N_k \to \infty} \left[E_\lambda^0 + \frac{1}{N_k} \bar{h}_{\lambda\lambda} \right] \to E_\lambda^0$ applies, this limit is numerically very difficult or even impossible to achieve for practical applications.

In summary, it can be said that the continuum approximation as we have postulated it (Eq. (5.65)) and the results (Fig. 5.13, Fig. 5.15) obtained indicate, the continuum approximation is an accurate approximation of the eigenvalues for the class of continuum excitons. Furthermore, the numerical advantage of this approximation is that the eigenvalues for the continuum are known as soon as the Hamiltonian has been constructed. Therefore, there are no additional numerical costs to calculate these values.

Now that we have an approximation for the eigenenergies of the continuum excitons, it is time to move towards eigenvectors and the eigenenergies of the strongly bound excitons. For this reason, the next section of this work will focus in particular on the question: *Can we approximate the eigenvectors using perturbation theory?*

5.5   Eigenvector perturbation approximation

The aim of this section is twofold; firstly, in addition to the insight that our equations of the eigenvectors offer, we also want to draw a numerical advantage. So far, due to the matrix inversion within the eigenvectors, the numerical scaling is still at $O(N^3)$, where N represents the dimension of the Hamiltonian. The calculation of the eigenvectors are therefore significantly too expensive for practical use compared to standard algorithms such as Divide-and-conquer [31], QR [27–29], inverse iteration [30], Jacobi-Davidson [33], Arnoldi [108,120], Lanczos [26,111] or the CG+SR [24] algorithm, which are well suited for large or small eigenvalues [33,121] and scale like $\sim O(N^2)$. Secondly, we want to gain a deeper understanding of perturbation series in the context of excitons because they are a well-known tool in physics and applied mathematics used to find approximated solutions for eigenvectors and eigenenergies. Furthermore, a series expansion, as we have seen in the case of the Neumann series expansion of the absorption spectrum (Ch. 5.1.2), allows us to get a deeper insight into the physics behind the problem, which enables us to build a deeper understanding.

Therefore, we will study two different kinds of perturbation theory in this section: Rayleigh-Schroedinger and Brillouin-Wigner perturbation theory. Furthermore, based on our previous results, we will explain why a perturbation-theoretic approach will challenge both continuum and strongly bound excitons and we will offer a way to circumvent these problems by modifying the perturbation theory.

5.5.1 Rayleigh-Schroedinger perturbation theory

The standard perturbation theory, which is taught in quantum-mechanic courses, is the Rayleigh-Schroedinger perturbation theory. For this reason, we also want to start with this approach. To avoid getting into unnecessarily complicated equations, we only want to concentrate on the perturbation series of the eigenvector up to the second order since this is sufficient for our analysis purposes, as we will see later.

Without specifying the Hamiltonian H more precisely, we split it in two parts:

$$\underline{\underline{H}} = \underline{\underline{\tilde{H}}}_0 + \underline{\underline{V}},$$

where $\underline{\underline{\tilde{H}}}_0$ represents the unperturbated contribution of the system and $\underline{\underline{V}}$ the perturbation of the system. With this definition, the perturbation series up to the second order results in

$$
A_\lambda^\mu = (A_\lambda^0)^\mu + \sum_{\lambda'}(1-\delta_{\lambda\lambda'})\frac{\sum_{\mu'\mu''}(A_{\lambda'}^0)^{\mu'}V_{\mu'\mu''}(A_\lambda^0)^{\mu''}}{\tilde{E}_\lambda^0 - \tilde{E}_{\lambda'}^0}(A_{\lambda'}^0)^\mu + \sum_{\lambda'}(1-\delta_{\lambda\lambda'})
$$
$$
\sum_{\lambda''}(1-\delta_{\lambda'\lambda''})\frac{\left(\sum_{\mu'\mu''}(A_{\lambda'}^0)^{\mu'}V_{\mu'\mu''}(A_{\lambda''}^0)^{\mu''}\right)\left(\sum_{\mu'\mu''}(A_{\lambda''}^0)^{\mu'}V_{\mu'\mu''}(A_\lambda^0)^{\mu''}\right)}{(\tilde{E}_\lambda^0 - \tilde{E}_{\lambda'}^0)(\tilde{E}_\lambda^0 - \tilde{E}_{\lambda''}^0)}(A_{\lambda'}^0)^\mu - \qquad (5.66)
$$
$$
\sum_{\lambda'}(1-\delta_{\lambda\lambda'})\frac{\left(\sum_{\mu'\mu''}(A_{\lambda'}^0)^{\mu'}V_{\mu'\mu''}(A_\lambda^0)^{\mu''}\right)\left(\sum_{\mu'\mu''}(A_\lambda^0)^{\mu'}V_{\mu'\mu''}(A_\lambda^0)^{\mu''}\right)}{(\tilde{E}_\lambda^0 - \tilde{E}_{\lambda'}^0)^2}(A_{\lambda'}^0)^\mu + \dots
$$

with $(A_\lambda^0)^\mu$ the μ-th component of the unperturbated eigenvector corresponding to the unperturbated Hamiltonian $\underline{\underline{\tilde{H}}}_0$ and the corresponding eigenvalues $\tilde{E}_\lambda^0 = \langle A_\lambda^0|\hat{\tilde{H}}_0|A_\lambda^0\rangle$. The next step is to specify the Hamiltonian more precisely. Since this is also about practical use, we must choose the unperturbed Hamiltonian in such a way that the corresponding eigenvector $\underline{A}_\lambda^0$ can be calculated as efficiently as possible. For this purpose, we first choose in general that the unperturbed Hamiltonian matrix represents a diagonal matrix, and thus the following applies $(A_\lambda^0)^\mu = \delta_{\mu\lambda}$. If we insert the expression for the unperturbed eigenstates into the perturbation series (Eq. (5.66)), we obtain:

$$
A_\lambda^\mu = \delta_{\mu\lambda} + \frac{(1-\delta_{\mu\lambda})}{\tilde{E}_\lambda^0 - \tilde{E}_\mu^0}V_{\mu\lambda} + \frac{(1-\delta_{\mu\lambda})}{\tilde{E}_\lambda^0 - \tilde{E}_\mu^0}\sum_{\mu'}\frac{(1-\delta_{\mu'\lambda})}{\tilde{E}_\lambda^0 - \tilde{E}_{\mu'}^0}V_{\mu\mu'}V_{\mu'\lambda} - \frac{(1-\delta_{\mu\lambda})}{(\tilde{E}_\lambda^0 - \tilde{E}_\mu^0)^2}V_{\mu\lambda}V_{\lambda\lambda} + \cdots . \quad (5.67)
$$

Now that we have determined the structure of the perturbation-theoretical eigenvector, we must ask ourselves how we define the perturbation. A standard approach in physics would be to say that the excitonic interactions $\underline{\underline{V}} = \underline{\underline{\Xi}}^{AGW}$ represent the perturbation of the non-interacting system $\underline{\underline{H}}^0$. Thus, the eigenvalues $\tilde{E}_\lambda^0$ of this system would inevitably be given by the independent-particle transitions energies $\tilde{E}_\lambda^0 = E_\lambda^0$. However, as we have seen in numerous examples (Fig. 4.3, Fig. 5.15), the assumption that these energies are a sufficiently good approximation of the eigenenergies is not given, neither for bound states nor for the continuum. Therefore, we choose the off-diagonal elements of the interaction Hamiltonian $V_{\mu\mu'} = \tilde{h}_{\mu\mu'} = (1-\delta_{\mu\mu'})h_{\mu\mu'}$ to represent the perturbation for the entire diagonal part of the Hamiltonian $(\underline{\underline{\tilde{H}}}_0)_{\mu\mu'} = \delta_{\mu\mu'}H_{\mu\mu'}^0 + \delta_{\mu\mu'}h_{\mu\mu'}$. In this case, we could say that our unperturbed system is described by the independent-particle transition energies and the diagonal term of the electron-hole interaction matrix, while the perturbation of our system is defined as the off-diagonal part of the electron-hole interaction matrix. This assumption would lead to the eigenenergies being approximated in the zero-order by the continuum approximation (Eq.

(5.65)) $\tilde{E}_\lambda^0 = E_\lambda^0 + h_{\lambda\lambda}$ and as we know (Ch. 5.4), this is a significantly better approximation of the eigenenergies than the independent-particle transition energy approximation of the eigenenergies. If we insert the latter assumption into the perturbation series (Eq. (5.67)) of the eigenvector, we get

$$A_\lambda^\mu = \delta_{\mu\lambda} + \frac{(1 - \delta_{\mu\lambda})}{(E_\lambda^0 + h_{\lambda\lambda}) - (E_\mu^0 + h_{\mu\mu})}\tilde{h}_{\mu\lambda}$$
$$+ \frac{(1 - \delta_{\mu\lambda})}{(E_\lambda^0 + h_{\lambda\lambda}) - (E_\mu^0 + h_{\mu\mu})} \sum_{\mu'} \frac{(1 - \delta_{\mu'\lambda})}{(E_\lambda^0 + h_{\lambda\lambda}) - (E_{\mu'}^0 + h_{\mu'\mu'})}\tilde{h}_{\mu\mu'}\tilde{h}_{\mu'\lambda} + \dots.$$

(5.68)

The last term of the second order in equation (5.67) disappears, since $\tilde{h}_{\lambda\lambda} = (1 - \delta_{\lambda\lambda})h_{\lambda\lambda} = 0$ holds. Furthermore, the structure of equation (5.68) is all too familiar to us; namely it represents the Neumann series up to the second order for our eigenvector equation, obtained from the S-vector approach (Eq. (5.27)) in the continuum approximation ($E_\lambda = E_\lambda^0 + h_{\lambda\lambda}$). To see this we rewrite equation (5.68) into a matrix-vector equation

$$\underline{A}_\lambda = \underline{1}_\lambda + \underline{\underline{\Delta\tilde{E}}}_\lambda^{-1}\underline{\tilde{h}}_\lambda + \underline{\underline{\Delta\tilde{E}}}_\lambda^{-1}\underline{\underline{\tilde{h}}}\,\underline{\underline{\Delta\tilde{E}}}_\lambda^{-1}\underline{\tilde{h}}_\lambda = \underline{1}_\lambda + \underline{\underline{\Delta\tilde{E}}}_\lambda^{-1} \sum_{k=0}^{1} \left(\underline{\underline{\Delta\tilde{E}}}_\lambda^{-1}\underline{\underline{\tilde{h}}}\right)^k \underline{\tilde{h}}_\lambda,$$

(5.69)

where we defined $\left(\Delta\tilde{E}_{\mu\mu'}(E_\lambda)\delta_{\mu\mu'}\right)^{-1} := \frac{(1-\delta_{\mu\lambda})}{(E_\lambda^0 + h_{\lambda\lambda}) - (E_\mu^0 + h_{\mu\mu})}$.

However, as we know from our analysis of the eigenvector structure (Ch. 5.3.1), the perturbation series in the current form will not work straightforward for a dense k-grid since the denominators of $\left(\Delta\tilde{E}_{\mu\mu'}(E_\lambda)\delta_{\mu\mu'}\right)^{-1}$ diverges and therefore the series expansion breaks down. Although it seems like a degeneration problem or the like, it is not since the problem lies in what we want to describe in a perturbation series. The class of perturbation theories we are discussing here are also called bound-state perturbation theories [122], which indicates what they are made for: Describing eigenstates, which are far enough from the continuum of eigenstates in a series expansion. In our case, however, the energies $E_\lambda^0 + h_{\lambda\lambda}$ lie in the continuum and are therefore not well separated, which is why the series diverges. On the other hand, if the energies did not lie within the continuum, nothing should formally stand in the way of a perturbation theory approach. This leads us to the more general form of perturbation theory, the Brillouin-Wigner perturbation theory.

5.5.2 Brillouin-Wigner perturbation theory

As we have seen in the last section, the predominant problem of the Rayleigh-Schroedinger perturbation theory for calculating excitonic eigenstates lies in the fact that we are tied to the continuum approximation of the eigenenergies, which unfortunately lie close together. It therefore makes sense to search for alternative options. For this reason, we will look at a type of perturbation theory that is less well known than Rayleigh-Schroedinger and is called Brillouin-Wigner perturbation theory [122–124]. For deriving the perturbation series, we use the procedure from [122], which is very clear and instructive. In the following, we will only use the non-degenerate perturbation series and keep the notation as before, except that we use the Dirac notation to keep the derivation as straightforward as possible. We will assume as in the case of the Rayleigh-Schroedinger perturbation theory that the Hamiltonian $\underline{\underline{H}}$ is splitted into the sum of the independent Hamiltonian $\underline{\underline{\tilde{H}}}_0$ with eigenvectors $\underline{A}_\lambda^0$ and eigenvalues $\tilde{E}_\lambda^0$, and the perturbation $\underline{\underline{V}}$. The eigenvalue equation for

our system is given as before by

$$\hat{H}\,|A_\lambda\rangle = E_\lambda\,|A_\lambda\rangle. \tag{5.70}$$

Before we start with the derivation of the perturbation series, we want to define some operators for the further calculations:

$$\hat{P}_\lambda := |A_\lambda^0\rangle\langle A_\lambda^0|\,, \tag{5.71}$$

$$\hat{Q}_\lambda := \sum_{\lambda'\neq\lambda} |A_{\lambda'}^0\rangle\langle A_{\lambda'}^0| \tag{5.72}$$

and

$$\hat{R}_\lambda := \sum_{\lambda'\neq\lambda} \frac{|A_{\lambda'}^0\rangle\langle A_{\lambda'}^0|}{E_\lambda - \tilde{E}_{\lambda'}^0}. \tag{5.73}$$

The first step in the direction of a perturbation series is to rewrite the eigenvalue equation (Eq. (5.70)) into the following form:

$$\left(E_\lambda - \hat{\tilde{H}}^0\right)|A_\lambda\rangle = \hat{V}\,|A_\lambda\rangle. \tag{5.74}$$

Next, we apply the $\hat{R}_\lambda$ operator (Eq. (5.73)) from the left to rewrite the eigenvalues equation (Eq. (5.74)):

$$\hat{R}_\lambda\left(E_\lambda - \hat{\tilde{H}}^0\right)|A_\lambda\rangle = \hat{Q}_\lambda\,|A_\lambda\rangle = \hat{R}_\lambda\hat{V}\,|A_\lambda\rangle. \tag{5.75}$$

Since our goal is to obtain an equation for $|A_\lambda\rangle$, we add $\hat{P}_\lambda\,|A_\lambda\rangle$ on both sides of equation (5.75) and use the fact that $\hat{P}_\lambda + \hat{Q}_\lambda = 1$ holds, which leads us to the following equation for $|A_\lambda\rangle$:

$$|A_\lambda\rangle = \hat{Q}_\lambda\,|A_\lambda\rangle + \hat{P}_\lambda\,|A_\lambda\rangle = \hat{P}_\lambda\,|A_\lambda\rangle + \hat{R}_\lambda\hat{V}\,|A_\lambda\rangle. \tag{5.76}$$

The obtained equation shows the well-known Dyson-like structure, which, as we have seen many times in this work, can be solved by subtraction and inversion, which then gives us the exact equation for $|A_\lambda\rangle$ and thus our starting equation for the perturbation series:

$$|A_\lambda\rangle = (1 - \hat{R}_\lambda\hat{V})^{-1}\hat{P}_\lambda\,|A_\lambda\rangle = (1 - \hat{R}_\lambda\hat{V})^{-1}\langle A_\lambda^0|A_\lambda\rangle\,|A_\lambda^0\rangle. \tag{5.77}$$

A convention that is usually adopted is to choose the normalization and phase of $|A_\lambda\rangle$ in such a way that: $\langle A_\lambda^0|A_\lambda\rangle = 1$. The last and decisive step in order to obtain a series expansion for the eigenvector (Eq. (5.77)) is then to expand $(1 - R_\lambda V)^{-1}$ into a Neumann series (Eq. (5.14)), which results in the perturbation series for the eigenvector $\underline{A}_\lambda$ and can be written in matrix-vector notation as:

$$\underline{A}_\lambda = \sum_{k=0}^{m_{PT}} \left(\underline{\underline{R}}_\lambda\underline{\underline{V}}\right)^k \underline{A}_\lambda^0. \tag{5.78}$$

In order to see how the obtained perturbation series (Eq. (5.78)) differs from the Rayleigh-Schroedinger series, we proceed as before:

The perturbation matrix is given as the off-diagonal matrix of the interaction part $V_{\mu\mu'} = \tilde{h}_{\mu\mu'} = (1 - \delta_{\mu\mu'})h_{\mu\mu'}$ and the unperturbed system as the diagonal matrix consisting of independent-particle transition energies and the diagonal part of the interaction matrix $(\underline{\underline{\tilde{H}}}_0)_{\mu\mu'} = \tilde{E}_\mu^0\delta_{\mu\mu'} = \delta_{\mu\mu'}H_{\mu\mu'}^0 + \delta_{\mu\mu'}h_{\mu\mu'}$. Under this assumption, the following equation results for the Brillouin-Wigner perturbation series in component notation up to second order:

$$
\begin{aligned}
A_\lambda^\mu = {}&\delta_{\mu\lambda} + \frac{(1 - \delta_{\mu\lambda})}{E_\lambda - (E_\mu^0 + h_{\mu\mu})}\tilde{h}_{\mu\lambda} \\
&+ \frac{(1 - \delta_{\mu\lambda})}{E_\lambda - (E_\mu^0 + h_{\mu\mu})}\sum_{\mu'} \frac{(1 - \delta_{\mu'\lambda})}{E_\lambda - (E_{\mu'}^0 + h_{\mu'\mu'})}\tilde{h}_{\mu\mu'}\tilde{h}_{\mu'\lambda} + \dots.
\end{aligned}
\tag{5.79}
$$

where we used $(A^0_\lambda)^\mu = \delta_{\mu\lambda}$.

If we compare the Brillouin-Wigner (Eq. (5.79)) and the Rayleigh-Schroedinger perturbation series (Eq. (5.68)), we immediately see that the only difference between the two series is that in the Brillouin-Wigner we find the exact eigenvalue E_λ in the denominator, while in the Rayleigh-Schroedinger we find the continuum approximation $E_\lambda = E^0_\lambda + h_{\lambda\lambda}$ (Eq. (5.65)) of the eigenenergies, which is the zeroth order energy of the expansion obtained by multiplying equation (5.74) from the left by $\langle A^0_\lambda|$ and substituting equation (5.79) for $|A_\lambda\rangle$:

$$E_\lambda = E^0_\lambda + h_{\lambda\lambda} + \sum_{\mu'}(1 - \delta_{\mu'\lambda})\frac{\tilde{h}_{\lambda\mu'}\tilde{h}_{\mu'\lambda}}{E_\lambda - (E^0_{\mu'} + h_{\mu'\mu'})} + \dots . \tag{5.80}$$

Even though this difference between the two series expansions seems to be small, we know that this is not true for strongly bound states (Fig. 5.13 d)). In the case of strongly bound excitons, the difference between the exact eigenvalue E_λ and $E^0_\lambda + h_{\lambda\lambda}$ is of the order of eV and therefore much larger than in the continuum (meV for a reasonable k-grid). For this reason, a series expansion in the Brillouin-Wigner perturbation theory for bound states may work. Nevertheless, we cannot work that easily with the series expansion since we do not know the eigenenergies and even worse we cannot use equation (5.80) to calculate corrections to the eigenenergies, since both sides depend on the exact eigenenergy E_λ. Even though we were to substitute the equation into itself

$$E_\lambda = E^0_\lambda + h_{\lambda\lambda} + \sum_{\mu'}(1 - \delta_{\mu'\lambda})\frac{\tilde{h}_{\lambda\mu'}\tilde{h}_{\mu'\lambda}}{(E^0_\lambda + h_{\lambda\lambda}) - (E^0_{\mu'} + h_{\mu'\mu'})} + \dots , \tag{5.81}$$

we cannot calculate higher order corrections since we would have the problem of divergences because the energies in the continuum are too close together.

Summarizing what we have found: The perturbation theory in the representation of Brillouin and Wigner could, in principle, work for strongly bound excitonic states; however, since the exact eigenvalues are unknown, we cannot easily use the method for practical applications. Furthermore, since the eigenvalues for continuum excitons are insufficiently separated, a perturbation theoretical description cannot work in a straightforward way. Nevertheless, this does not mean that it is impossible to modify the perturbation theory in such a way that it can work at least for an approximation of the eigenvectors in the continuum.

5.5.3 The modified perturbation theory

As the last section made clear to us, we will not get very far with an ordinary perturbation series in the continuum of energies; the reason for this is that the eigenvectors are formed by a divergence. However, we have already shown that a perturbation theoretical description of absorption spectra for materials with weak excitonic interactions is, in principle, possible (Ch. 5.1.2). This was possible since we introduced the parameter broadening, which prevented the frequency ω from being too close to a value $E^0_\mu + h_{\mu\mu}$. Therefore, the main idea of this section is to introduce a modified perturbation theory, which is formally exact but has the opportunity to reduce the divergence such that an approximation for the eigenvectors is possible. This should hold for bound as well as continuum excitons. Since we want to have the perturbation series modified by a control parameter of the divergence, so to say an artificial broadening, our natural choice of an exact

starting equation is the G-matrix approach of the eigenvectors (Eq. (5.39)), as this approach already includes a broadening factor γ.

However, as can be shown (Appendix G), it makes no difference on the optically visible scale whether we start from the G-matrix approach and develop a perturbation series or whether we artificially introduce a broadening parameter into the Brillouin-Wigner perturbation series. To avoid unnecessarily complicated formulations, we want to introduce the broadening parameter directly into the Brillouin-Wigner perturbation series and call this series the modified perturbation theory:

$$
\begin{aligned}
A_\lambda^\mu = \delta_{\mu\lambda} + \frac{(1 - \delta_{\mu\lambda})}{(E_\lambda + i\gamma) - (E_\mu^0 + h_{\mu\mu})}\tilde{h}_{\mu\lambda} \\
+ \frac{(1 - \delta_{\mu\lambda})}{(E_\lambda + i\gamma) - (E_\mu^0 + h_{\mu\mu})} \sum_{\mu'} \frac{(1 - \delta_{\mu'\lambda})}{(E_\lambda + i\gamma) - (E_{\mu'}^0 + h_{\mu'\mu'})}\tilde{h}_{\mu\mu'}\tilde{h}_{\mu'\lambda} + \ldots,
\end{aligned}
\tag{5.82}
$$

which in vector-matrix representation would be given as:

$$
\underline{A}_\lambda = \underline{1}_\lambda + \underline{\underline{\Delta\tilde{E}}}_\lambda^{-1}(\gamma) \sum_{k=0}^{m_{PT}} \left(\underline{\underline{\tilde{h}}}\,\underline{\underline{\Delta\tilde{E}}}_\lambda^{-1}(\gamma)\right)^k \underline{\tilde{h}}_\lambda
\tag{5.83}
$$

After we have introduced the modified perturbation series, we now want to see what results we can achieve with it in the limit of weak excitonic effects and in the limit of strongly bound excitons, for which the Brillouin-Wigner perturbation theory should be made.

5.6   Modified perturbation theory: From weak to strong excitonic effects

After having discussed in the last chapter why a perturbation theory approach using the Rayleigh-Schroedinger or Brillouin-Wigner perturbation theory cannot be applied in a straightforward way for material with weak excitonic interactions, we want to test the modified perturbation theory (Eq. (5.83)) for these materials in the following. The advantage of the modified perturbation theory is that it has a parameter γ with which we can reduce the divergence of $\underline{\underline{\Delta\tilde{E}}}_\lambda^{-1}(\gamma)$, which as we know from section 5.1.2 is significantly involved into the magnitude of the spectral radius $S_r\left(\underline{\underline{\tilde{h}}}\,\underline{\underline{\Delta\tilde{E}}}_\lambda^{-1}(\gamma)\right)$. It is clear that if the modified perturbation theory for $\gamma \to 0$ converges, we would have a working Brillouin-Wigner perturbation series and, thus, an exact series that would give us back the eigenvectors.

We expect, that for continuum excitons, we will only obtain an approximation of the eigenvectors. However, we will show that in the case of strongly bound states in LiF, the limit case $\gamma \to 0$ applies, and therefore, a quite accurate calculation is possible.

5.6.1   Modified perturbation theory appproach for excitonic states in Silicon

One of the significant problems with the straightforward perturbation series is that it is only suitable for well-separated eigenenergies. However, many bulk materials are materials with weak excitonic interactions and thus only have weak bound and continuum excitons. It is precisely for these materials that a perturbation series is a bad approach, as divergences will occur in the series

expansion, and thus, the series expansion will collapse. In the following, we want to show that with the help of the modified perturbation series (Eq. (5.83)), it is possible to find the excitonic eigenvectors of materials with weak excitonic effects in an approximated form.

Since the modified perturbation series (Eq. (5.83)) depends on the exact eigenvalues E_λ, which we do not know we will apply the continuum approximation to the eigenenergies (Eq. (5.65)) $E_\lambda = E_\lambda^0 + h_{\lambda\lambda}$, which guarantees us an accuracy of the eigenenergies down to a few meV (Ch. 5.4). Further, we need to know our broadening parameter γ for the practical application of the modified perturbation series, but we do not have a simple mathematical condition with which we can calculate the ideal value, we have to determine it numerically beforehand. The aim here must be to choose it as small as possible (high accuracy) but so large that the condition $S_r\left(\tilde{\underline{\underline{h}}}\,\underline{\underline{\Delta\tilde{E}}}_\lambda^{-1}(\gamma)\right) < 1$ is fulfilled. Furthermore, the spectral radius should not be driven too close to one; otherwise, the number of iteration steps until convergence will be large.

In order to keep the numerical costs as low as possible, we replace E_λ with $\tilde{\omega}$ in $\underline{\underline{\Delta\tilde{E}}}_\lambda^{-1}(\gamma)$ to calculate the spectral radius $S_r\left(\tilde{\underline{\underline{h}}}\,\underline{\underline{\Delta E}}_{\tilde{\omega}}^{-1}(\gamma)\right)$ for just a few energies and check how it behaves. This approach will help us to obtain an estimate of the broadening at low numerical cost.
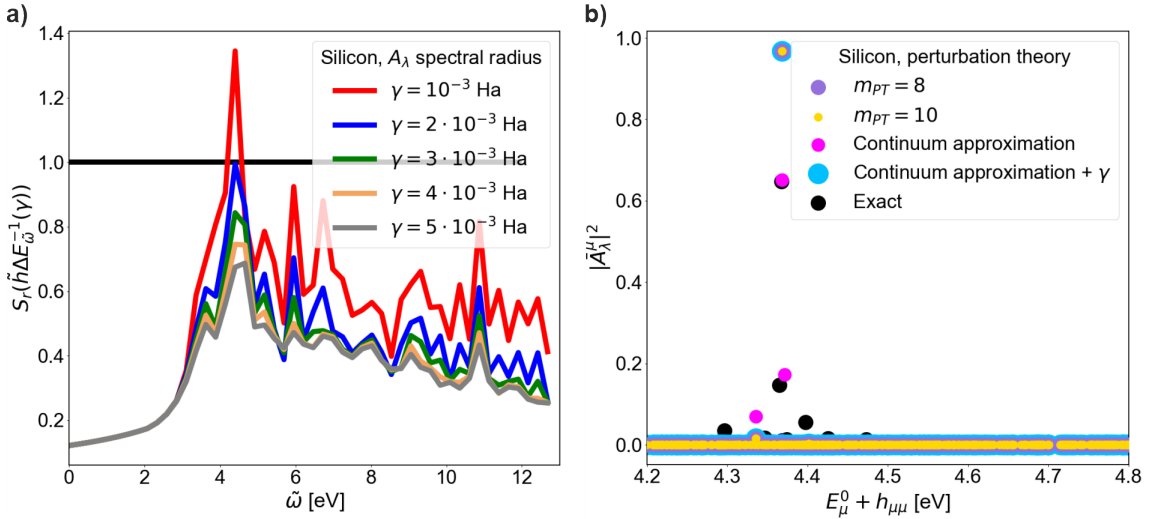

Figure 5.16: Modified perturbation theory for the excitonic eigenvectors $\underline{\bar{A}}_\lambda$ (Eq. (5.83)) belonging to the eigenvalues $E_\lambda = E_\lambda^0 + h_{\lambda\lambda}$ (Eq. (5.65)) in the continuum eigenvalue approximation for silicon. Shown is **a)** the spectral radius within the modified perturbation series depending on the selected broadening γ, against selected energies $\tilde{\omega}$ and **b)** the norm of the normalized eigenvector-components, $|\bar{A}_\lambda^\mu|^2$, against the diagonal elements $E_\mu^0 + h_{\mu\mu}$ of the Hamiltonian $\underline{\underline{H}}^{2p,reso}$ (Eq. (4.13)) for different orders in the perturbation series as well as the exact eigenvector (black) and the non-perturbativ eigenvector in the continuum eigenvalue approximation without (magenta) and with broadening (light blue). The result a) shows that with a sufficiently significant broadening ($\gamma = 3 \cdot 10^{-3}$ Ha), a perturbation theoretical description of the eigenvectors is possible. Furthermore, the obtained eigenvectors b) show a significantly stronger localized and pronounced peak structure than the exact and continuum-approximated eigenvectors. Numerical values are given in Appendix A; calculations were performed on a 5k grid (500 k-points) using a Kohn-Sham band structure, with 6 transition bands for the optical calculation.

As the result for the spectral radius depending on the broadening (Fig. 15 a)) shows us, if we are interested in the entire spectrum of eigenvectors, the broadening parameter $\gamma = 3 \cdot 10^{-3}$ Ha is a good choice because $\gamma = 2 \cdot 10^{-3}$ Ha already leads to a spectral radius very close to one in a certain range. If we use this broadening to calculate the eigenvectors in the continuum approximation of eigenenergies (Eq. (5.65)) and compare the result with the exact eigenvector (black), as well as those in the continuum approximation of eigenenergies without (magenta) and with broadening (light blue), then we see that the approximated eigenvectors have a more localized and pronounced peak structure compared to the exact or continuum approximation. At first glance, this seems contradictory because one would expect that if the eigenvectors in the continuum are determined by the divergence of $\underline{\underline{\Delta \tilde{E}}}_{\lambda}^{-1}(\gamma)$, then with greater broadening the divergence should become flatter and thus the eigenvectors should deviate more from the pronounced peak structure. It is certainly true that the divergence makes an important contribution. However, it is probably the case that the inversion of the matrix $\left(\underline{\underline{1}} - \underline{\underline{\tilde{h}}} \, \underline{\underline{\Delta \tilde{E}}}_{\lambda}^{-1}(\gamma) \right)^{-1}$ provides exactly the opposite, i.e. weakens the divergence. At least that is what we observed in the Neumann series of the absorption spectrum (Fig. 5.6). The divergence increases the spectrum in all areas, but the off-diagonals from higher orders of the Neumann series reduce this effect. We cannot observe this effect to the necessary extent for the modified perturbation theory of the eigenvectors, because besides the reduction of the divergence, we suppress the off-diagonal elements too strongly. For this reason, the eigenvectors show a more localized and pronounced peak structure than the exact eigenvectors.

In general, what we can observe is that each approximation we make changes a little more information in the eigenvector. Already the eigenvector of the continuum approximation is different from the exact eigenvector and with the additional introduction of a non-vanishing broadening even more information changes.

Indeed, this may seem like a major setback for using the modified perturbation series. However, not only the component in the strongest part of the eigenvector determines its usefulness. Because when we calculate the eigenvector to use it for applications such as absorption spectra, exciton-exciton excitation-spectra, EELS, and the like, it is not the pure eigenvector that plays the decisive role, but for example, the multiplication between the eigenvector and another vector or a matrix. So, not a single component of the eigenvector plays the decisive role, but the entire eigenvector. Furthermore, most application quantities depend on the sum of many eigenvectors and eigenenergies. Therefore, it is still possible to get a good result for the quantities we are interested in with a reasonable result for the eigenvectors at first sight. To demonstrate this, we will look at the absorption spectrum for Si calculated with the approximated eigenvectors from the modified perturbation theory.
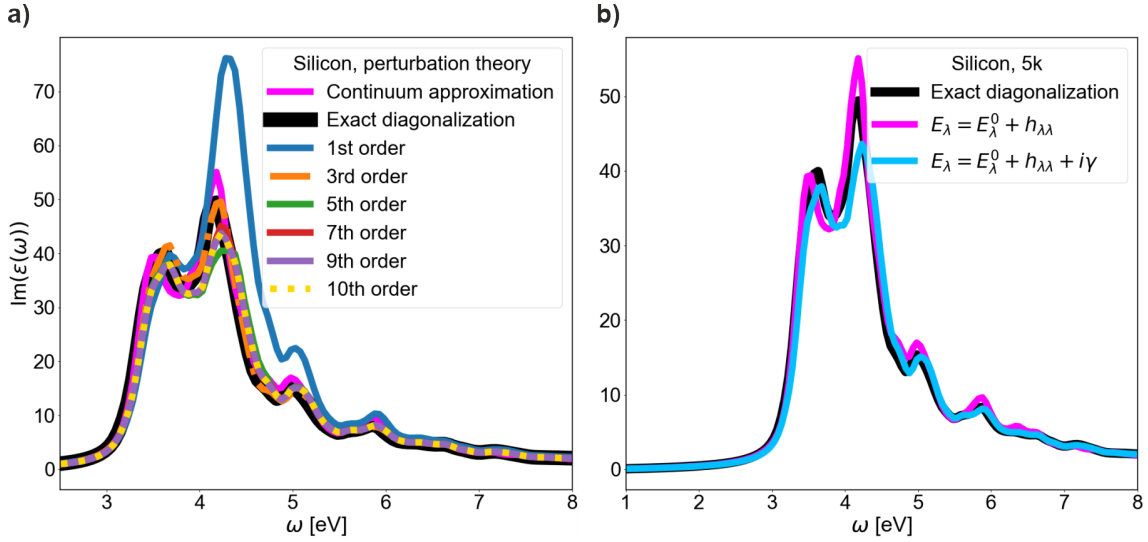
a)

b)

Figure 5.17: Absorption spectrum $(\mathrm{Im}(\epsilon(\omega))$ for the modified perturbation theory expansion of the exact eigenvector $\underline{\bar{A}}_\lambda$ (Eq. (5.83)), belonging to the Hamiltonian $\underline{\underline{H}}^{2p,reso}$ (Eq. (4.13)) for silicon. The expansion is performed for a broadening of $\gamma = 3 \cdot 10^{-3}$ Ha. Shown is **a)** the absorption spectrum for the series expansion in the eigenvector up to tenth order as well as the absorption spectrum for the unperturbated eigenvectors with the exact eigenenergies E_λ (black) and the continuum approximation for the eigenenergies $E_\lambda \approx E_\lambda^0 + h_{\lambda\lambda}$ (Eq. (5.65)) (magenta) and **b)** the unperturbated results for black, the exact diagonalization, magenta the continuum approximation and light blue the continuum approximation with a broadening of $\gamma = 3 \cdot 10^{-3}$ Ha. The result shows that after nine orders of the perturbation series, a useful and on the optical scale converged result is achieved for the spectrum. Numerical values are given in Appendix A; calculations were performed on a 5k grid (500 k-points) using a Kohn-Sham band structure, with 6 transition bands for the optical calculation and a Lorentzian broadening of $3.67 \cdot 10^{-3}$ Ha in the dielectric function.

In figure 5.17, we see the spectrum calculated in the continuum approximation with the eigenvectors from the modified perturbation series for $\gamma = 3 \cdot 10^{-3}$ Ha. The first observation we can make is that the spectrum is constructed in exactly the same way as in the case of the perturbation series of the absorption spectrum (Fig. 5.6): The first part of the eigenvector $\underline{A}_\lambda = \underline{1}_\lambda$ creates a shifted IP spectrum, the next order given by the diagonal matrix $\underline{\underline{\Delta \tilde{E}}}_\lambda^{-1}(\gamma)$ and the vector $\underline{\tilde{h}}_\lambda$ shifts the spectrum further and generally increases the spectrum (1st order, blue). The following orders of the series, which contain information from diagonals and especially off-diagonal elements of the Hamiltonian, then ensure a reduction in the effects so that a spectrum that is very close to the target spectrum is created after nine orders. The second observation we can make is that the spectrum no longer corresponds exactly to the one obtained by diagonalization. The reason for this is that through the introduction of a non-vanishing broadening parameter (γ) and using the continuum approximation, we change information in the eigenvectors, which means that the spectrum no longer corresponds exactly to the target spectrum but is an approximation to it. We can see the information change in the spectrum in figure 5.17 b), the continuum approximation without broadening (magenta) overestimates the second peak of the exact diagonalization. In contrast, we

underestimate the second peak with the introduced broadening and the continuum approximation (light blue). Thus, it is not the perturbation series that causes the deviation between the exact and the expanded spectrum but the broadening factor that modifies the eigenvectors and, thus, the spectrum.

We can conclude, according to the observation we made, that the modified perturbation series can give practical, usable approximations to the eigenvectors of continuum excitons for materials with weak to moderate excitonic effects, even though the obtained structure of the approximated eigenvectors shows a more localized and pronounced peak structure than the exact one.

Now that we have seen that the modified perturbation series works reasonably well in the continuum, we can expect that the modified perturbation series will work much better for materials with strongly bound excitons since, in this case, the exact and zero-order energies are probably separated, and thus, the divergence is weakened. However, this is precisely why we are confronted with another problem: We do not know the eigenenergies since we cannot apply the continuum approximation to them. This raises the question: *Can we use the modified perturbation series to determine the eigenvalues and eigenvectors of strongly bound excitons?*

5.6.2 Modified perturbation theory approach for excitonic states in Lithium fluoride

Unlike continuum excitons, strongly bound excitons have the advantage that their associated eigenenergies are separated from the continuum of the eigenenergy spectrum. For this reason, the strongly bound excitons could represent a much better field for approximations than the continuum excitons since divergence from $\underline{\underline{\Delta \tilde{E}}}_\lambda^{-1}$ will have a much milder effect. However, unlike the continuum excitons, we do not know a good analytical approximation for the eigenenergies, so we are faced with the problem of approximating both eigenvalues and eigenvectors. However, as we will show in the following, this is not an unsolvable big problem.

In the following, our starting equation will be that of the modified Brillouin-Wigner perturbation theory (Eq. (5.82)):

$$A_\lambda^\mu = \delta_{\mu\lambda} + \frac{(1 - \delta_{\mu\lambda})}{(e_\lambda [E_\lambda^0 + h_{\lambda\lambda}] + i\gamma) - (E_\mu^0 + h_{\mu\mu})} \tilde{h}_{\mu\lambda}$$

$$+ \frac{(1 - \delta_{\mu\lambda})}{(e_\lambda [E_\lambda^0 + h_{\lambda\lambda}] + i\gamma) - (E_\mu^0 + h_{\mu\mu})} \sum_{\mu'} \frac{(1 - \delta_{\mu'\lambda})}{(e_\lambda [E_\lambda^0 + h_{\lambda\lambda}] + i\gamma) - (E_{\mu'}^0 + h_{\mu'\mu'})} \tilde{h}_{\mu\mu'} \tilde{h}_{\mu'\lambda} + \ldots,$$

$$(5.84)$$

where we have expressed the eigenvalue by our representation $E_\lambda = e_\lambda [E_\lambda^0 + h_{\lambda\lambda}]$ (Eq. (5.40)). With this representation for the eigenenergies, we at least know that in the case of strongly bound excitons $e_\lambda < 1$ will apply, which gives us an upper limit below which we have to search for the eigenvalues. Before we want to deal with how to obtain the eigenvalues, we first want to check whether it is possible at all that we can approximate the eigenvectors because, from the point of view of the perturbation series, it is not very intuitive that for a strong perturbation $\underline{\underline{\tilde{h}}}$ an approach using a perturbation series should work. However, it is the interplay of perturbation and divergence that determines the spectral radius $S_r \left(\underline{\underline{\tilde{h}}} \, \underline{\underline{\Delta \tilde{E}}}_\lambda^{-1} (\gamma) \right)$, so there is hope that a perturbation-theoretic approach can work.

To ensure certainty, we must, however, examine the spectral radius more closely. To do this, we want to look at the maximum value e_λ that could be achieved such that the spectral radius is below one ($\max e_\lambda(S_r < 1)$) for LiF on a 3k grid. Since we are here only interested in the eigenvalues for the strongly bound excitons, we will have a look at the first three eigenvalues $\lambda = 1, 2, 3$ with $\gamma = 10^{-6}$ Ha.

λ	e_λ belonging to E_λ	$\max e_\lambda(S_r < 1)$	Δe_λ in %
1	0.86679	0.86677	0.0023
2	0.86472	0.86405	0.0775
3	0.86007	0.85884	0.1430

Table 5.1: Ratio (Δe_λ) between the exact e_λ associated with the first three eigenenergies of lithium fluoride and the maximum e_λ for which the spectral radius within the perturbation series of the associated eigenvector (Eq. (5.84)) remains less than one ($\max e_\lambda(S_r < 1)$).

From the table, we can read that although we chose the broadening to be extremely small ($\gamma = 10^{-6}$ Ha), the matrix structure factors associated with the first three eigenvalues, which would provide a spectral radius smaller than one ($\max e_\lambda(S_r < 1)$), differ from the exact matrix structure factors e_λ by significantly less than one per cent. This, in turn, shows us that a modified perturbation series (Eq. (5.84)) is possible even in the limit of small γ for the strongly bound excitons of LiF. Even knowing that the series would converge, we still do not know the eigenvalues. For them, we have to dig a little into the mathematical bag of tricks. Since math can sometimes be dry, we will first briefly explain the idea: Our problem is that we do not know the exact value for e_λ to any of the strongly bound excitons. Therefore, we have to estimate them somehow. Since for an estimated $\tilde{e}_\lambda$ we only have the vectors $|\tilde{A}_\lambda(\tilde{e}_\lambda)\rangle$ and energies $\tilde{E}_\lambda(\tilde{e}_\lambda) = \langle\tilde{A}_\lambda(\tilde{e}_\lambda)|\hat{H}|\tilde{A}_\lambda(\tilde{e}_\lambda)\rangle$, the question arises: *What is the difference between any energy $\tilde{E}_\lambda(\tilde{e}_\lambda)$, which is not an eigenenergy, and the energy of an eigenvalue?*

Since it is not trivial to make a general statement, nor to give a general proof for all eigenvalues, we will limit ourselves in this work to the lowest eigenvalue E_1, for which we can provide a proof. For all other eigenvalues, we will consider the result as an empirical fact but in no way claim that the method proposed below is valid in any case except for the lowest eigenvalue.

The statement is as follows: For all energies $\tilde{E}_1(\tilde{e}_1) = \langle\tilde{A}_1(\tilde{e}_1)|\hat{H}|\tilde{A}_1(\tilde{e}_1)\rangle \in M(\tilde{E}_\lambda(\tilde{e}_\lambda))$ calculated with the hermitian Hamiltonian $\hat{H}^{2p,reso}$ regardless of whether eigenenergy or not, the lowest eigenvalue E_1 is the global minimum $E_1 = \min\left(M(\tilde{E}_\lambda(\tilde{e}_\lambda))\right)$.

Proof:

Suppose we want to calculate the energy $\tilde{E}_\lambda$ for an arbitrary vector $\underline{\tilde{A}}_\lambda := \underline{\tilde{A}}_\lambda(\tilde{e}_\lambda)$ that lies in the Hilbert space $\mathscr{H}$ of the solutions of the positive definite, hermitian Hamiltonian H, which is spanned by the orthonormal eigenvectors $\underline{A}_\lambda$ of the Hamiltonian. In such a case, the energy is

given as:

$$\tilde{E}_\lambda(\tilde{e}_\lambda) = \langle \tilde{A}_\lambda(\tilde{e}_\lambda)|\hat{H}|\tilde{A}_\lambda(\tilde{e}_\lambda)\rangle = \sum_{\lambda'\lambda''} \langle \tilde{A}_\lambda(\tilde{e}_\lambda)|A_{\lambda'}\rangle \langle A_{\lambda'}|\hat{H}|A_{\lambda''}\rangle \langle A_{\lambda''}|\tilde{A}_\lambda(\tilde{e}_\lambda)\rangle = \sum_{\lambda'} |c_{\lambda\lambda'}|^2 E_{\lambda'},$$
(5.85)

where $c_{\lambda\lambda'} := \langle \tilde{A}_\lambda(\tilde{e}_\lambda)|A_{\lambda'}\rangle \in \mathbb{C}$ represents the overlap between a general vector ($|\tilde{A}_\lambda(\tilde{e}_\lambda)\rangle$) and an eigenvector ($|A_\lambda\rangle$).

Furthermore, since we are working with normalized vectors, the following applies:

$$\sum_{\lambda'} |c_{\lambda\lambda'}(\tilde{e}_\lambda)|^2 = \sum_{\lambda'} \langle \tilde{A}_\lambda(\tilde{e}_\lambda)|A_{\lambda'}\rangle \langle A_{\lambda'}|\tilde{A}_\lambda(\tilde{e}_\lambda)\rangle = \langle \tilde{A}_\lambda(\tilde{e}_\lambda)|\tilde{A}_\lambda(\tilde{e}_\lambda)\rangle = \delta_{\lambda\lambda}. \qquad (5.86)$$

Since we are interested in the eigenvalue E_1, which is by definition the smallest eigenvalue and the eigenvalues are real-valued numbers, we can construct any eigenvalue starting from E_1 by adding to E_1 a λ dependent positive value $\eta_\lambda \in \mathbb{R}_{>0}$: $E_\lambda = E_1 + \eta_\lambda$. If we insert this into equation (5.85), we get:

$$\tilde{E}_1(\tilde{e}_1) = \sum_{\lambda'} |c_{1\lambda'}|^2 E_{\lambda'} = \sum_{\lambda'} |c_{1\lambda'}|^2 (E_1 + \eta_{\lambda'}) = E_1 \sum_{\lambda'} |c_{1\lambda'}|^2 + \sum_{\lambda'} |c_{1\lambda'}|^2 \eta_{\lambda'} = E_1 + \sum_{\lambda'} |c_{1\lambda'}|^2 \eta_{\lambda'}.$$
(5.87)

However, since both $\eta_{\lambda'}$ and $|c_{\lambda\lambda'}|^2$ are positive real numbers, the following applies:

$$\tilde{E}_1(\tilde{e}_1) = E_1 + \sum_{\lambda'} |c_{1\lambda'}|^2 \eta_{\lambda'} \geqslant E_1$$

and thus E_1 is the global minimum for all energies $\tilde{E}_1(\tilde{e}_1) \in M(\tilde{E}_\lambda(\tilde{e}_\lambda))$ and not just for the eigenvalues E_λ.

Just before we want to apply this knowledge, it is worth pausing for a moment to consider whether the procedure is already familiar to us. We have an unknown parameter $\tilde{e}_\lambda$ which we vary in the vectors $\underline{\tilde{A}}_\lambda(\tilde{e}_\lambda)$ until the expectation value of the Hamiltonian gives us the global minimum ($\delta_{\tilde{e}_\lambda} \tilde{E}(\tilde{e}_\lambda) = 0$). This is exactly the procedure that is also used in the vartional method of quantum mechanics. However, in the general case even less is known than in our case, namely no general structure of the eigenvectors. This is why one generally uses so-called trial functions, which depend on the variation parameter(s) and serve as an approximation of the eigenfunction. As we will briefly explain later, the relationship between our method and the variations principle can provide an important indication of how we can extend this method to higher energies than the lowest.

With this knowledge, we can now write an algorithm in which we perform a series expansion in the modified perturbation series using different values for $\tilde{e}_1$ and then calculate the energies $\tilde{E}_1(\tilde{e}_1)$. The minimum of these energies then corresponds to our eigenenergy $E_1 = \min\left(M(\tilde{E}_\lambda(\tilde{e}_\lambda))\right)$.

Before we look at this procedure in practice, we should examine another case, namely what happens if we choose a $\tilde{e}_1$ for which the spectral radius becomes greater than one and thus a series expansion diverges. As established in section 5.3.1, the divergence of the eigenvector—specifically, a divergence of the series expansion—results in nothing else than a normalized vector with a pronounced peak ($|\tilde{A}_\lambda\rangle \approx |1_{\tilde{\lambda}}\rangle$). Consequently, although the series expansion collapses, the calculated value for the energy can, at most, assume the value of a continuum energy. This is due to the following relationship: $\tilde{E}_\lambda = \langle 1_{\tilde{\lambda}}|\hat{H}|1_{\tilde{\lambda}}\rangle = E_{\tilde{\lambda}}^0 + h_{\tilde{\lambda}\tilde{\lambda}}$. For this reason, as soon as the series diverges, depending on the strength of the divergence, the energy approaches a value in the continuum approximation

which is greater than the value of the energy for the bound exciton $E_1 < E_{\tilde{\lambda}}^0 + h_{\tilde{\lambda}\tilde{\lambda}} \forall_{\tilde{\lambda}}$. For this reason, even if we cannot achieve the exact value $\tilde{e}_1$ associated with an eigenvalue, we obtain an approximate value for the eigenenergy, which is still the minimum as long as this approximated value is smaller than $E_{\tilde{\lambda}}^0 + h_{\tilde{\lambda}\tilde{\lambda}}$.

Before we turn to the obtained results, it should be mentioned again that the results are only, as far as we know, valid for the smallest eigenvalue $\lambda = 1$; for all other eigenvalues, the results can only be used as empirical finding at the time of this work.
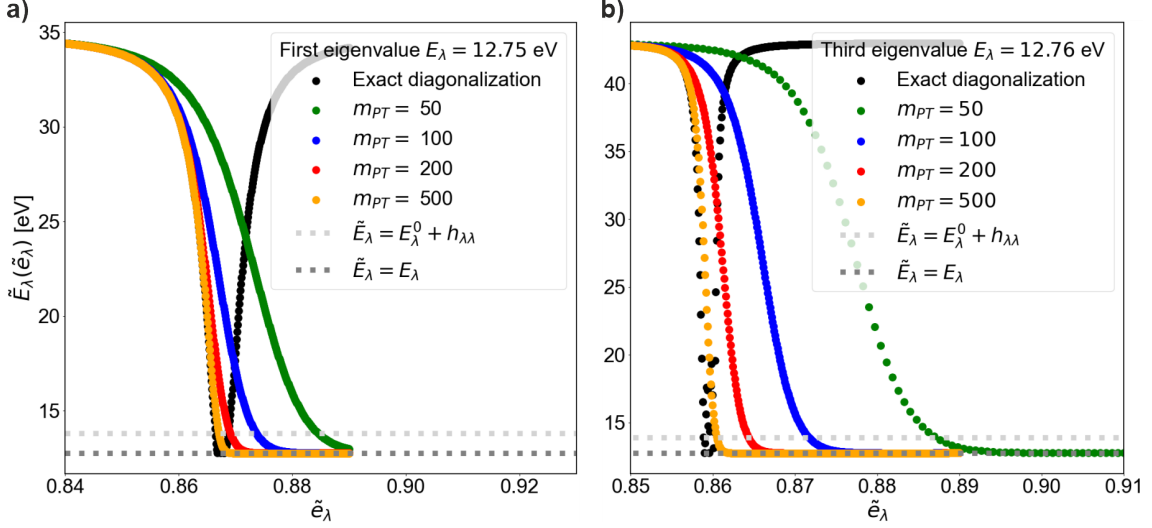
Figure 5.18: Calculation of the energies $\tilde{E}_\lambda(\tilde{e}_\lambda) = \langle \tilde{A}_\lambda(\tilde{e}_\lambda)|\hat{H}^{2p,reso}|\tilde{A}_\lambda(\tilde{e}_\lambda)\rangle$ (Eq. (5.85)) depending on the effective matrix structure factor $\tilde{e}_\lambda$ in the modified perturbation theory (Eq. (5.84), $\gamma = 10^{-6}$ Ha). Exact result for lithium fluoride in black, as well as the modified perturbation series expansion (Eq. (5.84)) depending on the order 50th (green), 100th (blue), 200th (red), 500th (orange) for **a)** the first eigenvalue ($\lambda = 1$) and **b)** the third eigenvalue ($\lambda = 3$). Shown as a horizontal dark grey line the exact value of the eigenenergy E_λ and in light grey the continuum approximation $E_\lambda = E_\lambda^0 + h_{\lambda\lambda}$ (Eq. (5.65)). It can be seen that, with increasing order, the energies obtained from the modified perturbation series converge close to the exact eigenvalue. Numerical values are given in Appendix A; calculations were performed on a 3k grid (108 k-points) using a Kohn-Sham band structure, with 6 transition bands for the optical calculation.

For both eigenvalues (Fig. 5.18 a) $\lambda = 1$, b) $\lambda = 3$), the energies approach the minimum, and as soon as they pass into the divergent region, they approach a "saddle point". The reason for this behavior is the beginning divergence of the series, which is why it slowly starts to converge towards the value $\tilde{E}_\lambda(\tilde{e}_\lambda) \rightarrow E_{\tilde{\lambda}} + h_{\tilde{\lambda}\tilde{\lambda}}$. Nevertheless, depending on the order one chooses for the series expansion (Eq. (5.84)), an accuracy up to 0.07% (5 meV) for the first eigenvalue can be achieved with 500 iteration steps.

λ	m_{PT}	E_λ [eV]	$\min \tilde{E}_\lambda$ [eV]	ΔE_λ in %	ΔE_λ [meV]
1	5	12.7515	13.5315	10.23	780
1	10	12.7515	13.1135	5.02	362
1	30	12.7515	12.8045	0.77	53
1	50	12.7515	12.7745	0.34	23
1	100	12.7515	12.7615	0.15	10
1	200	12.7515	12.7613	0.11	8
1	500	12.7515	12.7610	0.07	5

Table 5.2: Lowest excitonic energy of LiF determined using the variational principle and perturbation theory.

λ	m_{PT}	$\min E_\lambda$ [eV]	$\min \tilde{E}_\lambda$ [eV]	ΔE_λ in %	ΔE_λ [meV]
3	5	12.7606	13.6366	11.32	876
3	10	12.7606	13.2596	6.78	499
3	30	12.7606	12.8316	1.03	71
3	50	12.7606	12.7816	0.31	21
3	100	12.7606	12.7607	0.002	0.1
3	200	12.7606	12.7612	0.09	6
3	500	12.7606	12.7614	0.12	8

Table 5.3: Third excitonic energy of LiF determined using the variational principle and perturbation theory.

Before we turn to the deviation for the third eigenvalue and want to find out where the reasons could lie, we will first look at the corresponding eigenvectors.
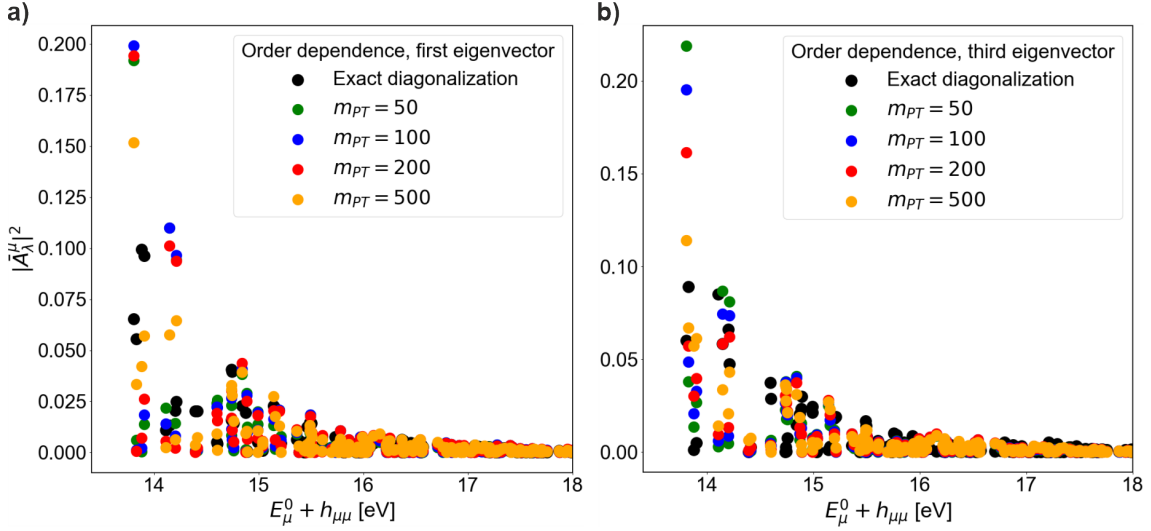
Figure 5.19: Modified perturbation theory expansion (Eq. (5.84), $\gamma = 10^{-6}$ Ha) of the normalized excitonic eigenvector $\underline{\bar{A}}_\lambda$ belonging to the Hamiltonian $\underline{\underline{H}}^{2p,reso}$ (Eq. (4.13)) for lithium fluoride. The expansion is performed in 50th (green), 100th (blue), 200th (red) and 500th (orange) order. Plotting the norm of the normalized eigenvector-components, $|\bar{A}_\lambda^\mu|^2$, belonging to the associated minimum energy $\min\left(M(\tilde{E}_\lambda(\tilde{e}_\lambda))\right)$ for **a)** the first eigenenergy ($\lambda = 1$) and **b)** the third eigenenergy ($\lambda = 3$), against the diagonal elements $E_\mu^0 + h_{\mu\mu}$ of the Hamiltonian. Additionally, the exact result is shown in black. It can be seen that the approximated eigenvectors follow the structure of the exact one relatively well. Numerical values are given in Appendix A; calculations were performed on a 3k grid (108 k-points) using a Kohn-Sham band structure and 6 transition bands for the optical calculation.

Like the eigenvalues, we see that the eigenvectors corresponding to the first and third eigenvalues approximate the exact result relatively well. The approximated eigenvectors reflect the broadened structure of the exact eigenvector very well, and the deviation between approximated and exact eigenvectors decreases with increasing order.

The results show that a perturbation-theoretic approximation of the bound eigenvectors is possible for negligible γ values in LiF. Furthermore, in the case of the first eigenvalue and eigenvector, we can be sure that the result will converge to the exact value if the spectral radius is less than one, according to our proof. However, in the case of the third eigenvector, we have to delve deeper into the variations principle, which tells us what we must fulfill in higher energies ($\lambda > 1$). According to the variations principle, the condition to be valid for higher states ($\lambda > 1$) is that the associated states are constrained by symmetry. This means the higher states are orthogonal to the other known states during the entire variation. It is obvious that in our case, where we start the variation from the exact structure of the eigenvectors, that if $\tilde{e}_\lambda = e_\lambda$ holds, the corresponding eigenvector $\underline{A}_\lambda(e_\lambda) = \underline{A}_\lambda$ is by definition orthogonal to any other eigenvector ($\langle A_\lambda | A_{\bar\lambda} \rangle = \delta_{\lambda\bar\lambda}$). At this point, however, we do not know whether the constructed eigenvectors maintain this behavior during the entire procedure or whether the orthogonality changes with the variation of $\tilde{e}_\lambda$ or during

the perturbation-theoretical expansion of the eigenvector. It is also possible that one of the two points is the reason why we observe the following behavior for the third eigenvalue: It first appears to converge ($m_{PT} = 5, 10, 30, \dots$) and then moves away from the minimum ($m_{PT} = 200, 500$). However, a more in-depth analysis is no longer part of this work, but our brief discussion should clarify that although the results seem very promising for the third eigenenergy and eigenvector, we have to treat them cautiously.

We have thus begun to tackle the last piece of the puzzle in this work, namely the strongly bound excitons, and have shown that under certain circumstances, it is possible to calculate the eigenvalues as well as the eigenvectors at the same time using a perturbation-theoretic description combined with the variations principle. The method clearly works very well for the lowest eigenvalue. For higher excitation energies, the variational principle will work if the corresponding states are constrained by symmetry. It remains to be seen whether this will also hold by definition for our constructed vectors during the entire procedure or whether one has to introduce orthogonality constraints. However, this is no longer part of this work.

Chapter 6

Conclusion and Outlook

"What we determine mathematically is only to a small extent an objective fact, to a greater extent an overview of possibilities."

Werner Heisenberg

6.1 Conclusion

In this work, our objective was to improve the understanding and computation of excitonic effects in the absorption spectra of materials. Absorption experiments measure a response function that can be formulated as an expectation value of an inverse operator describing the propagation of electrons and holes, as well as the electron-hole interaction. Our investigation started with the fundamental equations and commonly used approximations. In particular, we examined the Haydock-Lanczos algorithm, a standard tool for absorption spectra calculations (Ch. 5.1.1). Despite its numerical benefits, we identified limitations for detailed analyses due to its continued fraction series representation of the dielectric function. In order to better understand the role of excitonic effects for absorption spectra and to find powerful approximations, we, therefore, turned our attention to the Neumann series (Ch. 5.1.2). This allowed us to shed light on the role of the different contributions from the Hamiltonian for the formation of absorption spectra and to propose an approach with low numerical scaling. In addition to an explanation of how an absorption spectrum is systematically constructed from an independent-particle absorption spectrum to a spectrum with excitonic effects, we were able to show that this approach proved effective, especially for calculating the dielectric function at finite frequencies.

Expanding on this groundwork, we delved into the fundamental research topic of excitonic eigenvalues and eigenvectors (Ch. 5.2). Using the Gaussian algorithm, we derived three ways to display the exact eigenvector structures, which are crucial for analytical analysis. Furthermore, our investigation led to the development of an exact equation for eigenvalues, relying on the diagonal elements of the Hamiltonian and scaled by a scalar that we called the matrix structure factor. We found that the matrix structure factors converge extremely fast to the value one in the continuum

of excitations for materials with weak and strong excitonic effects. This observation and further analysis enabled an accurate approximation of the excitonic eigenvalues by the diagonal elements of the Hamiltonian, even for relatively sparse Brillouin sampling, making cost-effective calculations feasible for the continuum of various materials and bound excitons near the optical bandgap (Ch. 5.4).

Inspired by these findings, our understanding of the Neumann series and the different exciton types, we explored a perturbative approach for calculating eigenvectors in the continuum. In addition to adapting traditional perturbation theory (Ch. 5.5), we introduced modifications in the perturbation theory series expansion of the eigenvectors to mitigate divergence, which arises from the non-well-separated eigenenergies in the continuum. This approach allowed us to determine approximate eigenvectors for silicon (Ch. 5.6.1). Noteworthy, based on the insight gained in our exploration, we also proposed an original method to determine the energies of the lowest bound excitons of lithium fluoride with millielectronvolt accuracy (Ch. 5.6.2), using perturbation theory. Thus, the modified perturbation series gives access to the corresponding eigenvectors and absorption spectra with an $O(N^2)$ numerical efficiency, where N is the number of independent-particle transitions used as basis functions for the electron-hole problem.

In summary, our work introduces alternative approaches for calculating spectroscopic quantities, offering numerical advantages and a deeper understanding of excitonic effects. Furthermore, we have derived exact equations for eigenvectors and eigenvalues beyond those of the eigenvalue equation, which may allow one to find new and cost-effective approaches for numerical calculations. However, we recognize that this study is merely one step forward, acknowledging Isaac Newton's wisdom that our knowledge is a drop compared to the vast ocean of the unknown. In the outlook, we discuss potential paths for future exploration.

6.2   Outlook

6.2.1   The understanding of bound excitons and the thermodynamic limit

In this work, we have derived a number of equations that have allowed and will continue to allow a deep insight into excitonic effects. One question that will be very important for further approximation is how the thermodynamic limit affects the eigenvalues and eigenvectors. In particular, this requires a deeper understanding of the differences between bound and continuum excitons and, to be even more precise, between strongly bound excitons and the continuum. At the time of this work, we did not yet know how the weakly bound excitons behave, i.e., those excitons that are bound but whose energy is very close to the optical band edge.

One approach for the further study of these excitons could be a simple two or four-band Wannier-Mott model [97,98] with which we can perform various analyses. For example, how the non-constant part of the matrix structure factor behaves depends on the number of k-points and the strength of the excitonic effects. Furthermore, this model could be useful for investigating weakly bound excitons and strongly bound excitons.

In addition to this already somewhat more complex model, further simple models would be welcome with which we could perform analytical calculations and thus understand the thermodynamic limit. One such model could be a constant interaction matrix that scales with C/N_k.

6.2.2   Haydock-Lanczos for the eigenvector

Since we are interested in the eigenvalues and the eigenvectors, one of the most critical next steps is to use a more efficient algorithm than the Neumann algorithm for approximating the inversion of the matrix. In particular, an algorithm that is less dependent on broadening since this dependency leads to limitations.

We already know one possible algorithm, the Haydock-Lanczos algorithm (Ch. 5.1). Since the algorithm represents a continued fraction series, it is not limited by divergence and, therefore, not by a particular broadening factor, as is the Neumann algorithm. In the case of the Haydock-Lanczos algorithm, increasing broadening follows in more iteration steps necessary to achieve convergence. So, instead of using the Neumann algorithm to approximate the matrix inversion, we can also use the Haydock-Lanczos algorithm:

$$((E_\lambda + i\gamma) - \hat{H})^{-1} |\tilde{h}_\lambda\rangle = \sum_n c_n(E_\lambda) |f_n\rangle \approx \sum_n^{m_{Hay}} c_n(E_\lambda) |f_n\rangle, \ |f_0\rangle = |\tilde{h}_\lambda\rangle / ||\tilde{h}_\lambda||.$$

This equation can now have several advantages, firstly that it converges in any case, since the coefficients are given by $c_n(E_\lambda) = \phi_n(E_\lambda) b_n c_{n-1}(E_\lambda)$ and $\phi_n(E_\lambda) = [(E_\lambda + i\gamma) - a_n - b_{n+1}^2 \phi_{n+1}(E_\lambda)]^{-1}$, representing a contiued fraction. Secondly, it may not be necessary to start for each eigenvector again with the choice for the starting vector $|f_0\rangle$. It is maybe enough to calculate all components a_n and b_n once for one eigenvector, corresponding to the index $\tilde{\lambda}$, and reuse the values for eigenvectors in an energy range $E_\lambda \in [E_{\tilde{\lambda}} - \Delta, E_{\tilde{\lambda}} + \Delta]$. This can be justified if the structure and symmetry characteristics of the eigenvector prime is determined by the matrix inversion $\underline{\underline{\Delta E}}_\lambda^{-1}(\gamma)$ and not

by the vector $\underline{\tilde{h}}_\lambda$ multiplied by it. An illustration of this statement can be found in Appendix D (Fig. 11).

If it could be proven that it is always, or at least for many eigenvectors, possible to start from the same starting vector, i.e., $|f_0\rangle = |\tilde{h}_{\tilde{\lambda}}\rangle/||\tilde{h}_{\tilde{\lambda}}||$, where $\tilde{\lambda}$ is fixed for all or a group of λs, then the numerical costs for all or a group of eigenvectors would be reduced to $O(N^2)$, since $c_n(E_\lambda)$ only depends on a_n and b_n, as shown in section 5.1.1 and therefore has to be constructed once.

Such an approach could eventually lead to a very accurate and efficient algorithm for the eigenvectors in the continuum and for weakly bound excitons, which we could only approximate in this work using the Neumann series.

6.2.3   Machine learning and neuronal networks

An important and hot topic that is also suitable for our purposes is machine learning and neuronal networks. This technique has become indispensable in physics, especially when large matrices are involved, like in the case of screening calculations [104], or information is so nested that a simple physical approximation is no longer possible, like for the functionals in DFT [125, 126]. It is also used to calculate eigenvalues and eigenvectors [127], as well as to invert matrices [128].

However, a different approach would be helpful for our application because the matrices to be inverted for the eigenvectors can take on large dimensions and are, therefore, difficult to handle, even in the case of machine learning.

For our purposes, we could use the properties of the beta-matrix approach (Eq. (5.32)). The remarkable thing about the beta-matrix approach is that instead of inverting a complex $N \times N$ matrix for each λ, only the inversion of a diagonal, real-valued matrix, which has only been empirically proven at this point, can be carried out. Therefore, the costs of the beta-matrix approach arise from the construction of the matrix and not from the inversion per se.

For this reason, the beta-matrix approach is the ideal starting point for neural networks because a matrix is nothing more than an image, and we can, therefore, use the complex mapping between input and target spaces that neural networks provide to determine beta matrices.

For example, it may make sense to feed a neuronal network with $\underline{\underline{\Delta \tilde{E}}}_\lambda^{-1}$ as predictors and obtain the associated beta matrices. A similar procedure to calculate molecule spectra was described in [129].

Another way in which machine learning can help us is to determine eigenenergies. Since, with equation (5.40), $E_\lambda = e_\lambda \left[E_\lambda^0 + h_{\lambda\lambda} \right]$, we have an exact equation for the eigenvalues, which depends on two parameters, firstly the diagonal elements of the Hamiltonian and secondly on a scalar e_λ, which has to be determined for each λ. We know that e_λ converges to one in the continuum. However, for strongly bound excitons, e_λ is different from one because the matrix structure requires off-diagonal elements to be considered. Since the structure and strength of the Hamiltonian determines the eigenvalues, it can make sense to use machine learning to obtain the factors e_λ directly with the predictor Hamiltonian for the first view eigenvalues, similar to what was done in [130] for bandgaps.

6.2.4   Exciton-exciton-transition

In the following, we will look at possible applications of the developed approach rather than how it can be further improved.

An important application topic is exciton-exciton transitions since it is currently blossoming because ultrafast optics and improved measurement setups also make it possible to carry out excitations between excitonic states [4, 5, 131]. From a theoretical point of view, however, this field is rather complex to enter because it generally opens the field of non-equilibrium physics, which requires an exact dynamic description and is, therefore, very costly in numerical terms [17].

The numerical costs are particularly problematic as the scientific community has focused on materials with a Rydberg series of strongly bound excitons. The problem, from an ab inito point of view, is that such materials require an extremely dense k-lattice ($N_k > 10^5$) to achieve convergence [20, 24, 25]. This, in turn, leads to the fact that the associated matrix becomes large ($N \sim N_c N_v N_k = N_c N_v \cdot 10^5$ [20]) and thus, a calculation of the eigenvectors and eigenvalues is costly or impossible. The dynamic calculation is even more problematic.

At least for the dilemma of dynamic effects, there is a way out. If quasi-static excitonic states are assumed, it is possible to write down an equation for the exciton-exciton transitions based on linear-response [17]

$$\chi_{\mu_\alpha \mu_\beta}^{\lambda' q}(\omega) = \frac{2}{V} \sum_\lambda \frac{\mu_{\lambda' \lambda}^\alpha(q) \mu_{\lambda \lambda'}^\beta(q)}{E_\lambda(q) - E_{\lambda'}(q) - \Omega},$$

which, however, requires the static eigenvectors $\underline{A}_\lambda$ and eigenenergies E_λ as input [17]

$$\mu_{\lambda \lambda'}(q) = \sum_{v,c,c',k} (A_{\lambda q}^{cvk})^* A_{\lambda' q}^{c'vk} r_{cc'k} - \sum_{c,vv',k} (A_{\lambda q}^{cvk})^* A_{\lambda' q}^{cv'k} r_{v'vk-q} + \sum_{vck} (A_{\lambda q}^{cvk})^* A_{\lambda' q}^{cvk} \sum_{n \in occ} r_{nnk},$$

with $r_{nmk} = \langle nk|\hat{r}|mk \rangle, n \neq m$ defines the interband dipol-matrix element and $r_{nnk} = \langle nk|\hat{r}|nk \rangle$ the intraband dipol-matrix element. Thus, only the dynamic transition problem, not the diagonalization problem, can be bypassed. For these purposes, our newly developed approach could be an ideal tool for this field of application because it is not the exact information needed but a summarized form of information. Thus, a sufficiently good approximation of the eigenvectors and eigenvalues can lead to very good results, especially with low numerical costs ($O(N^2)$).

6.2.5   Time-dependent Bethe-Salpeter equation and beyond Tamm-Dancoff approximation

As we have seen in this work (Ch. 4.5.4) but have neglected throughout, in the non-adiabatic GW approximation, the BSE can also be transferred to an effective two-particle Hamiltonian, which, however, is frequency dependent [17]. This means that for each frequency, there is a Hilbert space of the same size as the static one, and thus, the numerical scaling for solving the problem has increased by the multiplicative factor N_ω. Since our approaches for the eigenvector (Eq. (5.27), (5.32), (5.39)) can be applied to any diagonalizable matrix, we can also use it in the dynamic case and thus possibly reduce the numerical costs.

Furthermore, it will be a similar story for the complete matrix (Eq. (4.12)) with coupling terms, i.e., beyond the TDA. These components are very often neglected because they lead to a reduction

of the Hamiltonian dimension by a factor of 2. However, as we have already discussed in section 4.5.1, this approximation will lead to problems in the case of very strongly bound excitons and especially in the case of EELS. For this reason, this case could also become an important point of application for our approaches.

Chapter 7

Appendix

A. Convergence parameters and material information

This part of the appendix contains the convergence parameters of the materials for ground state calculations in KS-DFT (Kohn-Sham bandstructure) and BSE calculations, as well as the used pseudopotentials and the scissor operators.

Nomenclature

a_{cell}: length scales by which dimensionless primitive translations vectors are to be multiplied

E_{cut}: kinetic energy cutoff which controls the number of plane waves at a given k point

$E_{cut,eps}$: determines the cutoff energy of the plane wave set used to represent the independent-particle susceptibility

$E_{cut,wfn}$: determines the cutoff energy of the plane wave set used to represent the wavefunctions in the formula that generates the independent-particle susceptibility

matsh: number of shells of G-vectors for the ϵ-matrix

wfnsh: number of shells of plane waves to describe the wavefunctions

N_{bands}: number of the last band (highest energy) to be included in the calculation

lomo: number of the first band (lowest energy) to be included in the calculation

Convergence criteria

Ecut was converged to a total energy difference of smaller than 10^{-5} Ha.

All parameters responsible for absorption spectra were systematically converged to the visible absorption spectrum until no direct change in the spectrum was observable. Each parameter was tested twice and in combination with the other parameters. After convergence was achieved, all parameters were tested again.

For the k-grid, a shifted grid was used in all cases to achieve faster convergence; the chosen shift is (0.11, 0.21, 0.31).

Germanium

Pseudopotential: non-relativistic LDA [132]

Scissor: 0.7 eV [133]

Groundstate-calculation

a_{cell}[Bohr]	E_{cut} [Ha]	$E_{cut,eps}$ [Ha]	$E_{cut,wfn}$ [Ha]	N_{band}
10.656 [134]	40	7	21	50

EXC-Calculation

matsh	wfnsh	N_{bands}	lomo	Broadening [Ha]
14	32	18	12	$3.67 \cdot 10^{-3}$

Silicon

Pseudopotential: non-relativistic PBE [132]

Scissor: 0.68 eV (comparison between experiment [40] and calculation)

Groundstate-calculation

a_{cell}[Bohr]	E_{cut} [Ha]	$E_{cut,eps}$ [Ha]	$E_{cut,wfn}$ [Ha]	N_{band}
10.217 [134]	30	5	16	50

EXC-Calculation

matsh	wfnsh	N_{bands}	lomo	Broadening [Ha]
20	27	18	2	$3.67 \cdot 10^{-3}$

DP-Calculation (EELS)

matsh	wfnsh	N_{bands}	lomo	Broadening [Ha]
25	35	52	1	$3.67 \cdot 10^{-3}$

Zinc Sulphide

Pseudopotential: non-relativistic PBE [132]

Scissor: 0.97 eV [135] (comparison between $G0W0$ and groundstate calculation)

Groundstate-calculation

a_{cell}[Bohr]	E_{cut} [Ha]	$E_{cut,eps}$ [Ha]	$E_{cut,wfn}$ [Ha]	N_{band}
10.29 [136]	40	6	18	120

EXC-Calculation

matsh	wfnsh	N_{bands}	lomo	Broadening [Ha]
12	32	24	4	$3.67 \cdot 10^{-3}$

Lithium Fluoride

Pseudopotential: non-relativistic LDA [132]

Scissor: 5.9 eV (comparison between experiment [41] and calculation)

Groundstate-calculation

a_{cell}[Bohr]	E_{cut} [Ha]	$E_{cut,eps}$ [Ha]	$E_{cut,wfn}$ [Ha]	N_{band}
7.583 [134]	50	19	34	90

EXC-Calculation

matsh	wfnsh	N_{bands}	lomo	Broadening [Ha]
26	32	24	2	$3.67 \cdot 10^{-3}$

B. Convergence silicon and lithium fluoride

The convergence shown refers to the absorption spectrum of lithium fluoride and silicon, which are the main materials in this work. Since the spectrum hardly changes noticeably for nearly converged parameters, only zoom-in on the optical band edge for silicon and the strongly bound exciton for lithium fluoride are shown. The only exceptions are the lomo parameter and the k-grid.

The convergence was made as follows: E_{cut} of the ground state calculation was converged until a deviation smaller than 10^{-5} Ha can be guaranteed for the total energy. Then, the grid was set to 4k (256 k-points) so that a clear absorption spectrum was visible. Subsequently, the parameters of the screening $E_{cut,eps}$, $E_{cut,wfn}$ and N_{band} were converged until there was no noticeable change in the spectrum. The parameters were then tested alternately so that no further change occurred. Then the parameters, matsh, wfnsh, N_{bands} and lomo, were converged in EXC for the spectrum and checked mutually.

After all parameters had converged, the k-grid was converged for the spectrum until no noticeable change could be found.
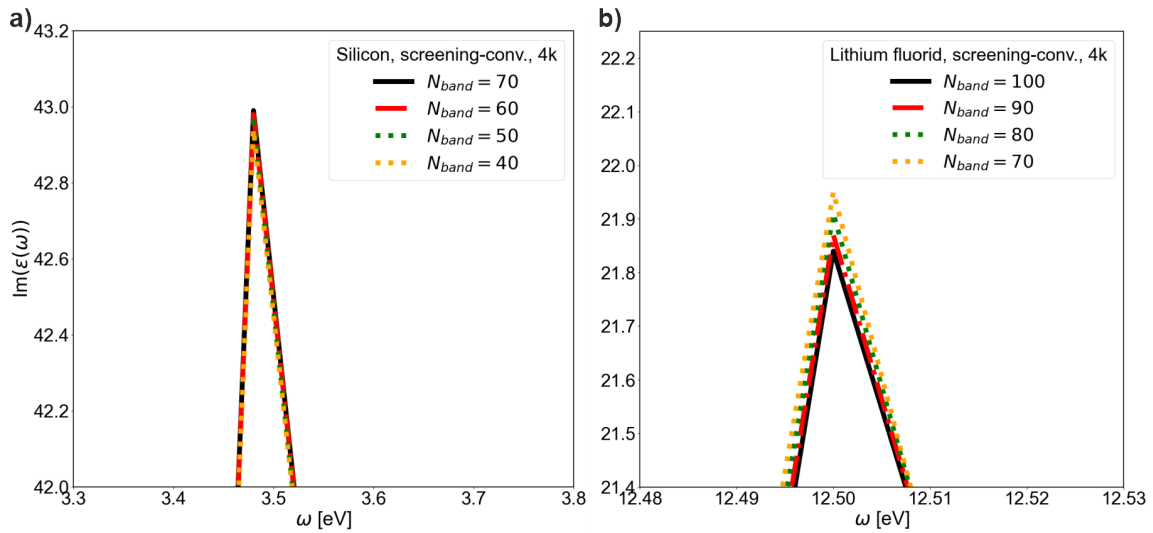
Figure 1: Convergence in terms of the number of bands in the screening calculation. Shown for **a)** silicon and **b)** lithium fluoride. Convergence is achieved for silicon at 50 bands and lithium fluoride at 90 bands. The calculations were carried out on a 4k grid (256 k-points) using a Kohn-Sham band structure.
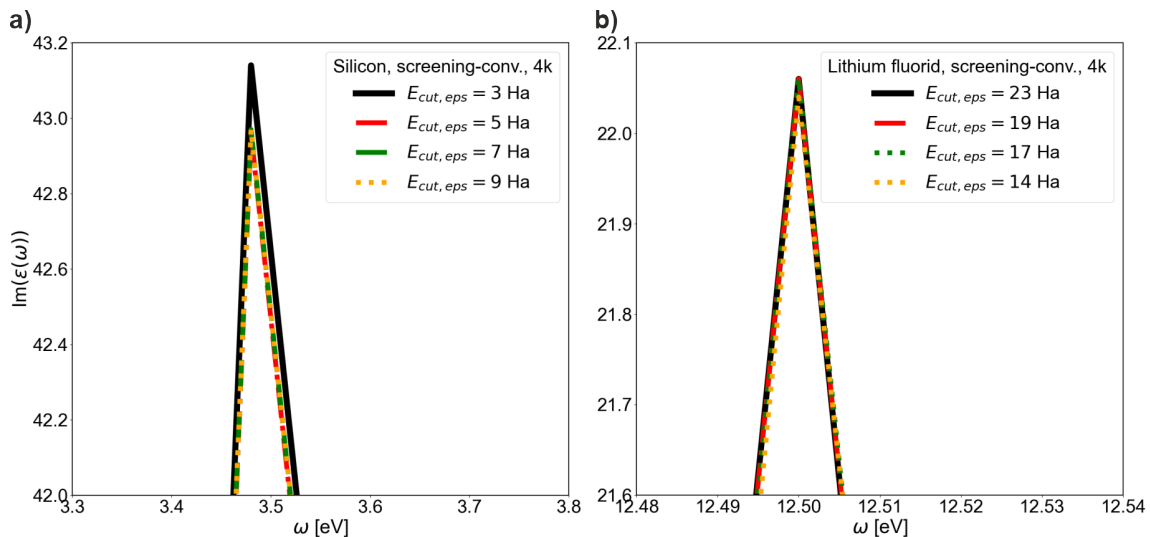
Figure 2: Convergence in terms of the cut-off energy of the planewave set used to represent the independent-particle susceptibility in the screening. Shown for **a)** silicon and **b)** lithium fluoride. Convergence is achieved for silicon at $E_{cut,eps} = 5$ Ha and lithium fluoride at $E_{cut,eps} = 19$ Ha. The calculations were carried out on a 4k grid (256 k-points) using a Kohn-Sham band structure.
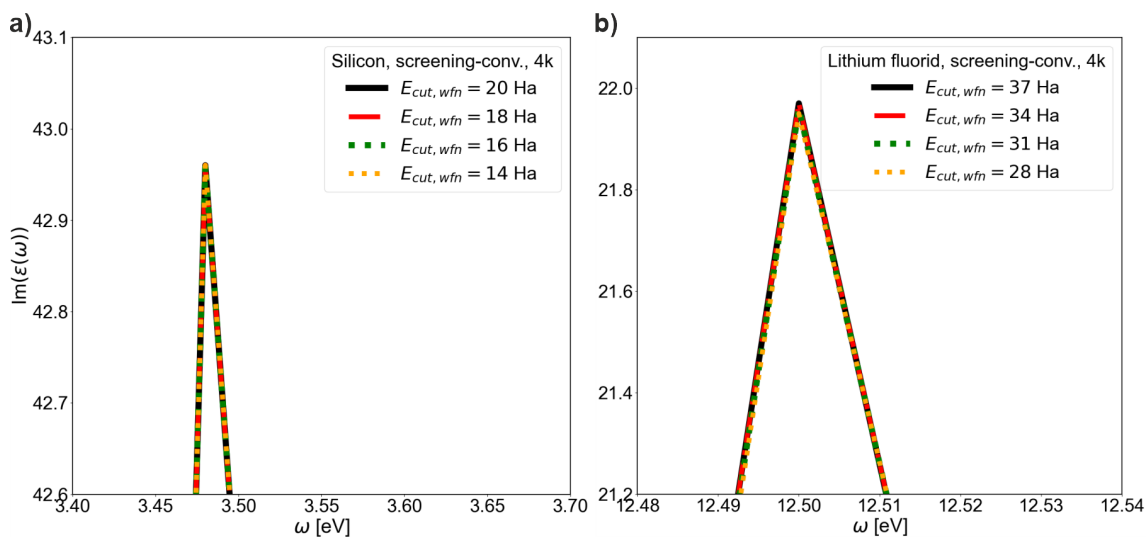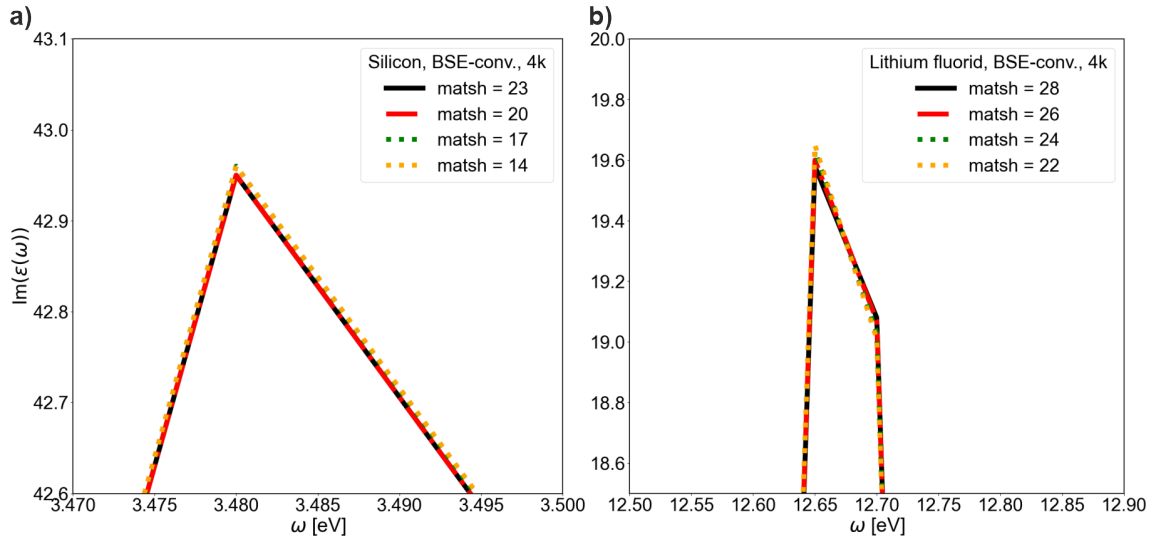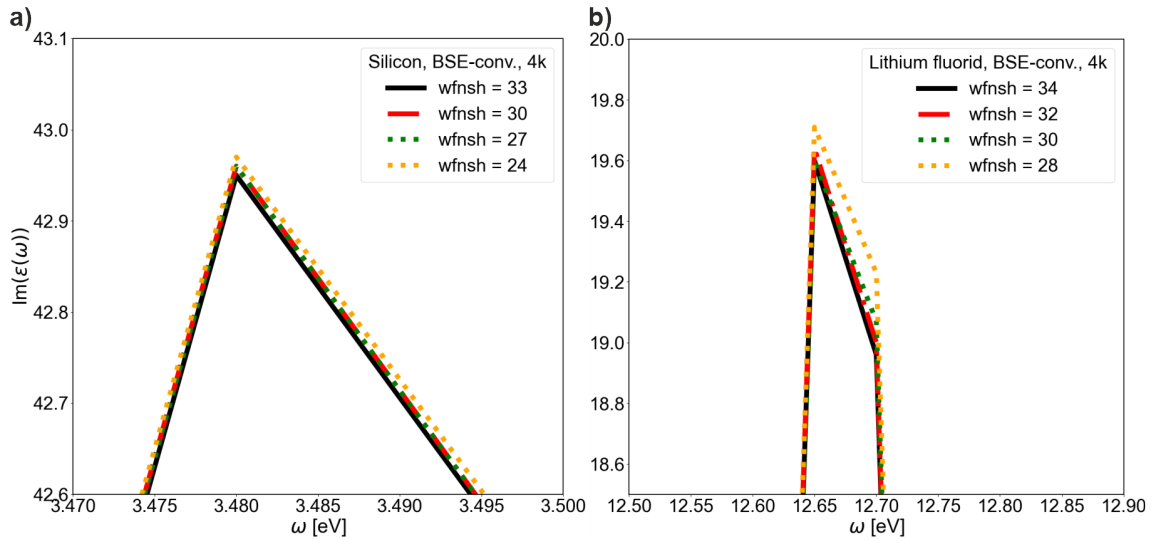

Figure 3: Convergence in terms of the cut-off energy of the planewave set used to represent the wavefunctions in the formula that generates the independent-particle susceptibility. Shown for **a)** silicon and **b)** lithium fluoride. Convergence is achieved for silicon at $E_{cut,wfn} = 16$ Ha and lithium fluoride at $E_{cut,wfn} = 34$ Ha. The calculations were carried out on a 4k grid (256 k-points) using a Kohn-Sham band structure.

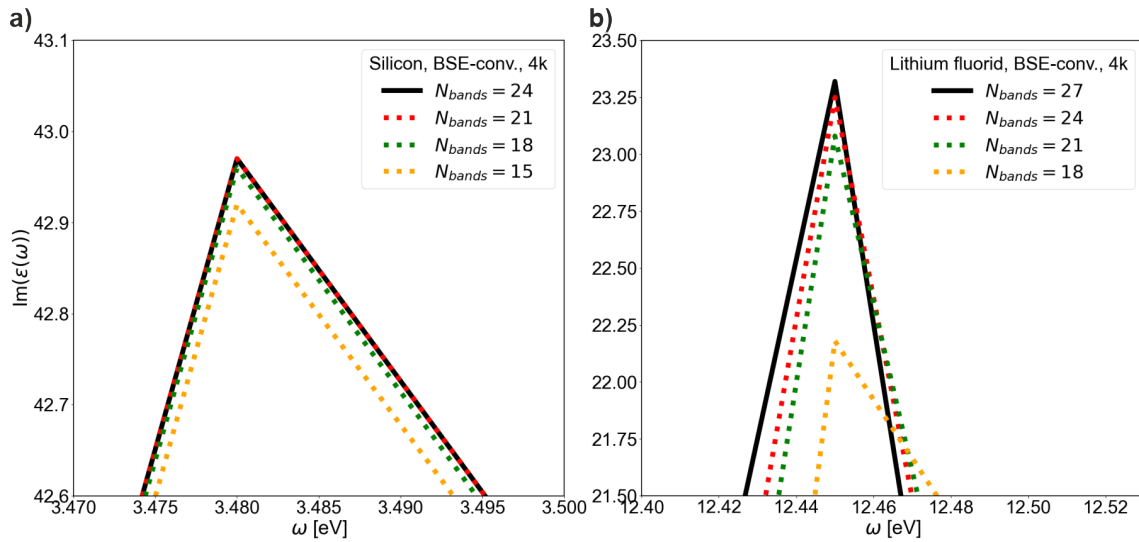**Figure 4:** Convergence in terms of the number of shells of G-vectors for the ϵ-matrix. Shown for **a)** silicon and **b)** lithium fluoride. Convergence is achieved for silicon at matsh = 20 and lithium fluoride at matsh = 26. The calculations were carried out on a 4k grid (256 k-points) using a Kohn-Sham band structure.

Figure 5: Convergence in terms of the number of shells of plane waves to describe the wave-functions. Shown for **a)** silicon and **b)** lithium fluoride. Convergence is achieved for silicon at wfnsh = 27 and lithium fluoride at wfnsh = 32. The calculations were carried out on a 4k grid (256 k-points) using a Kohn-Sham band structure.

Figure 6: Convergence in terms of the number of transition bands. Shown for **a)** silicon and **b)** lithium fluoride. Convergence is achieved for silicon at $N_{bands} = 18$ and lithium fluoride at $N_{bands} = 24$. The calculations were carried out on a 4k grid (256 k-points) using a Kohn-Sham band structure.
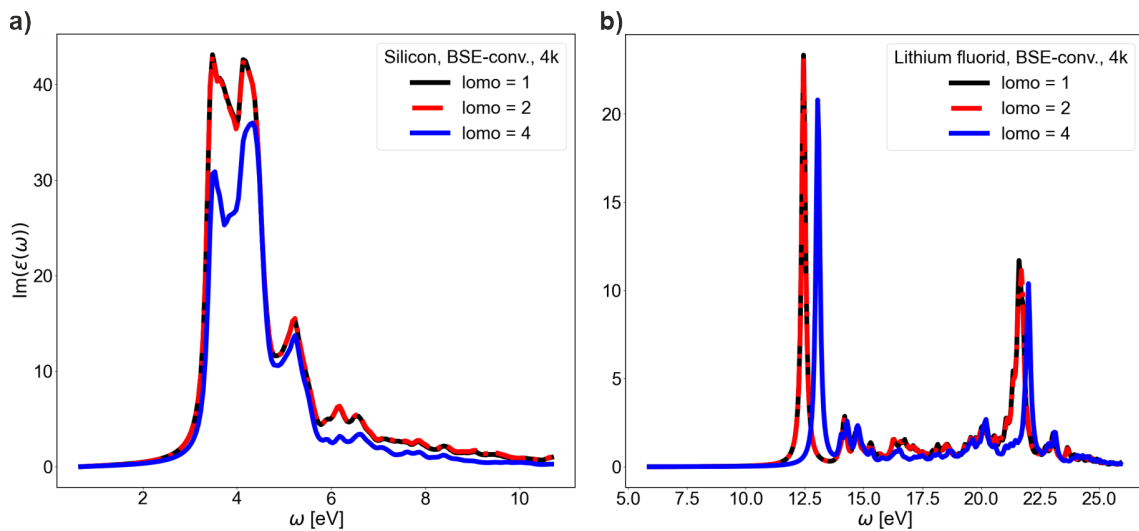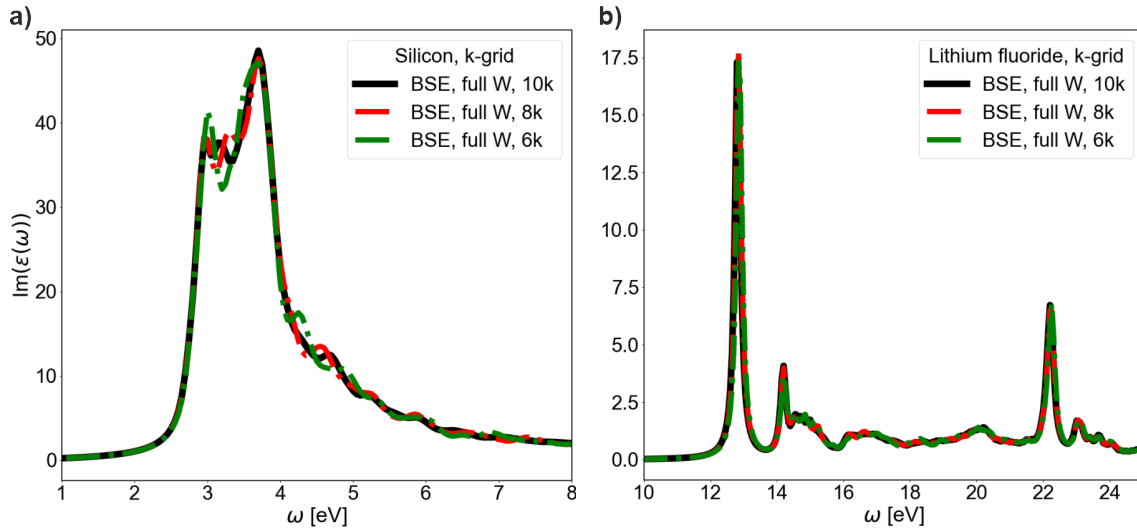

Figure 7: Convergence in terms of the number of the first band included in the calculation. Shown for **a)** silicon and **b)** lithium fluoride. Convergence is achieved for silicon at lomo = 2 and lithium fluoride at lomo = 2. The calculations were carried out on a 4k grid (256 k-points) using a Kohn-Sham band structure.

Figure 8: Convergence in terms of the k-grid density. Shown for **a)** silicon and **b)** lithium fluoride. Convergence is achieved for silicon at 8k and lithium fluoride at 8k using a Kohn-Sham band structure. The calculations were done with full converged parameters Appendix A for silicon and with 6 transition bands for lithium fluoride.

C. Low band calculation silicon and lithium fluoride

The following shows the difference between the fully converged spectrum for silicon and lithium fluoride with 22 transition bands (Nbands-lomo) and 6 transition bands calculated on an 8k grid (2048 k-points) using a Kohn-Sham band structure.
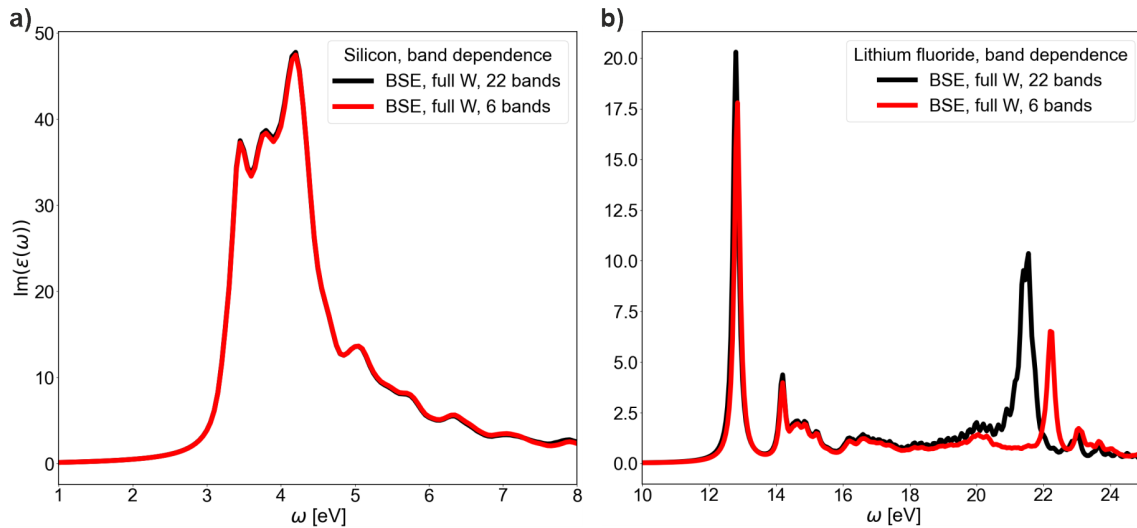

Figure 9: Comparison of the absorption spectrum $Im(\epsilon(\omega))$ for **a)** silicon and **b)** lithium fluoride. Shown 22 transition bands (Nbands-lomo) (black) and 6 transition bands (red). The results show that a very accurate absorption spectrum (red) can be generated for silicon by neglecting the bands. In the case of lithium fluoride, it is less accurate but sufficiently good. Calculations were performed with full converged parameters on an 8k grid (2048 k-points) using a Kohn-Sham band structure.

D. Unshown results

This part of the appendix shows results which are not directly relevant to this work but which emphasize the results in an appropriate place.
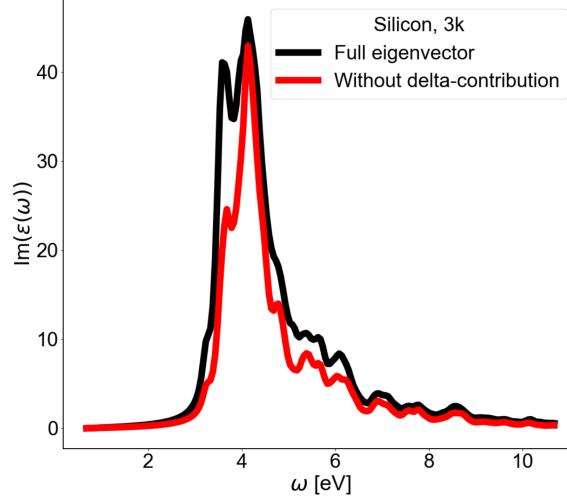

Figure 10: Comparison of the absorption spectrum $\mathrm{Im}(\epsilon(\omega))$ for the full eigenvector structure $\underline{A}_\lambda = \underline{1}_\lambda + \underline{\underline{\Delta E}}_\lambda^{-1}(\gamma)\left(\underline{\underline{1}} - \underline{H\Delta E}_\lambda^{-1}(\gamma)\right)^{-1}\tilde{\underline{h}}_\lambda$ (black) and the eigenvector without the delta-contribution $\underline{A}_\lambda = \underline{\underline{\Delta E}}_\lambda^{-1}(\gamma)\left(\underline{\underline{1}} - \underline{H\Delta E}_\lambda^{-1}(\gamma)\right)^{-1}\tilde{\underline{h}}_\lambda$ (red). Results show that it is essential to consider the delta structure in the eigenvectors. Numerical values are given in Appendix A; calculations were performed on a 3k grid (108 k-points) using a Kohn-Sham band structure, with 6 transition bands.

To illustrate the statement: This can be justified if the structure of the eigenvector prime is determined by the matrix inversion and not by the vector that is multiplied by it.

We show a zoom in in (Fig. 5.18), which shows for the exact diagonalization (black) three minima at different places for $\tilde{e}_\lambda$.
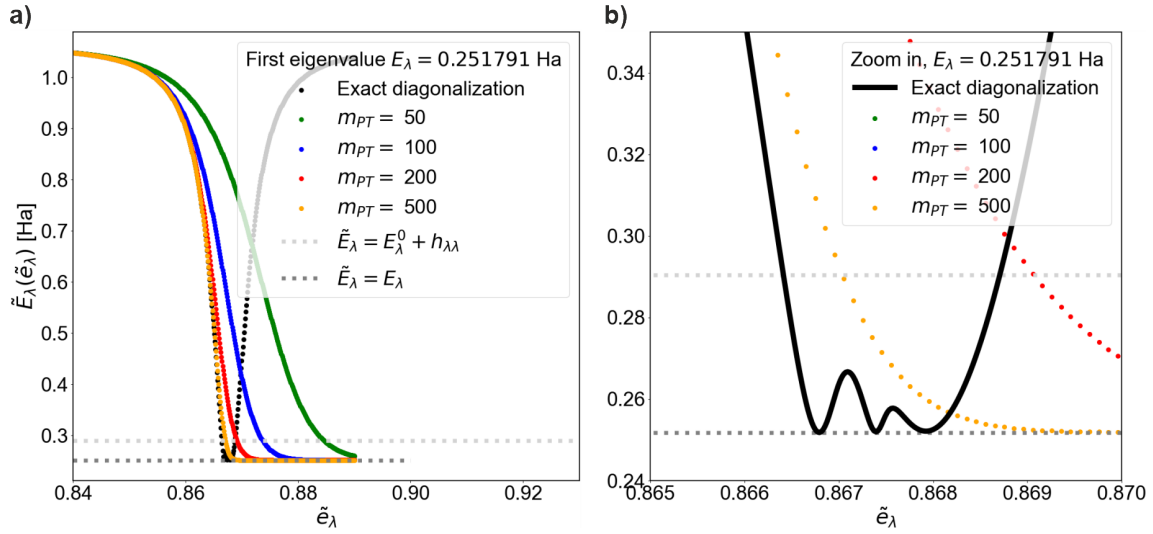

Figure 11: Calculation of the energies $\tilde{E}_\lambda = \langle \tilde{A}_\lambda(\tilde{e}_\lambda) | \hat{H}^{2p,reso} | \tilde{A}_\lambda(\tilde{e}_\lambda) \rangle$ depending on the effective matrix structure factor $\tilde{e}_\lambda$. Exact result for lithium fluoride in black, as well as modified perturbation theory expansion of the exact eigenvector (Eq. (5.83)) depending on the order 50 (green), 100 (blue), 200 (red), 500 (orange) for **a)** the first eigenvalue ($\lambda = 1$) and **b)** zoom-in of a). Shown as a horizontal dark grey line the exact value of the eigenenergy E_λ and in light grey the continuum approximation $E_\lambda \approx E_\lambda^0 + h_{\lambda\lambda}$ (Eq. (5.65)). In the zoom-in, it can be seen that for the exact eigenvector three minima form instead of just one. Numerical values are given in Appendix A; calculations were performed on a 3k grid (108 k-points) using a Kohn-Sham band structure, with 6 transition bands.

As shown in the table below the value for e_λ of the minima corresponds precisely to the e_λ that would be necessary to obtain the first three eigenvalues from the multiplication of the diagonal element of the matrix $E_\lambda^0 + h_{\lambda\lambda}$, belonging to the first eigenvalue ($\lambda = 1$), and the obtained minima $\tilde{e}_{\lambda=1,\tilde{\lambda}}^{min}, \tilde{\lambda} \in \{1, 2, 3\}$.

	E_λ [Ha]	$\tilde{e}_{\lambda=1,\tilde{\lambda}}^{min}$	$\tilde{e}_{\lambda=1,\tilde{\lambda}}^{min} \cdot (E_1^0 + h_{1,1})$ [Ha]	ΔE_λ [meV]
$\lambda, \tilde{\lambda} = 1$	0.251791	0.866793	0.251792	0.027211
$\lambda, \tilde{\lambda} = 2$	0.251967	0.867387	0.251965	0.054422
$\lambda, \tilde{\lambda} = 3$	0.252121	0.867931	0.252123	0.054422

We can see in the table that the difference ΔE_λ between the exact E_λ and approximated value $\tilde{e}_{\lambda=1,\tilde{\lambda}}^{min} \cdot (E_1^0 + h_{1,1})$ is in the order of 10^{-2} meV. The reason why we have three minima instead of just one, as we would expect, is unclear at this stage and needs to be studied in more detail. However, it is already an indication that the exact structure of the vector $\underline{\tilde{h}}_\lambda$ plays a subordinate role.

E. Gaussian algorithm

Since the goal is to transform a linear system of equations into a triangular form so that the system of equations can be solved from bottom to top, rows must be subtracted from each other. In an algorithm the following subtraction from top to bottom are performed: $(R_j - (a_{j,i}/a_{i,i}))R_i)$, where R_j represents the j-th row and $a_{j,i}$ is the element of the j-th row and i-th column. Then, a double loop is performed over $i : 1 \rightarrow n-1$ and $j : i+1 \rightarrow n$ to transform the matrix into a triangular shape.

The procedure of the algorithm is shown in a flowchart in figure 12. For the flowchart we assume that the matrix is not singular and therefore the gaussian algorithm is applicable.
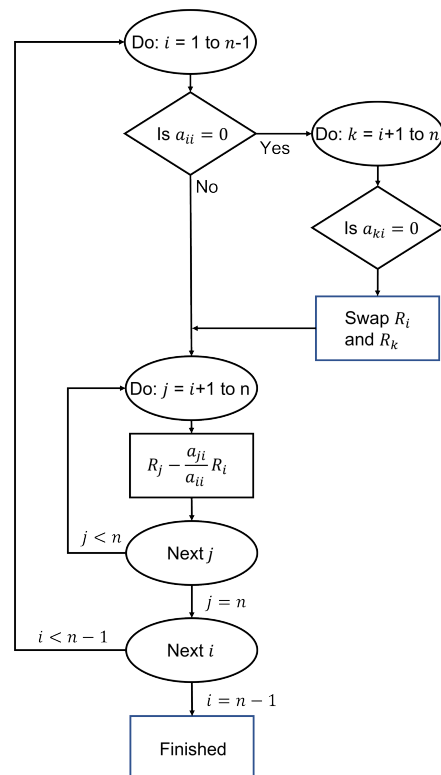

Figure 12: Flowchart Gaussian algorithm.

F. Heatmap of the resonant part of the two-particle Hamiltonian for silicon and lithium fluoride

To give an idea of the structure of the Hamiltonian, which is hermitian, we plot the matrix as a $|H^{2p,reso}|^2$-heat map for Si and LiF in the following.
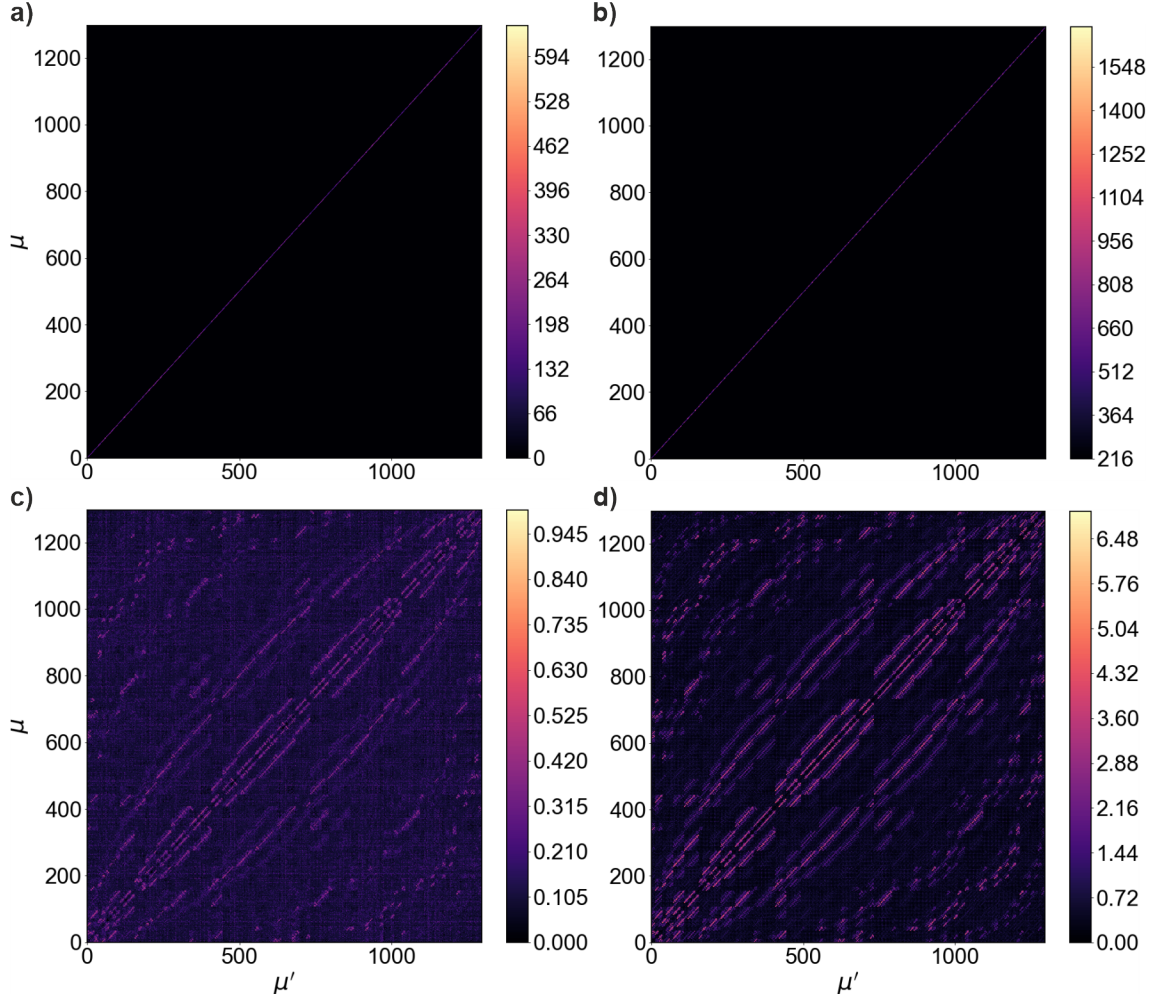

Figure 13: Heatmap of the norm for the Hamiltonian-components, $|H^{2p,reso}_{\mu\mu'}|^2$ (Eq. (4.13)), for the rows μ and columns μ'. Results shown for silicon **a/c)** and lithium flouride **b/d)**, scaled by 10^3. Heatmaps show that first, the elements in lithium fluoride are significantly stronger, up to a factor of 2.38 for the diagonals **a/b)** and up to a factor of 6.86 for the off-diagonals **c/d)**. Second, the general structure of the matrices is very similar. Numerical values are given in Appendix A; calculations were performed on a 3k grid (108 k-points) using a Kohn-Sham band structure, with 6 transition bands for silicon and lithium fluoride.

G. Broadening dependent G-matrix perturbation theory vs. Brillouin-Wigner perturbation theory

Our starting equation for developing a working perturbation series for the eigenvectors will be equation (5.39):

$$\underline{A}_\lambda = \underline{1}_\lambda + \lim_{\gamma \to 0} \left((E_\lambda + i\gamma)\underline{\underline{1}} - \underline{\underline{H}} \right)^{-1} \tilde{\underline{H}}_\lambda = \underline{1}_\lambda + \lim_{\gamma \to 0} \underline{\underline{\Delta E}}_\lambda^{-1}(\gamma) \left(\underline{\underline{1}} - \tilde{\underline{\underline{h}}} \, \underline{\underline{\Delta E}}_\lambda^{-1}(\gamma) \right)^{-1} \tilde{\underline{H}}_\lambda, \quad (1)$$

with $\tilde{\underline{H}}_\lambda = \underline{\underline{H}}\underline{1}_\lambda - E_\lambda \underline{1}_\lambda$. We want to modify our starting equation before expanding the equation into a perturbation series. The change will be that we replace E_λ by $E_\lambda^0 + h_{\lambda\lambda}$ inside the vector $\tilde{\underline{H}}_\lambda$. The same would automatically happen when we replace the eigenvalues with the continuum approximation, which we will do for practical applications. Under this assumption, our eigenvector equation becomes:

$$\underline{A}_\lambda = \underline{1}_\lambda + \lim_{\gamma \to 0} \underline{\underline{\Delta E}}_\lambda^{-1}(\gamma) \left(\underline{\underline{1}} - \tilde{\underline{\underline{h}}} \, \underline{\underline{\Delta E}}_\lambda^{-1}(\gamma) \right)^{-1} \tilde{\underline{h}}_\lambda, \quad (2)$$

To understand the effects of this modification, we will look at the bound excitonic states for Si and LiF, as the approximation would have the greatest impact on these states since the continuum approximation for these states deviates furthest from the associated energies.
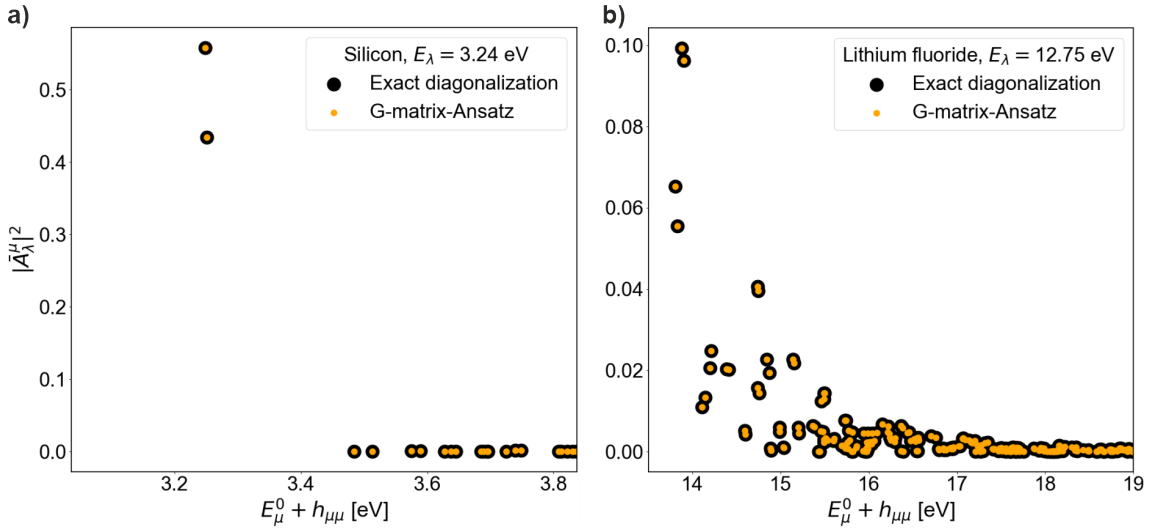

Figure 14: Normalized eigenvector $\underline{\bar{A}}_\lambda$, belonging to the smallest eigenvalue E_λ, in the G-matrix approach (Eq. (5.39), $\gamma = 10^{-6}$ Ha) with the assumption $\tilde{\underline{H}}_\lambda \approx \tilde{\underline{h}}_\lambda$ (orange) compared to the exact diagonalization eigenvector (black). Plotting the norm of the normalized eigenvector-components, $|\bar{A}_\lambda^\mu|^2$, against the diagonal elements $E_\mu^0 + h_{\mu\mu}$ of the Hamiltonian $\underline{\underline{H}}^{2p,reso}$ (Eq. (4.13)) for **a)** silicon and **b)** lithium fluoride. It turns out that the assumption $\tilde{\underline{H}}_\lambda \approx \tilde{\underline{h}}_\lambda$ leads to minor deviations (order of 0.0001-0.0008 %). Numerical values given in Appendix A calculations were performed on a 3k grid (108 k-points) using a Kohn-Sham band structure, with 6 transition bands for the optical calculation.

As figure 14 shows the approximation $\tilde{\underline{H}}_\lambda \approx \tilde{\underline{h}}_\lambda$ does not lead to any optically visible deviation of the approximated eigenvector (orange) compared to the one obtained from diagonalization (black). Looking at the most significant deviation between the eigenvector obtained from the diagonalization of the Hamiltonian $\underline{\underline{H}}^{2p,reso}$ and the eigenvector from the G-matrix approach with the

approximation $\underline{\tilde{H}}_\lambda \approx \underline{\tilde{h}}_\lambda$, we obtain 0.0001 % for Si and 0.0008 % for LiF. Thus, without fear of making a large mistake, we can assume that we can calculate the eigenvector using equation (3). After the first modification, the next step is to express the matrix inversion inside the eigenvector (Eq. (3)) through the Neumann series, as we did for the Brillouin-Wigner perturbation theory (Eq. (5.79)) so that we get in matrix-vector representation the general formular:

$$\underline{A}_\lambda = \underline{1}_\lambda + \underline{\underline{\Delta E}}_\lambda^{-1}(\gamma) \sum_{k=0}^{m_{PT}} \left(\underline{\underline{\tilde{h}}} \, \underline{\underline{\Delta E}}_\lambda^{-1}(\gamma) \right)^k \underline{\tilde{h}}_\lambda, \tag{3}$$

or in component representation up to the first order in the series expansion

$$A_\lambda^\mu = \delta_{\mu\lambda} + \frac{1}{(E_\lambda + i\gamma) - (E_\mu^0 + h_{\mu\mu})} \tilde{h}_{\mu\lambda}$$

$$+ \frac{1}{(E_\lambda + i\gamma) - (E_\mu^0 + h_{\mu\mu})} \sum_{\mu'} \frac{1}{(E_\lambda + i\gamma) - (E_{\mu'}^0 + h_{\mu'\mu'})} \tilde{h}_{\mu\mu'} \tilde{h}_{\mu'\lambda} + \dots . \tag{4}$$

with the convergence condition $S_r \left(\underline{\underline{\tilde{h}}} \, \underline{\underline{\Delta E}}_\lambda^{-1}(\gamma) \right) < 1$. The broadening parameter is only one crucial difference to the Brillouin-Wigner perturbation theory (Eq. (5.79)), much more severe is the difference that we include the element $\mu = \lambda$, which does not usually occur in the series expansion. However, the difference between the two series stems from the fact that we postulated the G-matrix approach. Nevertheless, as shown in figure 15 there would be no visible difference between introducing the broadening into the Brillouin-Wigner perturbation series or using the perturbation series from equation (Eq. (4)).
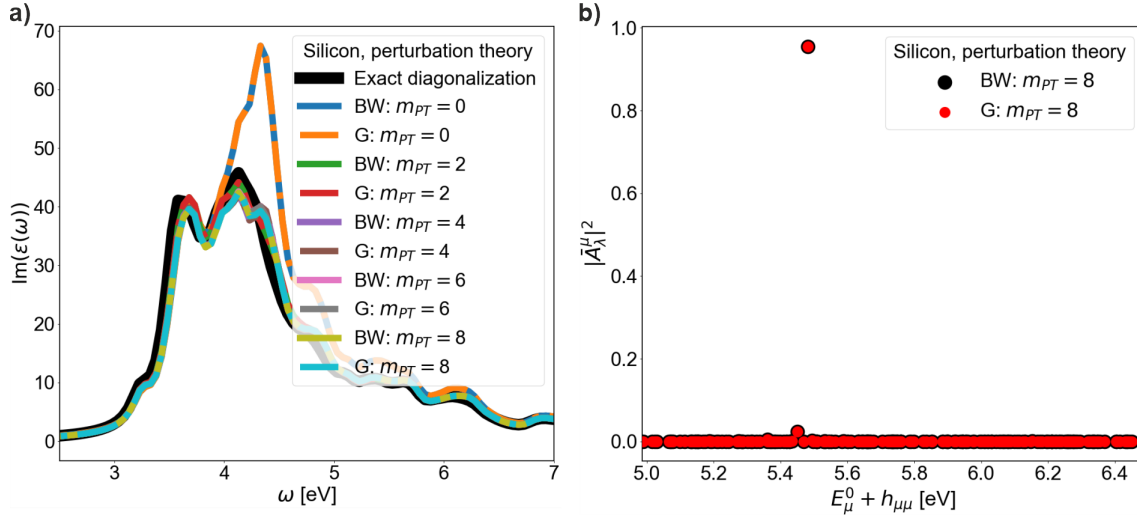

Figure 15: Modified perturbation theory ($\gamma = 3 \cdot 10^{-3}$) from the G-matrix approach (G, Eq. (3)) and the Brillouin-Wigner approach (BW, Eq. (5.83)) for the excitonic eigenvectors $\underline{\bar{A}}_\lambda$ belonging to the eigenvalues $E_\lambda = E_\lambda^0 + h_{\lambda\lambda}$ (Eq. (5.65)) in the continuum eigenvalue approximation for silicon. Shown is **a)** the absorption spectrum $\text{Im}(\epsilon(\omega))$ obtained from the modified perturbation series **b)** the norm of the normalized eigenvector-components, $|\bar{A}_\lambda^\mu|^2$, against the diagonal elements $E_\mu^0 + h_{\mu\mu}$ of the Hamiltonian $\underline{\underline{H}}^{2p,reso}$ (Eq. (4.13)). The results show that there is no optically visible difference, neither for individual eigenvectors (b)) nor for the absorption spectrum obtained (a)). Numerical values are given in Appendix A; calculations were performed on a 3k grid (108 k-points) using a Kohn-Sham band structure, with 6 transition bands for the optical calculation and a Lorentzian broadening of $3.67 \cdot 10^{-3}$ Ha.

Bibliography

[1] T. Mueller and E. Malic, "Exciton physics and device application of two-dimensional transition metal dichalcogenide semiconductors," <u>npj 2D Materials and Applications</u>, vol. 2, no. 1, p. 29, Sep 2018.

[2] G. Wang, X. Marie, I. Gerber, T. Amand, D. Lagarde, L. Bouet, M. Vidal, A. Balocchi, and B. Urbaszek, "Giant enhancement of the optical second-harmonic emission of wse$_2$ monolayers by laser excitation at exciton resonances," <u>Phys. Rev. Lett.</u>, vol. 114, p. 097403, Mar 2015.

[3] C. Palacios-Berraquero, M. Barbone, D. M. Kara, X. Chen, I. Goykhman, D. Yoon, A. K. Ott, J. Beitner, K. Watanabe, T. Taniguchi, A. C. Ferrari, and M. Atatüre, "Atomically thin quantum light-emitting diodes," <u>Nature Communications</u>, vol. 7, no. 1, p. 12978, Sep 2016.

[4] R. Huber, B. A. Schmid, R. A. Kaindl, and D. S. Chemla, "Femtosecond thz studies of intra-excitonic transitions," <u>physica status solidi (b)</u>, vol. 245, no. 6, pp. 1041–1048, 2008.

[5] C. Soonyoung, S. Ji Ho, S. Sangwan, P. Jun, H. Hoseok, H. Moon-Ho, and C. Hyunyong, "1s-intraexcitonic dynamics in monolayer MoS$_2$ probed by ultrafast mid-infrared spectroscopy." <u>Nat. Commun.</u>, vol. 7, p. 10768, 2016.

[6] P. Steinleitner, P. Merkl, A. Graf, P. Nagler, K. Watanabe, T. Taniguchi, J. Zipfel, C. Schüller, T. Korn, A. Chernikov, S. Brem, M. Selig, G. Berghäuser, E. Malic, and R. Huber, "Dielectric engineering of electronic correlations in a van der waals heterostructure," <u>Nano Letters</u>, vol. 18, no. 2, pp. 1402–1409, Feb 2018.

[7] P. Merkl, F. Mooshammer, P. Steinleitner, A. Girnghuber, K.-Q. Lin, P. Nagler, J. Holler, C. Schüller, J. M. Lupton, T. Korn, S. Ovesen, S. Brem, E. Malic, and R. Huber, "Ultrafast transition between exciton phases in van der waals heterostructures," <u>Nature Materials</u>, vol. 18, no. 7, pp. 691–696, Jul 2019.

[8] Y. Liu, A. Elbanna, W. Gao, J. Pan, Z. Shen, and J. Teng, "Interlayer excitons in transition metal dichalcogenide semiconductors for 2d optoelectronics," <u>Advanced Materials</u>, vol. 34, no. 25, p. 2107138, 2022.

[9] J. Holler, M. Selig, M. Kempf, J. Zipfel, P. Nagler, M. Katzer, F. Katsch, M. V. Ballottin, A. A. Mitioglu, A. Chernikov, P. C. M. Christianen, C. Schüller, A. Knorr, and T. Korn,

"Interlayer exciton valley polarization dynamics in large magnetic fields," Phys. Rev. B, vol. 105, p. 085303, Feb 2022.

[10] V. Villafañe, M. Kremser, R. Hübner, M. M. Petrić, N. P. Wilson, A. V. Stier, K. Müller, M. Florian, A. Steinhoff, and J. J. Finley, "Twist-dependent intra- and interlayer excitons in moiré $MoSe_2$ homobilayers," Phys. Rev. Lett., vol. 130, p. 026901, Jan 2023.

[11] C. Poellmann, P. Steinleitner, U. Leierseder, P. Nagler, G. Plechinger, M. Porer, R. Bratschitsch, C. Schüller, T. Korn, and R. Huber, "Resonant internal quantum transitions and femtosecond radiative decay of excitons in monolayer WSe_2," Nature Materials, vol. 14, no. 9, pp. 889–893, Sep 2015.

[12] G. Wang, A. Chernikov, M. M. Glazov, T. F. Heinz, X. Marie, T. Amand, and B. Urbaszek, "Colloquium: Excitons in atomically thin transition metal dichalcogenides," Rev. Mod. Phys., vol. 90, p. 021001, Apr 2018.

[13] C. Trovatello, F. Katsch, N. J. Borys, M. Selig, K. Yao, R. Borrego-Varillas, F. Scotognella, I. Kriegel, A. Yan, A. Zettl, P. J. Schuck, A. Knorr, G. Cerullo, and S. D. Conte, "The ultrafast onset of exciton formation in 2d semiconductors," Nature Communications, vol. 11, no. 1, p. 5277, Oct 2020.

[14] T. Zhang and J. Wang, "Defect-enhanced exciton–exciton annihilation in monolayer transition metal dichalcogenides at high exciton densities," ACS Photonics, vol. 8, no. 9, pp. 2770–2780, 2021.

[15] S. Latini, E. Ronca, U. De Giovannini, H. Hübener, and A. Rubio, "Cavity control of excitons in two-dimensional materials," Nano Letters, vol. 19, no. 6, pp. 3473–3479, 2019.

[16] X. Zhang, J. A. Leveillee, and A. Schleife, "Effect of dynamical screening in the bethe-salpeter framework: Excitons in crystalline naphthalene," Phys. Rev. B, vol. 107, p. 235205, Jun 2023.

[17] D. Sangalli, M. D'Alessandro, and C. Attaccalite, "Exciton-exciton transitions involving strongly bound excitons: An ab initio approach," Phys. Rev. B, vol. 107, p. 205203, May 2023.

[18] H. Y. Chen, D. Sangalli, and M. Bernardi, "First-principles ultrafast exciton dynamics and time-domain spectroscopies: Dark-exciton mediated valley depolarization in monolayer WSe_2," Phys. Rev. Res., vol. 4, p. 043203, Dec 2022.

[19] E. E. Salpeter and H. A. Bethe, "A relativistic equation for bound-state problems," Phys. Rev., vol. 84, pp. 1232–1242, Dec 1951.

[20] D. Y. Qiu, F. H. da Jornada, and S. G. Louie, "Screening and many-body effects in two-dimensional crystals: Monolayer MoS_2," Phys. Rev. B, vol. 93, p. 235435, Jun 2016.

[21] M. Vanzini, F. Sottile, I. Reshetnyak, S. Ciuchi, L. Reining, and M. Gatti, "Design of auxiliary systems for spectroscopy," Faraday Discuss., vol. 224, pp. 424–447, 2020.

[22] A. Lorin, M. Gatti, L. Reining, and F. Sottile, "First-principles study of excitons in optical spectra of silver chloride," Phys. Rev. B, vol. 104, p. 235149, Dec 2021.

[23] F. Bechstedt, Many-Body Approach to Electronic Excitations. Springer, Berlin, Heidelberg, 2016.

[24] F. Fuchs, C. Rödl, A. Schleife, and F. Bechstedt, "Efficient $\mathcal{O}(N^2)$ approach to solve the Bethe-Salpeter equation for excitonic bound states," Phys. Rev. B, vol. 78, p. 085103, Aug 2008.

[25] I. M. Alliati, D. Sangalli, and M. Grüning, "Double k-grid method for solving the bethe-salpeter equation via lanczos approaches." Front. Chem., vol. 9, p. 763946, 2022.

[26] C. Lanczos, "An iteration method for the solution of the eigenvalue problem of linear differential and integral operators," Journal of Research of the National Bureau of Standards, vol. 45, pp. 255–282, 1950.

[27] J. G. F. Francis, "The qr transformation a unitary analogue to the lr transformation—part 1," The Computer Journal, vol. 4, no. 3, pp. 265–271, 01 1961.

[28] ——, "The qr transformation part 2," The Computer Journal, vol. 4, no. 4, pp. 332–345, 01 1962.

[29] V. Kublanovskaya, "On some algorithms for the solution of the complete eigenvalue problem," USSR Computational Mathematics and Mathematical Physics, vol. 1, no. 3, pp. 637–657, 1962.

[30] G. Peters and J. H. Wilkinson, "The calculation of specified eigenvectors by inverse iteration," Handbook for Automatic Computation, vol. 2, pp. 418–439, 1971.

[31] J. J. M. Cuppen, "A divide and conquer method for the symmetric tridiagonal eigenproblem," Numerische Mathematik, vol. 36, no. 2, pp. 177–195, Jun 1980.

[32] I. C. F. Ipsen, "Computing an eigenvector with inverse iteration," SIAM Review, vol. 39, no. 2, pp. 254–291, 1997.

[33] G. L. G. Sleijpen and H. A. Van der Vorst, "A jacobi-davidson iteration method for linear eigenvalue problems," SIAM Review, vol. 42, no. 2, pp. 267–293, 2000.

[34] P. Bientinesi, I. S. Dhillon, and R. A. van de Geijn, "A parallel eigensolver for dense symmetric matrices based on multiple relatively robust representations," SIAM Journal on Scientific Computing, vol. 27, no. 1, pp. 43–66, 2005.

[35] M. Rohlfing and S. G. Louie, "Electron-hole excitations in semiconductors and insulators," Phys. Rev. Lett., vol. 81, pp. 2312–2315, Sep 1998.

[36] D. Y. Qiu, F. H. da Jornada, and S. G. Louie, "Solving the bethe-salpeter equation on a subspace: Approximations and consequences for low-dimensional materials," Phys. Rev. B, vol. 103, p. 045117, Jan 2021.

[37] M. Marsili, A. Molina-Sánchez, M. Palummo, D. Sangalli, and A. Marini, "Spinorial formulation of the GW-BSE equations and spin properties of excitons in two-dimensional transition metal dichalcogenides," Phys. Rev. B, vol. 103, p. 155152, Apr 2021.

[38] M. Gatti, Ab-Initio Calculations of Many body effects in electronic spectra. PhD Thesis, 2007. [Online]. Available: https://etsf.polytechnique.fr/node/1744

[39] F. Sottile, Response functions of semiconductors and insulators: from the Bethe-Salpeter equation to time-dependent density functional theory. PhD Thesis, 2003. [Online]. Available: https://etsf.polytechnique.fr/node/1671

[40] D. E. Aspnes and A. A. Studna, "Dielectric functions and optical parameters of Si, Ge, GaP, GaAs, GaSb, InP, InAs, and InSb from 1.5 to 6.0 ev," Phys. Rev. B, vol. 27, pp. 985–1009, Jan 1983.

[41] K. K. Rao, T. J. Moravec, J. C. Rife, and R. N. Dexter, "Vacuum ultraviolet reflectivities of lif, naf, and kf," Phys. Rev. B, vol. 12, pp. 5937–5950, Dec 1975.

[42] R. Gross and A. Marx, Festkörperphysik. De Gruyter, Oldenbourg, 2018.

[43] M. Born and R. Oppenheimer, "Zur quantentheorie der molekeln," Annalen der Physik, vol. 389, no. 20, pp. 457–484, 1927.

[44] G. Czycholl, Theoretische Festköperphysik Band 2. Springer Spektrum, Berlin, Heidelberg, 2017.

[45] G. D. Mahan, Many-Particle-Physics. Springer New York, NY, 2000.

[46] R. M. Martin, L. Reining, and D. M. Ceperley, Interacting Electrons: Theory and Computational Approaches. Cambridge University Press, 2016.

[47] S. L. Adler, "Quantum theory of the dielectric constant in real solids," Phys. Rev., vol. 126, pp. 413–420, Apr 1962.

[48] N. Wiser, "Dielectric constant with local field effects included," Phys. Rev., vol. 129, pp. 62–69, Jan 1963.

[49] A. Marini, "Ab initio finite-temperature excitons," Phys. Rev. Lett., vol. 101, p. 106405, Sep 2008.

[50] M. Zacharias, C. E. Patrick, and F. Giustino, "Stochastic approach to phonon-assisted optical absorption," Phys. Rev. Lett., vol. 115, p. 177401, Oct 2015.

[51] M. Zacharias and F. Giustino, "One-shot calculation of temperature-dependent optical spectra and phonon-induced band-gap renormalization," Phys. Rev. B, vol. 94, p. 075125, Aug 2016.

[52] R. A. Jishi, Feynman Diagram Techniques in Condensed Matter Physics. Cambridge University Press, 2013.

[53] J. Schwinger, "On the green's functions of quantized fields: I." <u>Proc. Natl. Acad. Sci. U. S. A.</u>, vol. 37, pp. 452–455, 1951.

[54] S. Albrecht, <u>Optical Absorption Spectra of Semiconductors and Insulators: ab initio calculations of many-body effects</u>. PhD Thesis, 1999. [Online]. Available: https://etsf.polytechnique.fr/node/3442

[55] E. Runge and E. K. U. Gross, "Density-functional theory for time-dependent systems," <u>Phys. Rev. Lett.</u>, vol. 52, pp. 997–1000, Mar 1984.

[56] F. Bruneval, <u>Exchange and Correlation in the Electronic Structure of Solids, from Silicon to Cuprous Oxide: GW Approximation and beyond</u>. PhD Thesis, 2005. [Online]. Available: https://etsf.polytechnique.fr/node/1670

[57] L. Hedin, "New method for calculating the one-particle green's function with application to the electron-gas problem," <u>Phys. Rev.</u>, vol. 139, pp. A796–A823, Aug 1965.

[58] A. Schindlmayr and R. W. Godby, "Systematic vertex corrections through iterative solution of hedin's equations beyond the GW approximation," <u>Phys. Rev. Lett.</u>, vol. 80, pp. 1702–1705, Feb 1998.

[59] P. Romaniello, D. Sangalli, J. A. Berger, F. Sottile, L. G. Molinari, L. Reining, and G. Onida, "Double excitations in finite systems," <u>The Journal of Chemical Physics</u>, vol. 130, no. 4, p. 044108, 01 2009.

[60] F. Bruneval, N. Vast, L. Reining, M. Izquierdo, F. Sirotti, and N. Barrett, "Exchange and correlation effects in electronic excitations of Cu_2O," <u>Phys. Rev. Lett.</u>, vol. 97, p. 267601, Dec 2006.

[61] H. H. Li, "Refractive index of silicon and germanium and its wavelength and temperature derivatives," <u>Journal of Physical and Chemical Reference Data</u>, vol. 9, no. 3, pp. 561–658, 07 1980.

[62] M. G. Mead, "Comparison of the optical and dielectric properties of crystalline and molten lithium fluoride," <u>Journal of Physics C: Solid State Physics</u>, vol. 7, no. 2, p. 445, jan 1974.

[63] W. Hanke and L. J. Sham, "Many-particle effects in the optical spectrum of a semiconductor," <u>Phys. Rev. B</u>, vol. 21, pp. 4656–4673, May 1980.

[64] G. Onida, L. Reining, and A. Rubio, "Electronic excitations: density-functional versus many-body green's-function approaches," <u>Rev. Mod. Phys.</u>, vol. 74, pp. 601–659, Jun 2002.

[65] T. Sander, E. Maggio, and G. Kresse, "Beyond the tamm-dancoff approximation for extended systems using exact diagonalization," <u>Phys. Rev. B</u>, vol. 92, p. 045209, Jul 2015.

[66] C. Rödl, <u>Elektronische und exzitonische Anregungen in magnetischen Isolatoren</u>. PhD Thesis, 2009. [Online]. Available: https://www.db-thueringen.de/receive/dbt_mods_00016531

[67] M. Rohlfing and S. G. Louie, "Electron-hole excitations and optical spectra from first principles," Phys. Rev. B, vol. 62, pp. 4927–4944, Aug 2000.

[68] H. Stolz, Einführung in die Vielelektronentheorie der Kristalle. Akademie-Verlag, Berlin, 1974.

[69] B. Zhu, X. Chen, and X. Cui, "Exciton binding energy of monolayer WS_2," Scientific Reports, vol. 5, no. 1, p. 9218, Mar 2015.

[70] A. Chernikov, T. C. Berkelbach, H. M. Hill, A. Rigosi, Y. Li, B. Aslan, D. R. Reichman, M. S. Hybertsen, and T. F. Heinz, "Exciton binding energy and nonhydrogenic rydberg series in monolayer WS_2," Phys. Rev. Lett., vol. 113, p. 076802, Aug 2014.

[71] W. G. Schmidt, S. Glutsch, P. H. Hahn, and F. Bechstedt, "Efficient $\mathcal{O}(N^2)$ method to solve the bethe-salpeter equation," Phys. Rev. B, vol. 67, p. 085307, Feb 2003.

[72] R. M. Martin, Electronic Structure: Basic Theory and Practical Methods. Cambridge University Press, 2004.

[73] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," Phys. Rev., vol. 136, pp. B864–B871, Nov 1964.

[74] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," Phys. Rev., vol. 140, pp. A1133–A1138, Nov 1965.

[75] W. Kohn, "v-representability and density functional theory," Phys. Rev. Lett., vol. 51, pp. 1596–1598, Oct 1983.

[76] F. Giustino, Materials modelling using Density Functional Theory. Cambridge University Press, 2020.

[77] J. P. Perdew and M. Levy, "Physical content of the exact kohn-sham orbital energies: Band gaps and derivative discontinuities," Phys. Rev. Lett., vol. 51, pp. 1884–1887, Nov 1983.

[78] R. W. Godby, M. Schlüter, and L. J. Sham, "Self-energy operators and exchange-correlation potentials in semiconductors," Phys. Rev. B, vol. 37, pp. 10 159–10 175, Jun 1988.

[79] M. S. Hybertsen and S. G. Louie, "Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies," Phys. Rev. B, vol. 34, pp. 5390–5413, Oct 1986.

[80] F. Bruneval, N. Vast, and L. Reining, "Effect of self-consistency on quasiparticles in solids," Phys. Rev. B, vol. 74, p. 045102, Jul 2006.

[81] I. D. White, R. W. Godby, M. M. Rieger, and R. J. Needs, "Dynamic image potential at an Al(111) surface," Phys. Rev. Lett., vol. 80, pp. 4265–4268, May 1998.

[82] O. Pulci, L. Reining, G. Onida, R. Del Sole, and F. Bechstedt, "Many-body effects on one-electron energies and wavefunctions in low dimensional systems," Computational Materials Science, vol. 20, p. 300, 2001.

[83] "The abinitproject: Impact, environment and recent developments," <u>Computer Physics Communications</u>, vol. 248, p. 107042, 2020.

[84] A. L. Fetter and J. D. Walecka, <u>Quantum Theory of Many-Particle Systems</u>. McGraw-Hill, New York, 1971.

[85] I. Tamm, "Relativistic interaction of elementary particles," <u>J. Rev. (USSR) 9:449</u>, 1945.

[86] S. M. Dancoff, "Non-adiabatic meson theory of nuclear forces," <u>Phys. Rev.</u>, vol. 78, pp. 382–385, May 1950.

[87] M. Grüning, A. Marini, and X. Gonze, "Exciton-plasmon states in nanoscale materials: Breakdown of the tamm-dancoff approximation," <u>Nano Letters</u>, vol. 9, no. 8, pp. 2820–2824, 2009.

[88] K. Shindo, "Effective electron-hole interaction in shallow excitons," <u>Journal of the Physical Society of Japan</u>, vol. 29, no. 2, pp. 287–296, 1970.

[89] S. Botti and M. A. L. Marques, "Strong renormalization of the electronic band gap due to lattice polarization in the GW formalism," <u>Phys. Rev. Lett.</u>, vol. 110, p. 226404, May 2013.

[90] C. Attaccalite, M. Grüning, and A. Marini, "Real-time approach to the optical properties of solids and nanostructures: Time-dependent bethe-salpeter equation," <u>Phys. Rev. B</u>, vol. 84, p. 245110, Dec 2011.

[91] A. Marini and R. Del Sole, "Dynamical excitonic effects in metals and semiconductors," <u>Phys. Rev. Lett.</u>, vol. 91, p. 176402, Oct 2003.

[92] R. Zimmermann, "Dynamical screening of the wannier exciton," <u>physica status solidi (b)</u>, vol. 48, no. 2, pp. 603–618, 1971.

[93] J. Singh, D. Birkedal, V. G. Lyssenko, and J. M. Hvam, "Binding energy of two-dimensional biexcitons," <u>Phys. Rev. B</u>, vol. 53, pp. 15 909–15 913, Jun 1996.

[94] S. Das, G. Shi, N. Sanders, and E. Kioupakis, "Electronic and optical properties of two-dimensional α-PbO from first principles," <u>Chemistry of Materials</u>, vol. 30, no. 20, pp. 7124–7129, 2018.

[95] K. Hummer and C. Ambrosch-Draxl, "Oligoacene exciton binding energies: Their dependence on molecular size," <u>Phys. Rev. B</u>, vol. 71, p. 081202, Feb 2005.

[96] A. B. Kunz, T. Miyakawa, and S. Oyama, "Electronic energy bands, excitons, and plasmons in lithium fluoride crystal," <u>physica status solidi (b)</u>, vol. 34, no. 2, pp. 581–589, 1969.

[97] G. H. Wannier, "The structure of electronic excitation levels in insulating crystals," <u>Phys. Rev.</u>, vol. 52, pp. 191–197, Aug 1937.

[98] G. Czycholl, <u>Theoretische Festköperphysik Band 1</u>. Springer Spektrum, Berlin, Heidelberg, 2017.

[99] C. Sommer, P. Krüger, and J. Pollmann, "Optical spectra of alkali-metal fluorides," Phys. Rev. B, vol. 86, p. 155212, Oct 2012.

[100] Y. M. Niquet, D. Rideau, C. Tavernier, H. Jaouen, and X. Blase, "Onsite matrix elements of the tight-binding hamiltonian of a strained crystal: Application to silicon, germanium, and their alloys," Phys. Rev. B, vol. 79, p. 245201, Jun 2009.

[101] I. Reshetnyak, Computing optical properties and photo-emission spectra : a new starting point. PhD Thesis, 2015. [Online]. Available: https://etsf.polytechnique.fr/node/3441

[102] G. Cappellini, R. Del Sole, L. Reining, and F. Bechstedt, "Model dielectric function for semiconductors," Phys. Rev. B, vol. 47, pp. 9892–9895, Apr 1993.

[103] M. L. Trolle, T. G. Pedersen, and V. Véniard, "Model dielectric function for 2D semiconductors including substrate screening," Scientific Reports, vol. 7, no. 1, p. 39844, Jan 2017.

[104] S. S. Dong, M. Govoni, and G. Galli, "Machine learning dielectric screening for the simulation of excited state properties of molecules and materials," Chem. Sci., vol. 12, pp. 4970–4980, 2021.

[105] R. Haydock, "Rounding errors in the recursion method: Inner products," Computer Physics Communications, vol. 30, no. 3, pp. 221–228, 1983.

[106] M. Grüning, A. Marini, and X. Gonze, "Implementation and testing of lanczos-based algorithms for random-phase approximation eigenproblems," Computational Materials Science, vol. 50, no. 7, pp. 2148–2156, 2011.

[107] M. P. Ljungberg, P. Koval, F. Ferrari, D. Foerster, and D. Sánchez-Portal, "Cubic-scaling iterative solution of the bethe-salpeter equation for finite systems," Phys. Rev. B, vol. 92, p. 075422, Aug 2015.

[108] Y. Saad, Iterative Methods for Sparse Linear Systems, 2nd ed. Society for Industrial and Applied Mathematics, 2003.

[109] M. Shao, F. H. da Jornada, L. Lin, C. Yang, J. Deslippe, and S. G. Louie, "A structure preserving lanczos algorithm for computing the optical absorption spectrum," SIAM Journal on Matrix Analysis and Applications, vol. 39, no. 2, pp. 683–711, 2018.

[110] D. Sangalli, A. Ferretti, H. Miranda, C. Attaccalite, I. Marri, E. Cannuccia, P. Melo, M. Marsili, F. Paleari, A. Marrazzo, G. Prandini, P. Bonfà, M. O. Atambo, F. Affinito, M. Palummo, A. Molina-Sánchez, C. Hogan, M. Grüning, D. Varsano, and A. Marini, "Many-body perturbation theory calculations using the yambo code," Journal of Physics: Condensed Matter, vol. 31, no. 32, p. 325902, may 2019.

[111] H. Weikert, H. Meyer, L. S. Cederbaum, and F. Tarantelli, "Block Lanczos and many-body theory: Application to the one-particle Green's function," The Journal of Chemical Physics, vol. 104, no. 18, pp. 7122–7138, 05 1996.

[112] L. X. Benedict and E. L. Shirley, "Ab initio calculation of $\epsilon_2(\omega)$ including the electron-hole interaction: Application to GaN and CaF$_2$," Phys. Rev. B, vol. 59, pp. 5441–5451, Feb 1999.

[113] D. Werner, Funktionalanalysis.   Springer Berlin, Heidelberg, 2005.

[114] C. Emminger, F. Abadizaman, N. S. Samarasingha, T. E. Tiwald, and S. Zollner, "Temperature dependent dielectric function and direct bandgap of ge," Journal of Vacuum Science and Technology B, vol. 38, no. 1, p. 012202, 12 2019.

[115] A. Bechiri, F. Benmakhlouf, and N. Bouarissa, "Calculation of electronic and optical properties of zn-based ii–vi semiconductors," Physics Procedia, vol. 2, no. 3, pp. 803–812, 2009, proceedings of the JMSM 2008 Conference.

[116] S. Ozaki and S. Adachi, "Optical constants of ZnSxSe$_1-$x ternary alloys," Journal of Applied Physics, vol. 75, no. 11, pp. 7470–7475, 06 1994.

[117] A. Aouina, M. Gatti, and L. Reining, "Strategies to build functionals of the density, or functionals of green's functions: what can we learn?" Faraday Discuss., vol. 224, pp. 27–55, 2020.

[118] M. Vanzini, A. Aouina, M. Panholzer, M. Gatti, and L. Reining, "Connector theory for reusing model results to determine materials properties," npj Computational Materials, vol. 8, no. 1, p. 98, May 2022.

[119] T. C. Berkelbach, Microscopic theories of excitons and their dynamics.   PhD Thesis, 2014. [Online]. Available: https://academiccommons.columbia.edu/doi/10.7916/D8BP010F

[120] W. E. Arnoldi, "The principle of minimized iterations in the solution of the matrix eigenvalue problem," Quarterly of Applied Mathematics, vol. 9, pp. 17–29, 1951.

[121] P. Benner, S. Dolgov, V. Khoromskaia, and B. N. Khoromskij, "Fast iterative solution of the bethe–salpeter eigenvalue problem using low-rank and qtt tensor approximation," Journal of Computational Physics, vol. 334, pp. 221–239, 2017.

[122] R. Littlejohn. (2020) Physics 221ab, quantum mechanics, fall 2019 and spring 2020, university of california, berkeley, notes 22: Bound-state perturbation theory. [Online]. Available: https://bohr.physics.berkeley.edu/classes/221/1920/221.html

[123] Z. Li and N. A. Smirnova, "Converging many-body perturbation theory for ab initio nuclear-structure: I. brillouin-wigner perturbation series for closed-shell nuclei," 2023.

[124] K. Carter-Fenk and M. Head-Gordon, "Repartitioned brillouin-wigner perturbation theory with a size-consistent second-order correlation energy," The Journal of Chemical Physics, vol. 158, no. 23, 2023.

[125] L. Fiedler, K. Shah, M. Bussmann, and A. Cangi, "Deep dive into machine learning density functional theory for materials science and chemistry," Phys. Rev. Mater., vol. 6, p. 040301, Apr 2022.

[126] R. Pederson, B. Kalita, and K. Burke, "Machine learning and density functional theory," Nature Reviews Physics, vol. 4, no. 6, pp. 357–358, Jun 2022.

[127] Z. Yi, Y. Fu, and H. J. Tang, "Neural networks based approach for computing eigenvectors and eigenvalues of symmetric matrix," Computers and Mathematics with Applications, vol. 47, no. 8, pp. 1155–1164, 2004.

[128] Y. Zhang, K. Chen, W. Ma, and X.-D. Li, "Matlab simulation of gradient-based neural network for online matrix inversion," in Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence, D.-S. Huang, L. Heutte, and M. Loog, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 98–109.

[129] K. Ghosh, A. Stuke, M. Todorović, P. B. Jørgensen, M. N. Schmidt, A. Vehtari, and P. Rinke, "Deep learning spectroscopy: Neural networks for molecular excitation spectra," Advanced Science, vol. 6, no. 9, p. 1801367, 2019.

[130] G. Pilania, A. Mannodi-Kanakkithodi, B. P. Uberuaga, R. Ramprasad, J. E. Gubernatis, and T. Lookman, "Machine learning bandgaps of double perovskites," Scientific Reports, vol. 6, no. 1, p. 19375, Jan 2016.

[131] L. Luo, L. Men, Z. Liu, Y. Mudryk, X. Zhao, Y. Yao, J. M. Park, R. Shinar, J. Shinar, K.-M. Ho, I. E. Perakis, J. Vela, and J. Wang, "Ultrafast terahertz snapshots of excitonic rydberg states and electronic coherence in an organometal halide perovskite," Nature Communications, vol. 8, no. 1, p. 15565, Jun 2017.

[132] D. R. Hamann, "Optimized norm-conserving vanderbilt pseudopotentials," Phys. Rev. B, vol. 88, p. 085117, Aug 2013.

[133] S. Botti, F. Sottile, N. Vast, V. Olevano, L. Reining, H.-C. Weissker, A. Rubio, G. Onida, R. Del Sole, and R. W. Godby, "Long-range contribution to the exchange-correlation kernel of time-dependent density functional theory," Phys. Rev. B, vol. 69, p. 155112, Apr 2004.

[134] P. Haas, F. Tran, and P. Blaha, "Calculation of the lattice constant of solids with semilocal functionals," Phys. Rev. B, vol. 79, p. 085104, Feb 2009.

[135] A. Riefer and W. G. Schmidt, "Solving the bethe-salpeter equation for the second-harmonic generation in zn chalcogenides," Phys. Rev. B, vol. 96, p. 235206, Dec 2017.

[136] A. D. Corso, F. Mauri, and A. Rubio, "Density-functional theory of the nonlinear optical susceptibility: Application to cubic semiconductors," Phys. Rev. B, vol. 53, pp. 15 638–15 642, Jun 1996.

Declaration of Authorship

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen Hilfsmittel als die angegebenen verwendet habe. Die Stellen, die anderen Werken (gilt ebenso für Werke aus elektronischen Datenbanken oder aus dem Internet) wörtlich oder sinngemäß entnommen sind, habe ich unter Angabe der Quelle und Einhaltung der Regeln wissenschaftlichen Zitierens kenntlich gemacht. Über wissenschaftlich korrektes Arbeiten wurde ich aufgeklärt. Diese Versicherung umfasst auch in der Arbeit verwendete bildliche Darstellungen, Tabellen, Kartenskizzen und gelieferte Zeichnungen. Ebenso versichere ich, dass die vorgelegten Druckexemplare und die abgegebene elektronische Version der Arbeit identisch sind, sowie, dass ich diese Arbeit nicht bereits an einer anderen Universität zur Erlangung eines akademischen Grades eingereicht habe. Mir ist bewusst, dass Täuschungen nach der für mich gültigen Studien- und Prüfungsordnung geahndet werden. Ich habe Kenntnis von den in § 26 Abs. 5 vorgesehenen Rechtsfolgen.

_____   _____

Unterschrift   Ort, Datum