12-2000

Electronic Excitations in Metals and Semiconductors: Ab Initio Studies of Realistic Many-Particle Systems

Wei Ku
University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a dissertation written by Wei Ku entitled "Electronic Excitations in Metals and Semiconductors: Ab Initio Studies of Realistic Many-Particle Systems." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Physics.

Adolfo G. Eguiluz, Major Professor

We have read this dissertation and recommend its acceptance:

Bennett C. Larson, John J. Quinn, Peter T. Cummings

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a thesis written by Wei Ku entitled "Electronic Excitations in Metals and Semiconductors: *Ab Initio* Studies of Realistic Many-Particle Systems." I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Physics.

Adolfo G. Eguiluz, Major Professor

We have read this dissertation
and recommend its acceptance:

Bennett C. Larson

John J. Quinn

Peter T. Cummings

Accepted for the Council:

Dr. Anne Mayhew

Interim Vice Provost and
Dean of The Graduate School

(Original signatures are on file in the Graduate Admissions and Records Office.)

Electronic Excitations in Metals and Semiconductors:

Ab Initio Studies of Realistic Many-Particle Systems

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Wei Ku

December 2000

This work is dedicated to my parents:

Kui-Din Ku

and

Chiu-Mei Chang Ku

and my lovely wife:

Li-Yu Yang.

Acknowledgements

I am grateful to Dr. Adolfo G. Eguiluz for his guidance and financial support during my graduate research. I greatly benefited from his availability to discussion of physics and career decisions, his serious attitude in physics research, and his help in exploring potential post-doc positions in several good places. The strong interaction between the group members in the (lengthy) weekly group meeting proves to be valuable and stimulating. In the process of extracting information from the WIEN package and developing algorithms for evaluating the transition matrix elements, collaboration with James M. Sullivan is crucial to the final success.

In my graduate courses, John J. Quinn has given me exciting lectures on condensed matter physics. Geoffrey S. Canright has given me inspirational lectures on classical mechanics and chaos theory, as well as guidance in the early stage of my graduate study. William M. Bugg and James E. Parks' support in providing graduate teaching assistantship and partial support from the Science Alliance are appreciated.

Many discussions with Ben C. Larson in the Solid State Division of Oak Ridge National Laboratory have proven to be insightful. The support from John F. Cooke in building the Hive cluster of the division is essential to speeding up the developing process of the all-electron GWA code of this work. Sharron K. King helped greatly in managing all the computers in the division and providing a smooth computing environment.

The importance of the training on high performance computing provided by JICS and the local numerical tech support from JICS cannot be exaggerated. Portion of the calculations is performed in NERSC.

Finally, this work will not be finished in the limited time period without the full support from my wife, Li-Yu Yang, who has to live with my irregular working hours for the past four years.

Abstract

Electronic excitations in metals and semiconductors, measured through inelastic x-ray scattering, electronic energy-loss spectroscopy, and angle-resolved photoemission, provide valuable information about the intrinsic many-body interactions between electrons in the crystal environment. These dynamical interactions are usually analyzed with intuitive simple models. With the problems studied, which correspond to linear response and quasi-particle excitation, it will be shown that *ab initio* approaches not only can provide more realistic understanding of microscopic processes and mechanisms, but also result in new theoretical interpretations that resolve a number of remarkable "anomalies." Specifically, these problems include the dispersion of the plasmon lifetime in potassium, the electron-hole excitations near the *d*-band threshold of zinc, the occupied bandwidth of sodium, and the band gap of silicon and germanium. These analyses are based on newly developed all-electron, full potential implementations of the time-dependent density functional theory and the conserving finite temperature many-body perturbation theory. These studies also illustrate the limitation of current knowledge of many-body approaches, and demonstrate the importance of the interplay between experiment and theory.

Contents

List of Tables

List of Figures

Chapter 1 Introduction

Electronic excitations in metals and semiconductors, measured through inelastic x-ray scattering, electronic energy-loss spectroscopy, and angle-resolved photoemission, provide valuable information about the intrinsic many-body interactions between electrons in the crystal environment. The information is obtained through comparison between experiment and theory in a wide range of phase space consisting of momentum and energy dependence.

While the comparison is commonly carried out with simplified theoretical models to gain physical intuition of the systems, it will be demonstrated that more elaborate *ab intio* approaches not only can supply more quantitative description, but they also offer insight into realistic and microscopic processes and mechanisms. Furthermore, new theoretical interpretations are given that resolve a number of remarkable "anomalies."

In particular, two kinds of electronic excitations are studied: linear response and quasi-particle excitation. For linear response, several case studies including the anomalous dispersion of the plasmon lifetime in potassium (K) and the electron-hole excitations near the *d*-band threshold of zinc (Zn) are studied within the time-dependent density functional theory. For quasi-particle excitation, the occupied quasiparticle bandwidth of sodium (Na) and band gap of silicon (Si) and germanium (Ge) are considered within the Baym-Kadanoff conserving scheme of finite-temperature many-body perturbation theory.

Chapter 2 Dynamical Linear Response

In certain experimental setups, e.g.: inelastic x-ray scattering (IXS) and electronic energy-loss spectroscopy (EELS), the incident particle possesses such high momentum that it only weakly interacts with the sample and the first Born approximation is well respected [1]. The corresponding differential cross section

$$\frac{d^2\sigma}{d\Omega d\omega} = \left(\frac{d\sigma}{d\Omega}\right)_0 S(\vec{q},\omega) \tag{1}$$

is proportional to the dynamical structure factor

$$S(\vec{q},\omega) \equiv \int d^3x \int d^3x' e^{-i\vec{q}\cdot(\vec{x}-\vec{x}')} \int_{-\infty}^{+\infty} dt\, e^{i\omega t} \left\langle \hat{n}(\vec{x},t)\hat{n}(\vec{x}',0)\right\rangle, \tag{2}$$

which depends only on the properties of the sample itself, where $\vec{q}/\omega$ is the momentum/energy transfer. The momentum and energy dependence of $S(\vec{q},\omega)$ contains rich information about the electronic interaction in the crystal environment and thus provides a "play ground" for theories of many-body mechanisms. In actual implementation, the following relation is employed to compare experiment and theory

$$S(\vec{q},\omega) \propto \mathrm{Im}\,\chi(\vec{q},\omega), \tag{3}$$

where χ is the density response function, defined as variation of density, *n*, with respect to the external potential V :

$$\chi(1,2) \equiv \frac{\delta n(1)}{\delta V(2)}. \tag{4}$$

Here, notation "1" represents four-dimensional variable $(\vec{x},t)$. It is also useful to define the "dielectric function", $\varepsilon(q,\omega)$, through

$$\varepsilon \equiv (1+v\chi)^{-1}, \tag{5}$$

whose imaginary part will be referred to as the "loss function".

The theoretical task is then to produce realistic χ, which is evaluated within the framework of Time-Dependent Density Functional Theory discussed in the next section, followed by case studies of the anomalous plasmon lifetime dispersion of K and the *d*-excitation threshold behavior of Zn.

2.1 Time-Dependent Density Functional Theory

One of the theoretical approaches to describe a many-body system is Time-Dependent Density Functional Theory (TDDFT) [2,3], in which the real system is "mapped" to a virtual non-interacting system, called Kohn-Sham (KS) system. The connection between the virtual system and the real system is that the density of the KS system is proven capable of representing the exact density of the real many-body system. Therefore, it is straightforward to relate the KS density response to exact density response of the real system.

A general formulation of the time evolution of an interacting electron system in an external potential $v_e(\vec{x},t)$ has been given by Runge and Gross [4,5]. These authors established the invertibility of the mapping $v_e(\vec{x},t) \to n(\vec{x},t)$, where $n(\vec{x},t)$ is the

time-dependent density for the interacting system. From that non-trivial result it is possible to demonstrate [4,5] that $n(\vec{x},t)$ can be obtained as

$$n(\vec{x},t) = \sum_{\nu}^{occupied} \left|\varphi_{\nu}(\vec{x},t)\right|^2 , \tag{6}$$

in terms of the solutions $\varphi_{\nu}(\vec{x},t)$ of the time-dependent Kohn-Sham equation,

$$i\hbar\frac{\partial}{\partial t}\varphi_{\nu}(\vec{x},t) = \left(-\frac{\hbar^2}{2m}\nabla^2 + v_s[n](\vec{x},t)\right)\varphi_{\nu}(\vec{x},t), \tag{7}$$

where the single-particle potential $v_s[n](\vec{x},t)$ contains, in addition to the external and Hartree potentials, the exchange-correlation potential $V_{xc}[n](\vec{x},t)$, whose functional dependence on the density is implied by the notation. The TDDFT formalism is particularly well suited for the study of the linear response of a many-electron system to an external potential variation $\delta V(\vec{x},t)$ [6]. In general, the density-response function χ is defined in Eq. (4), allowing $\delta n = \chi\delta V$. In TDDFT the linear change in density can also be calculated as $\delta n = \chi^{KS}\delta v_s$, where χ^{KS} is the response for the unperturbed Kohn-Sham system. From Eqs. (6) and (7) δv_s can be related to δV, and this leads to the following integral equation for the response function [6]:

$$\chi = \chi^{KS} + \chi^{KS}(v + f_{xc})\chi , \tag{8}$$

where *v* is the bare Coulomb interaction, and the exchange-correlation kernel f_{xc} is defined by the equation

$$f_{xc}[n](\vec{x},t;\vec{x}',t') = \frac{\delta V_{xc}[n](\vec{x},t)}{\delta n(\vec{x}',t')}, \tag{9}$$

where the functional derivative is to be evaluated at the unperturbed density. The explicit effect of f_{xc} will be referred to as "dynamical" correlation effect, while the one of V_{xc} as "static" one. It needs to be emphasized that this formulism is formally exact; the approximation comes when the functional form of $V_{xc}[n](\vec{x},t)$, and the corresponding f_{xc}, is chosen.

The spectral representation for the Kohn-Sham response χ^{KS} in terms of the eigenfunctions and eigenvalues of the unperturbed Kohn-Sham problem is of the usual form. For a periodic crystal it is convenient to work with the Fourier transform of χ^{KS}, given by the equation

$$\chi^{KS}_{\vec{K},\vec{K}'}(\vec{k},\omega) = \frac{1}{\Omega}\sum_{\vec{k}'}^{BZ}\sum_{j,j'}\frac{f_{\vec{k}',j} - f_{\vec{k}'+\vec{k},j'}}{E_{\vec{k}',j} - E_{\vec{k}'+\vec{k},j'} + \hbar(\omega + i\eta)}\left\langle \vec{k}', j \left| e^{-i(\vec{k}+\vec{K})\cdot\hat{\vec{x}}} \right| \vec{k}'+\vec{k}, j' \right\rangle \\ \text{x} \quad \left\langle \vec{k}'+\vec{k}, j' \left| e^{i(\vec{k}+\vec{K}')\cdot\hat{\vec{x}}} \right| \vec{k}', j \right\rangle \ , \tag{10}$$

where Ω is the normalization volume, K' is a vector of the reciprocal lattice, *j* is a band index, and both wave vectors, *k* and *k'*, are in the first Brillouin zone (BZ), or an equivalent parallelepiped zone.

2.1.1 Numerical Implementation

In the actual implementation in this work, Eq. (10) is evaluated from Kohn-Sham states and eigenvalues obtained with the local-density approximation (LDA) for the exchange-correlation potential $V_{xc}(\vec{x})$ [7,8]. The success and the failure of this approximation in the present response context will be addressed in the case of Na and Zn. Equation (8) is

then solved numerically as a matrix equation in the above Fourier representation. Finally, the dynamical structure factor is given by $S(\vec{q},\omega) \propto \mathrm{Im}\,\chi_{\vec{K},\vec{K}}\left(\vec{k},\omega\right)$, where $\vec{q} = \vec{k} + \vec{K}$.

Adopted in this work is a full potential, all-electron approach, in which all the angular components of the potential, v_s, are taken into account and the Kohn-Sham states are represented in a basis set called "linearized augmented plane wave" (LAPW) [9,10]. The LAPW has the powerful capability to faithfully describe both the localized strong oscillation near the atomic site and extensive smooth behavior in the interstitial region of the Kohn-Sham states. Thus, this approach does not depend on the use of artificial pseudo-potential and it allows realistic description of the Kohn-Sham wave functions within the atomic sphere. This "pseudopotential free" feature turns out to be very important in obtaining a reliable result in the present work [11].

The “local-orbital” extension [10] of the LAPW method serves an important role. First, it help to stretch the energy range of the linearization. That is, the quality of the high lying empty Kohn-Sham states, necessary in Eq. (10), is improved. Second, it allows accurate description of the shallow core states, called semi-core states, which are pretty localized except for the finite extended tails in their density that partially participate the bonding. This extension makes it possible to treat these semi-core states *on the same footing* as the valence states.

Another crucial technique employed is imaginary frequency sampling. The imaginary part of Eq. (10) is essentially a sum of δ-functions broadened with the

numerical parameter η. In systems with a sharp feature in the spectra, e.g.: long-lived plasmon mode in Na or the strong "onset" in Zn, this parameter needs to be smaller than the nature linewidth of the feature in the spectra, in order for a reliable extraction of the line shape. This renders the conventional direct summation [12] over the *k′* index in Eq. (10) impractical. The approach used in this work is to avoid summing over the δ-functions by first evaluating Eq. (10) on imaginary frequency axis, on which terms in the summation are smooth, and then numerically analytic-continuing the resulting function (χ^{KS} or χ) to slightly above (δ << nature linewidth) the real axis, via Padé approximants [13-16] (details are provided in appendix 0). The power of this technique is illustrated in Figure 1, compared in which are the loss function $\mathrm{Im}\,\chi_{\vec{K}=0,\vec{K}'=0}\left(\vec{k},\omega\right)$ obtained as just outlined with its counterpart obtained via the (more standard) direct real frequency approach. Figure 1 corresponds to a 16x16x16 wave vector mesh [17]; for the same $\vec{k}$-mesh, the real-axis approach cannot resolve the linewidth. It is intriguing to note that the loss peak calculated with this new technique is quite insensitive to a decrease of δ by an order of magnitude. Thus, the numerical broadening can be controlled to be 1% or less of features of interest.

2.2 Plasmon Lifetime Dispersion of Potassium

The existence and energy dispersion of plasmon can be easily derived within the homogenous electron gas model [18]. On the other hand, much less is known about an

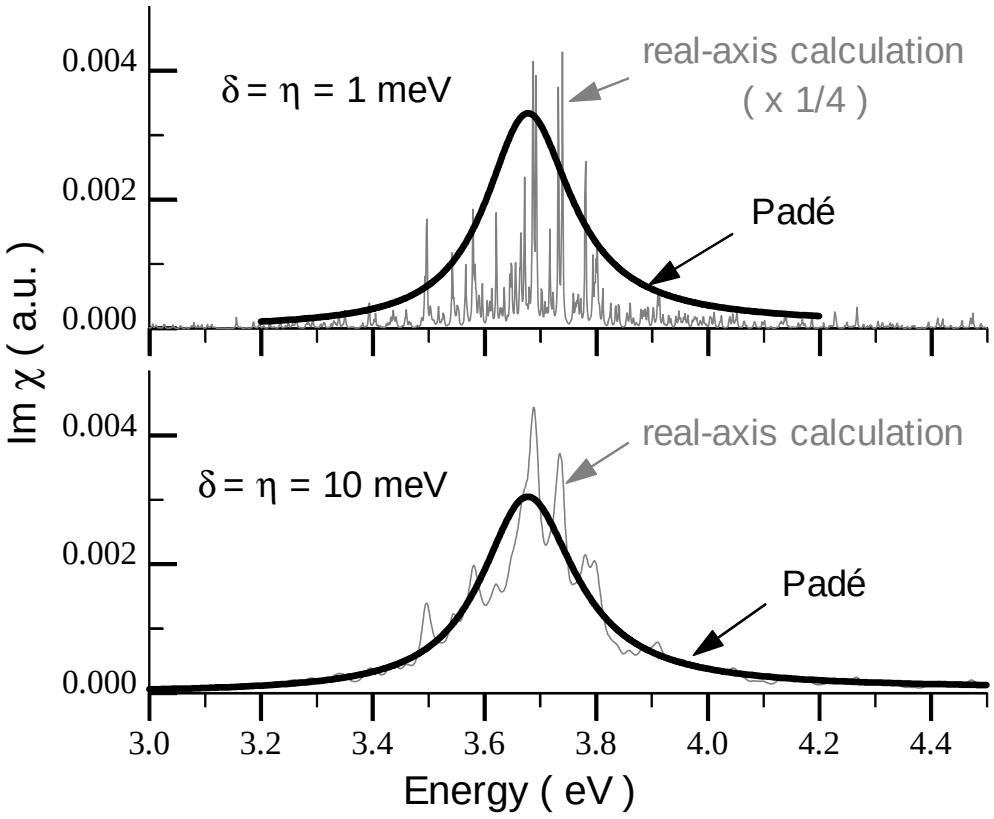


Figure 1 Imaginary Sampling Technique : loss function for K

Loss function for K for a small momentum transfer $\vec{q} = (1,1,0)\left(2\pi/\left(16a_0\right)\right)$, where $a_0 = 5.23\text{Å}$ is the lattice constant. Thick solid line: $\text{Im}\,\chi$ obtained from Eqs. (8) and (10) for imaginary ω's, followed by analytic continuation to a distance δ above the real-ω axis. Thin solid line: $\text{Im}\,\chi$ obtained directly for real ω's, using a broadening parameter η in Eq. (10). Top (bottom) panel corresponds to $\delta = \eta = 1\text{meV}$ (10 meV).

accurate description of the lifetime of this collective excitation. DuBois and Kivelson [19] first ruled out electron correlation as the leading mechanism by carefully considering the cancellation between contributions from different terms in the perturbation expansions. Later, Kloos [20] found strong momentum dependence of plasmon linewidth, ~ inverse lifetime, in a series of simple metals, which was later confirmed by Gibbons *et al*. [21]. This interesting momentum dependence was partially explained, for the case of aluminum and lithium, by Sturm as being controlled by the strength of the pseudopotential corresponding to periodic crystal [22,23]. However, in the case of potassium and sodium, the theoretical prediction qualitatively differs from the experimental data.

The failure of the previous theory on plasmon linewidth dispersion raised more attention by the more recent landmark electron energy-loss spectroscopy (EELS) measurements by vom Felde, Sprösser-Prou, and Fink [24,25], whose data revealed remarkable anomalies [26] in the excitation spectrum of the alkali metals. Consistent with earlier measurement, the dispersion of the plasmon linewidth in potassium (and to a lesser degree in sodium) was found to differ qualitatively from theoretical predictions obtained on the premise that the effects of the one-electron band structure are unimportant, except for the presence of a small gap just above the Fermi surface. On this basis, it was concluded that (*i*) the linewidth dispersion is anomalous, and (*ii*) the physics behind the anomaly must be due to strong dynamical short-range correlations [24]. Since these correlations vanish in the high-density limit $r_s \to 0$, which is far from being realized in potassium, in which the density is low ($r_s \sim 5$), conclusion (*ii*) seems

"natural," and is consistent with the absence of single-particle fine structure in the EELS data.

In this work, it is found [27] that the anomalous dispersion of the plasmon linewidth in K is, in fact, controlled by decay into single particle-hole pairs involving empty states of *d*-symmetry. By comparison with the EELS data it is concluded that dynamical many-body correlations (which we turn off), do not affect the linewidth dispersion. By contrast, the exchange-correlation effects built into the crucial "final-state" *d*-bands have a profound impact on the calculated linewidth. The novel features of the results are traced to the fact that the damping process is dominated by a subtle interplay between flat *d*-bands and the collective plasmon mode.

The aim of in this work is to disentangle the effects of the one-particle band structure from those of dynamical many-body correlations. The TDDFT linear-response framework is well suited for this purpose, as it allows one to turn off the latter by setting $f_{xc} = 0$ and thus to elucidate the impact of single-particle decay channels as discussed below.

2.2.1 Key Decay Mechanism

Shown in Figure 2 is the result of a well-converged calculation of the plasmon linewidth dispersion of K (solid circles), obtained using a 20x20x20 $\vec{k}$ -mesh and an energy cutoff of 20 eV in Eq. (10), corresponding to the inclusion of ~ 20 valence bands plus core states. This result is compared with the EELS data of vom Felde et al. [24] (empty

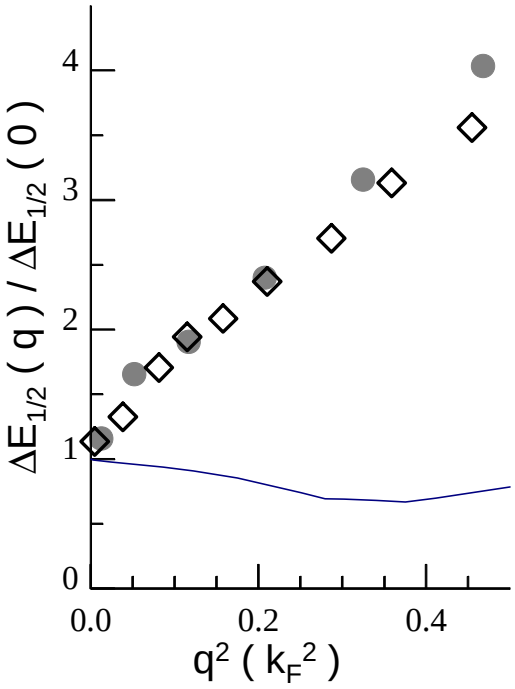


Figure 2 Plasmon Linewidth Dispersion for K

Comparison of our theoretical results with the EELS data of Ref. [25] (diamonds), and the theoretical results of Ref. [23] (solid line). Theory is for (1,1,0) propagation; the EELS data are for polycrystalline K.

diamonds). Note that the full-width at half-maximum of the plasmon peak, $\Delta E_{1/2}(\vec{q})$, is given relative to its extrapolated value for $\vec{q}=0$, a convention also adopted by previous authors [23,24]. Clearly, the calculated results are in excellent agreement with experiment; since this agreement is obtained for $f_{xc}=0$, the plasmon linewidth dispersion of K is not controlled by a dynamical many-body mechanism[*+].

The result of Figure 2 is striking, as "intuitive" expectations based on the fact that the gap just above the Fermi surface at the *N*-point is small[#] yield the result shown by the solid line in Figure 2, which corresponds to an evaluation of the dielectric function to second order in an empirical pseudopotential [23]. It is apparent that the use of nearly free electron states and their eigenvalues in the evaluation of χ^{KS} breaks down completely. Note that the fact that this simple-model result is so far off the EELS data is precisely what led to the proposal of the importance of dynamical correlations [24].

---

[*] This conclusion does not imply that the dynamical correlations vanish identically, but that any effects of f_{xc} would have little impact on the *dispersion* of the linewidth.

[+] If the calculated results of Fig. 2 are compared with the EELS data on *an absolute scale*, and the difference is imputed to many-body correlations, the latter effect may account, at most, for 20% of the plasmon linewidth; this is an *upper bound*, as other damping processes are present in the experiment (e.g., boundary scattering).

[#] The existence of this small *N-point* gap (cf. inset of Fig. 3) is ultimately responsible for the Umklapp process which leads to a finite plasmon width at zero wave vector.

Analyzed in Figure 3 is the role of key "final state" bands entering the evaluation of the single-particle response χ^{KS}; the relevant part of the band structure is shown in the inset of Figure 3 (left panel), in which the shaded strip is the ω-interval representing all the single-particle states which may couple to the plasmon (as determined by the conservation laws of energy and crystal momentum)[*]. Now *three* valence bands (thin solid lines) are needed in order to obtain a good dispersion curve for the plasmon *energy*. Keeping just these bands, the resulting plasmon linewidth dispersion curve (solid triangles) agrees well with the result of Ref. [23], which is understandable, as the states kept are, for the most part, nearly-free-electron-like. Also shown is the result (solid squares) which incorporates the contribution from three additional bands (thick solid lines in the inset). The inclusion of these bands brings about a *qualitative* change in the plasmon linewidth dispersion curve, which is now quite close to the EELS data[+]. *Clearly, these three bands provide the key decay channels for the plasmon of K.* It is

[*] The lower edge of the shaded strip is given by (bottom of occupied band $+\omega_p(0)$); the upper edge is ($E_F + \omega_p(q_c)$), where q_c is the wave vector for which the plasmon enters the continuum of particle-hole pair excitations; $\omega_p(0)$=3.7 eV.

[+] In Figure 3, the contribution to χ^{KS} from the semi-core states is excluded, since these are absent in the model calculation. On an absolute scale, the converged calculation of Figure 2 differs from the six-band calculation of Figure 3 mainly because of the effect of these states.

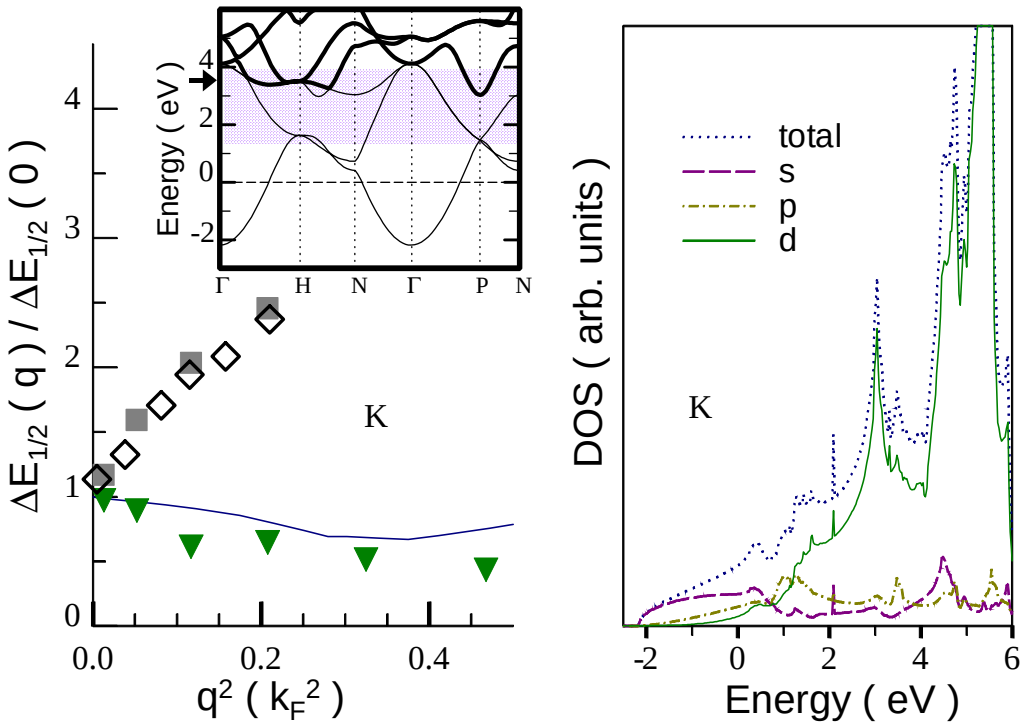


Figure 3 Key Decay Mechanism of the Plasmon in K

Left panel: Plasmon linewidth dispersion obtained upon keeping 3 (triangles) and 6 (squares) valence bands in χ^{KS} (the contribution to χ^{KS} from the semi-core states is excluded, since these are absent in the results of Ref. [23]. On an absolute scale, the converged calculation of Figure 2 differs from this 6-band calculation mainly because of the effect of these states.). Inset: LDA band structure of K; the arrow indicates the value of $\omega_p(0)$. Right panel: Calculated DOS for K —total DOS and contributions from states of *s*, *p*, and *d*- symmetry [9]; the zero of energy is the Fermi level.

significant that the bands in question are overwhelmingly of *d*-character, as evidenced by the angular momentum-resolved density of states (DOS) shown on the right panel of Figure 3.

The damping channels just identified yield a *positive* plasmon linewidth dispersion. This point can be visualized with the aid of the qualitative result that [23]

$$\Delta E_{1/2}(\vec{q}) \approx \mathrm{Im}\,\varepsilon(\vec{q};\omega)\left(\partial\,\mathrm{Re}\,\varepsilon(\vec{q};\omega)/\partial\omega\right)^{-1}\Big|_{\omega=\omega_p(\vec{q})}, \tag{11}$$

where ε is the dielectric function defined above. The first factor in Eq. (11) represents the availability of channels for decay into particle-hole pairs with energy $\hbar\omega_p(\vec{q})$ (given by the position of the peak in $\mathrm{Im}\,\chi(\vec{q};\omega)$); the second represents a "polarization effect" from interband transitions taking place at other energy. Then, since $\omega_p(\vec{q})$ disperses upwards, a larger portion of the flat bands in the shaded strip becomes available to the plasmon as $|\vec{q}|$ increases, which naturally leads to an enhancement of the first factor in Eq. (11) — thus the positive linewidth dispersion, in agreement with experiment.

2.2.2 Role of the "Static" Exchange Correlation Effect in V_{xc} within LDA

In the above analysis, dynamical many-body correlations are concluded unimportant in the present problem. However, there *is* a subtle many-body effect built into the result of Figure 2; the same has to do with the exchange-correlation effects included in the Kohn-Sham response χ^{KS}. It is worth emphasizing that, although the single-particle

states entering Eq. (10) —in particular, the crucial *d*-bands of interest here— do not have the meaning of quasiparticle states, Eq. (10) is *rigorous*, in the TDDFT linear-response framework [8]; the only approximation that has been made in the evaluation of χ^{KS} is the LDA.

Given in Figure 4 is the result from additional calculations within the random-phase approximation (RPA), in which *all* exchange-correlation effects are left out, including those in the band structure —which now corresponds to the Hartree approximation. The key physical change is that the flat Hartree *d*-bands are shifted upwards in relation to the LDA *d*-bands of Figure 3; the former lie almost entirely *above* the shaded strip in Figure 4. As a result, the plasmon linewidth dispersion curve acquires a much smaller slope, which differs significantly from experiment. This finding places the result of Figure 2 in an even more interesting perspective, since it means that exchange-correlation effects contained in the one-particle-like Kohn-Sham states of *d*-symmetry play a non-trivial role —via χ^{KS}— in the explanation of the plasmon damping mechanism.

2.2.3 Role of the Shallow "Semi-Core" States

Finally, it is worth noting that K has a polarizable core, due to the existence of shallow 3*p* "semi-core" states with binding energy of 16 eV, which impacts the damping process via the second factor in Eq. (11). A semi-analytical technique to treat such polarizable

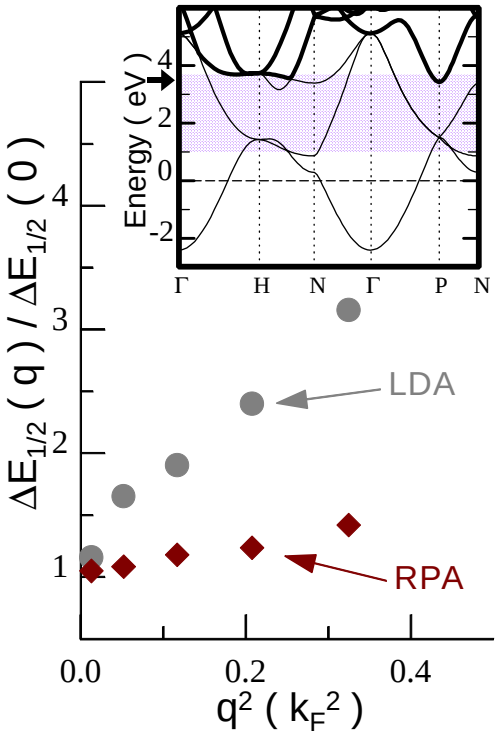


Figure 4 Role of Exchange-Correlation Effect in V_{xc} within LDA

Plasmon linewidth dispersion obtained on the basis of RPA (diamonds) and LDA (circles) single-particle response χ^{KS} and respective band structures. The inset shows Hartree bands used in the RPA case. Notice the impact of exchange-correlation effects on the location of the *d*-bands, compared with the one in Figure 3.

core has been given by Sturm *et al*. [28]. As mentioned above, the "local-orbital" extension of the LAPW method allows treating the contribution to χ^{KS} from semi-core states *on the same footing* with that from the valence states.

The circles (squares) in Figure 5 are the result obtained in the presence (absence) of this contribution to χ^{KS}. It is apparent that the absolute value of the linewidth is affected markedly by the availability of the core excitations. In fact, their effect on *the slope* of the linewidth dispersion curve is even larger than that of the many-particle correlations. Of course, this effect was included in the central result of Figure 2.

The effect of the semi-core states is through their participation of polarization processes. Even though the energy range of these polarization processes are higher than the plasmon energy, the corresponding long "tail" in the real part of the dielectric function is broadly extended and thus contributes to the linewidth through the second part of Eq. (11).

2.2.4 Summary

In summary, the "anomalous" dispersion of the linewidth of the K plasmon has been explained via *ab initio* TDDFT-based calculations. The key mechanism is shown to be decay into particle-hole pairs involving empty states of *d*-symmetry. The results highlight the role of the exchange-correlation effects built into the LDA single-particle band structure —which is closely related to the fact that the final-state bands are flat.

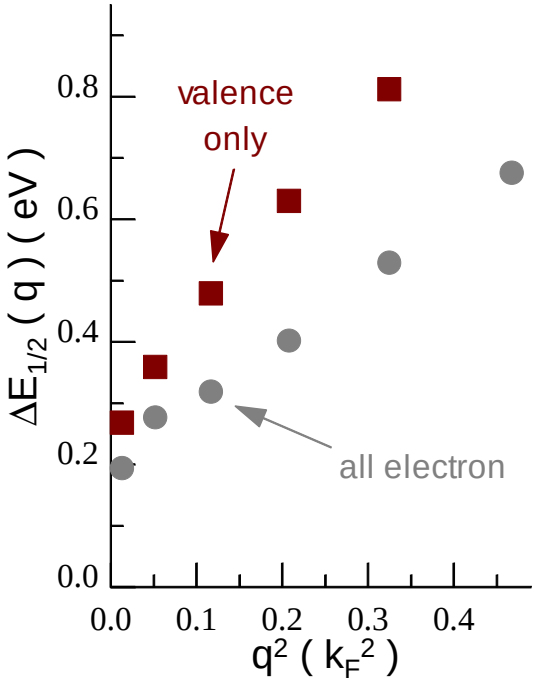


Figure 5 Role of the Semi-Core States: plasmon linewidth dispersion for K

Calculated plasmon linewidth dispersion in the presence (circles) or absence (squares) of the core contribution to the Kohn-Sham response χ^{KS} .The interesting non-negligible role of the semi-core states is also illustrated. The phase-space complexity introduced by these bands in the plasmon damping process renders simple models inapplicable.

The result in this work is made possible with the employment of imaginary frequency sampling technique, followed by analytical continuation via Padé approximants. This technique allows resolving the very narrow line shape in the case of Na, corresponding to the long lifetime of the plasmon mode, with a reasonably coarse mesh through a very small numerical broaden parameter.

2.3 Electron-Hole Excitation near the *d*-band threshold in Zinc

Fink's group has reported recently a series of electron energy-loss spectroscopy (EELS) measurements for post-transition metals Zn [29] and Cd [30], for which the threshold for excitation of electron-hole pairs involving hole states of *d*-like symmetry lies close to the nominal plasmon energy. Carefully examined in this chapter is the case of Zn, whose measured loss function is characterized by an intriguing line shape, which was interpreted [29] on the basis of a Lorentz-Drude model consisting of the standard intra-band Lindhard contribution (corresponding to the *sp*-electrons) and inter-band contribution from the shallow core states (corresponding to the *d*-electrons) that hybridizes with the former for large wave vectors. The unsuccessful model description, in the current literature, of the momentum dependence of the strength of features in the experiment spectra calls further theoretical support, and thus motivates this work.

Conceptually, Zn is a very interesting system, since, as the experiments of Widder *et al*. [29] strongly suggest, the *d*-like electrons —which reside in a band manifold whose width is ~ 1 eV— control the overall shape of the spectrum. Thus, this system is

a good realization of "narrow-band physics." Anticipating one of the main conclusions of this work, it will be shown that the true picture of the excitations, while determined by the availability of the *d*-electrons, is quite different from the picture presented in the original experimental article.

2.3.1 Breakdown of the LDA in a Dynamical Context

On the theory side, most of the modern calculations of dynamical response in extended systems, including the one discussed in the first section, refer to simple metals [12,14] and semiconductors [15] for which the pair-excitation process involves a hole in a broad *sp* band. The electronic structure of those materials is well described by the local-density approximation (LDA) [7]. Furthermore, the Kohn-Sham response, χ^{KS}, calculated on the basis of LDA Kohn-Sham eigenenergy and states, turns out to describe the electron-hole response quite well; in addition, a sophisticated description of the many-body kernel f_{xc} is not required. Thus, broad-band materials are systems for which the dynamical linear response can be calculated from "first-principles."

Recently, density-response calculations have been reported for noble metals Cu [31] and Ag [32], and for transition metal Ni [33]. For these systems the LDA-based response function also turns out to agree with experiment rather well —at least, in the description of "plasmonic" excitations of rather high energies (25-40 eV range) and relatively short lifetime. This is the case even in NiO [34]; of course, for this strongly correlated material the LDA-based response is meaningless for excitations energies comparable with the insulating gap (4 eV).

By contrast, in the case of Zn it is found in the present work that the LDA based dynamical linear response fails *qualitatively*. This failure is mainly attributed to the improper *d*-band energy location. (Related problems with the LDA have been encountered in the recent literature [35] in the case of semiconducting compounds such as CdTe, stemming from the presence of cadmium 4*d* electrons.) This problem is not surprising, in the sense that spurious self-interactions built into the LDA are known to affect the description of localized states. What *is* remarkable is the enormous impact of this limitation of the LDA on the overall shape of the calculated loss spectrum. The proximity of the *d*-threshold to the nominal plasmon energy, and the fact that the former lies *below* the latter, magnifies this failure of the LDA in the context of the electron dynamics in Zn.

Shown in Figure 6 is the comparison between the EELS data [30] and theoretical loss function of Zn, obtained as outlined in section 2.1. The calculation refers to a small wave vector along the (100) direction; the data, taken for a polycrystalline sample, refer to a comparable wave vector transfer.

Consider the case of the LDA-based response calculation, noted as χ^{LDA}, in which the Kohn-Sham density-response function is obtained on the basis of LDA band structure and wave functions, corresponding to the popular Perdew and Wang parameterization [8] and its associated f_{xc}. (This implementation is also called adiabatic LDA [36] or TDLDA.) It is apparent that the LDA-based loss function shown

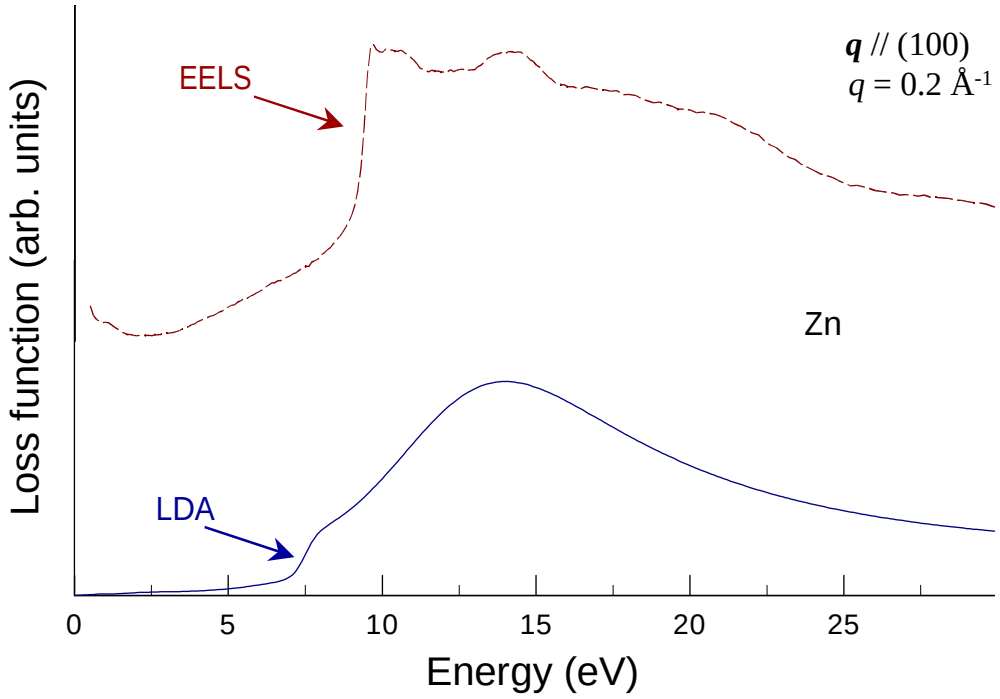


Figure 6 Failure of the LDA Based Dynamical Response: loss function of Zn

Comparison of the calculated loss spectrum of Zn with the EELS data of Ref. [29]. The theoretical spectra refer to LDA-based (solid line) loss function. Notice the weak "shoulder" in the LDA result, contrary to the strong onset measured in EELS.

in Figure 6 is in complete disagreement with observation. The dramatic 9.4 eV onset which dominates the experimental line shape is almost entirely absent. Not only it occurs ~ 2 eV below the measured threshold, but also the strength of the calculated onset is far too small —it amounts to a weak "shoulder". In fact, the loss function for a similar momentum transfer along the (001) direction (not shown) bears no hint of an onset!

The reason for this qualitative failure of the LDA is quickly surmised with reference to Figure 7, which shows the all-electron band structure for hcp Zn, obtained in the LDA. While at the Fermi level the bands are predominantly broad *sp*-like bands, a *d*-like manifold is quite apparent toward the bottom of the valence band. The threshold for this manifold is at ~ 7 eV below the Fermi energy, while *the EELS data suggest that it should lie appreciably lower*, at ~ 9.4 eV. From a physics standpoint, this problem is (at least in part) a reflection of the spurious self-interaction [37], which is not properly canceled out in the LDA treatment of exchange. Note that this problem does not affect the *ground-state* properties of Zn, which are described rather accurately in the LDA [38]. By contrast, Figure 6 demonstrates a spectacular failure of the LDA in a *dynamical* context.

A fundamental resolution to this problem clearly calls for the construction of a band structure that is at least self-interaction free. Techniques for achieving this have been proposed recently, in the context of the optimized-potential method [39]. However, in the present case of a narrow *d*-band it is an open question how to supply

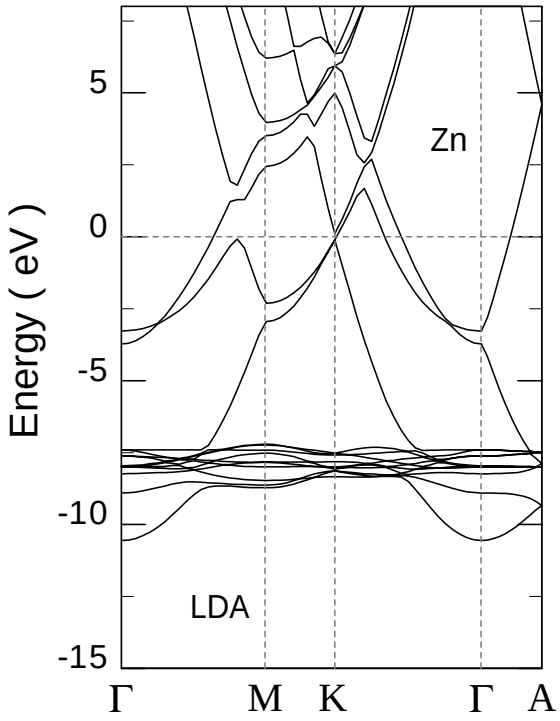


Figure 7 LDA Band structure of Zn

LDA band structure of Zn calculated within the all-electron full potential LAPW method [9].

an exact-exchange calculation [40] with an appropriate (approximate) treatment of correlation. Thus, in this work a more phenomenological approach is adopted, which leads to a new interpretation of the physics underlying the loss function near the *d*-electron threshold.

2.3.2 Phenomenological Approach

In this work, a phenomenological band structure is produced with Slater's $X\alpha$ approximation [41], which differs from its LDA counterpart by an overall downward shift of the *d*-like bands of about 2 eV. The ensuing loss function mimics the observed threshold behavior quite well. This result highlights the impact of *d*-band energy location on the loss spectrum of Zn. Furthermore, the result reveals that the threshold line shape turns out to be a result of subtle interplay between two coherent modes. This picture is very different from what was envisioned in the original article [29].

The phenomenological band structure employed in this work is obtained through Slater's $X\alpha$ approximation, in which $V_{xc}[n](\vec{x}) = (3/2)\alpha V_x^{LDA}[n](\vec{x})$. The motivation for this choice is that, for $\alpha > 2/3$, V_{xc} becomes stronger (more negative); in the corresponding self-consistent calculation the *d*-like states tend to become more localized, and this increases the exchange energy (its absolute value) and the binding energy of the *d*-manifold. It is important to note that the binding energy of the delocalized *sp* states is largely unaffected, in relation to the Fermi energy; in particular, there is no discernible change in the shape of the Fermi surface.

The $X\alpha$ band structure corresponding to $\alpha = 0.9$ is shown in the right panel of Figure 8. This "optimal" choice of the parameter α was arrived at upon performing the corresponding evaluation of the response function, noted as $\chi^{X\alpha}$, and an optimization of the calculated loss function vs. the EELS data, shown in Figure 9. It is apparent that the $X\alpha$-based spectrum features a sharp onset, and the same mimics the observed line shape very well.

It needs to be emphasized that the direct interplay between band structure —in particular, the position of the *d*-electron threshold— and the EELS data implicit in the above "optimization" of the $X\alpha$-based response is "legitimate," in the sense that the Kohn-Sham eigenvalues and wave functions enter *rigorously* in the non-interacting *electron-hole* response χ^{KS} given by Eq. (10). This situation is to be contrasted with the case of the quasi-particle excitations measured in photoemission (detail discussion in Chapter 3), in which no connection between the Kohn-Sham eigenenergy and quasiparticle energy has been established.

It should also be remarked that the above "direct mapping" between the response function and the Kohn-Sham band structure in the evaluation of the $X\alpha$-based response is based on the assumption that f_{XC} associated with $V_{xc}^{X\alpha}$ is very small compared to the Coulomb potential, which becomes singular at long wave length limit. For the momentum range of interest here, this seems to be a reasonable assumption, unless there exist some strongly frequency-dependent (at ~ 9 eV) correlation processes

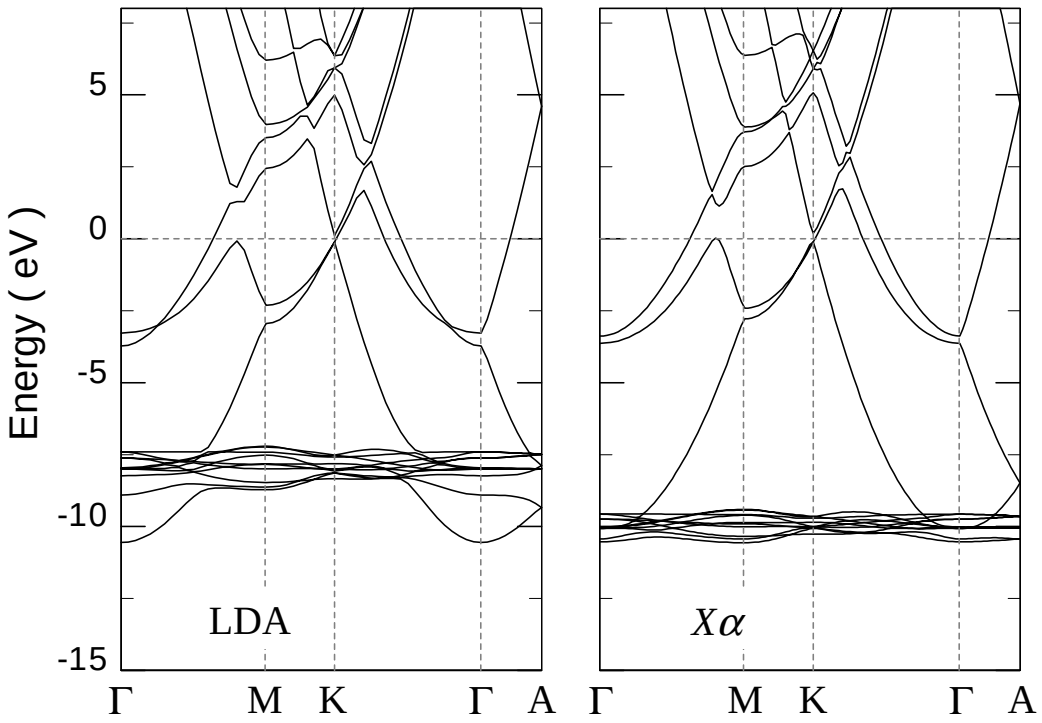


Figure 8 Phenomenological Band structure of Zn

Band structure of Zn calculated within the all-electron full potential LAPW method [9]. Left panel: LDA band structure; right panel: phenomenological $X\alpha$ band structure, corresponding to an "optimized" value of α ($\alpha = 0.9$) which yields the loss function shown in Figure 9

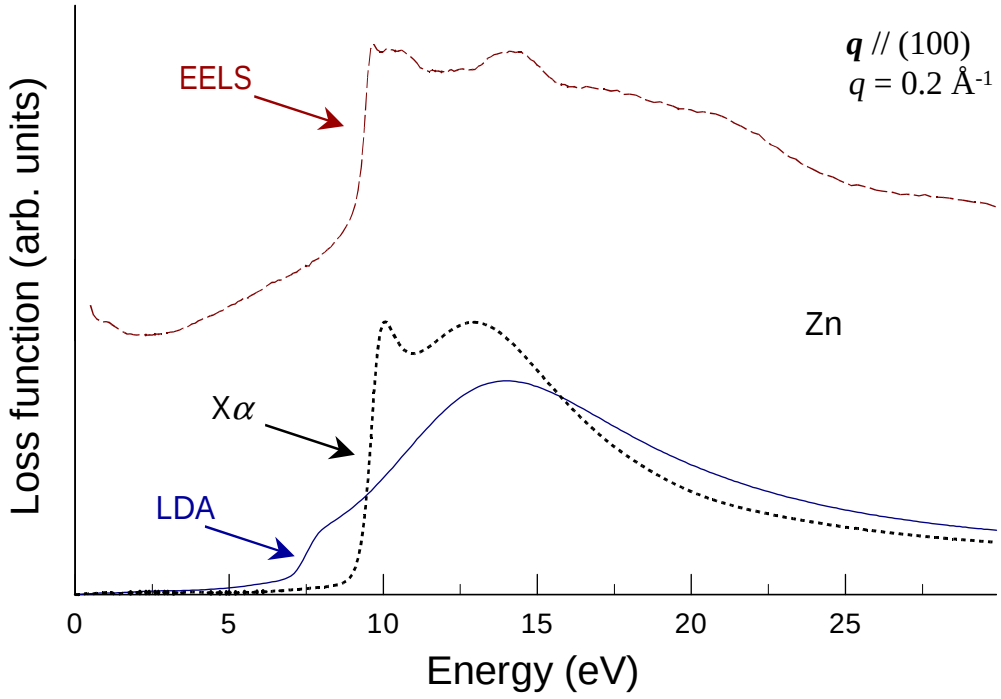


Figure 9 Phenomenological Simulation of Dynamical Response in Zn

Comparison of the calculated loss spectrum of Zn with the EELS data of Ref. [29]. The theoretical spectra refer to LDA-based (solid line) and $X\alpha$-based (short dashes) loss function. Notice the strong onset in the case of $X\alpha$ result.

involving the electrons in the flat *d*-bands of Zn that are large enough to compete with the Coulomb potential. To this author's knowledge, there is no existing model of the many-body kernel f_{XC} that contains this type of physics.

2.3.3 Physical Interpretation of the Threshold Excitation

Based on the phenomenological response function, it is now possible to study the microscopic mechanisms. Plotted in Figure 10 re the calculated ε^{LDA} and $\varepsilon^{X\alpha}$, through Eq. (5), which correspond to χ^{KS} and $\chi^{X\alpha}$ defined above. It can be quickly visualized that the sharp onset observed at ~ 9.4 eV in the loss function of χ^{KS} correlates directly with the small value of $\varepsilon^{X\alpha}$ (upper panel) which realizes in the neighborhood of that energy.

The physical origin of this small value of $\varepsilon^{X\alpha}$ near threshold is twofold: First, there is a significant modulation of the dielectric function brought about by *d*-electron excitation (Figure 10); second, the close proximity of the nominal plasmon energy (the actual zero of $\mathrm{Re}\,\varepsilon^{X\alpha}$), and the fact that this energy occurs *above* the threshold, has the effect that the modulation comes close to producing a new zero of $\varepsilon^{X\alpha}$ for an energy in the neighborhood of the threshold. Thus, the onset in the observed loss function is *coherent in nature*.

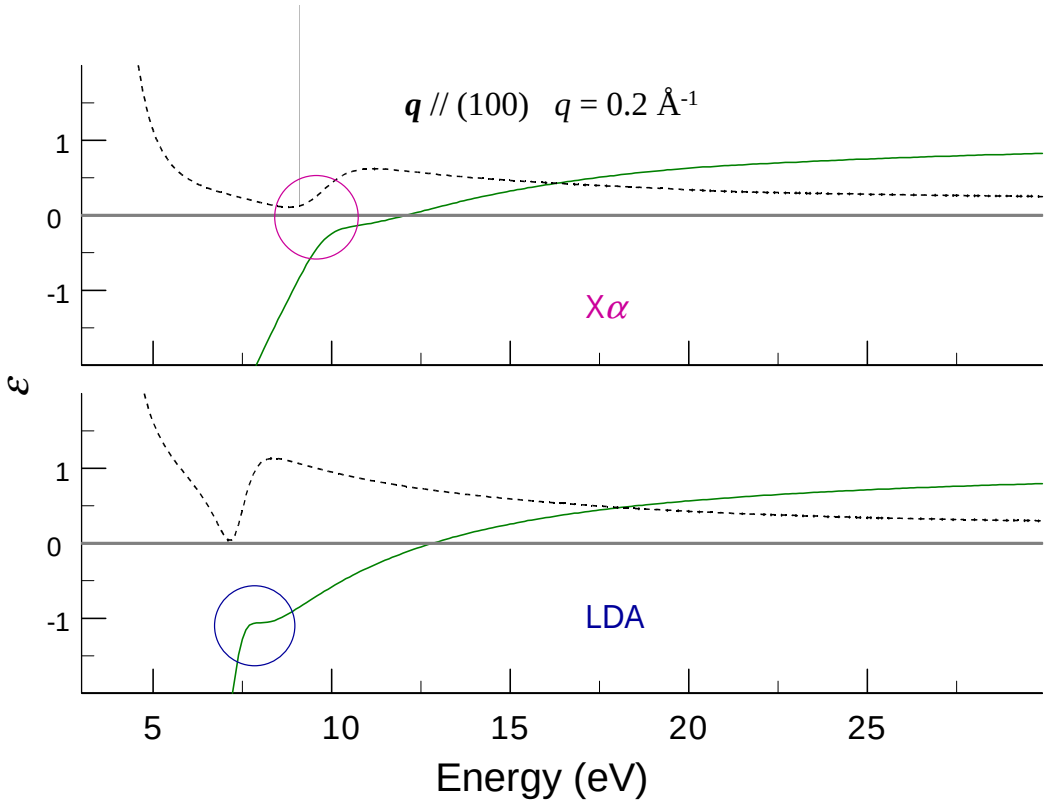


Figure 10 Collective Nature of the Threshold onset: dielectric function of Zn

Comparison of $X\alpha$ and LDA dielectric function. Physical interpretation of the sharp onset of the *d*-electron excitation threshold, which dominates the loss function of Zn (Figure 9), is made based on calculated dielectric functions. The figure shows the real part (solid line) and imaginary part (dashes) of the dielectric function correspond to $X\alpha$ (upper panel) and LDA (lower panel). The circles drawn in each panel highlight the modulation of the dielectric function due to excitation of electron-hole pairs near the *d*-electron threshold.

By contrast, in the case of ε^{LDA} (Figure 10, lower panel), the top of the *d*-band manifold lies ~ 2 eV lower than the measured onset; thus, it is sufficiently removed in energy from the over-damped plasmon that $\mathrm{Re}\,\varepsilon^{LDA}$ is too large (too negative) near the LDA *d*-threshold; thus, in this case the modulation of the dielectric function does not bring this function close to a zero value — thereby the weak shoulder which is (at best) observed in the loss function obtained from LDA (Figure 7).

One may ask why, if the *d*-electron-related onset is coherent in nature, the same does not manifest itself as a "real peak" in the loss spectrum —i.e., a peak to which one could assign a linewidth; cf. Figure 9. The answer to this question is found in the very nature of the Kohn-Sham states involved in the phenomenon at stake: they are of *d*-character, and they lie in a very narrow band slightly above the energy of the coherent mode, where they have a large density of (initial) states. Thus, as the energy transfer goes through the threshold, there is a dramatic "onset" in the excitation probability — which is reflected in the rapid raise in $\mathrm{Im}\,\varepsilon^{X\alpha}$ above the threshold, as observed in Figure 10. As a consequence, there is no actual peak in the loss function associated with *d*-electron excitation, but rather a sharp edge on the low-energy side of the threshold-related line shape.

The microscopic picture we have just developed is very different from the one invoked (with a Lorentz-Drude model) in the analysis of the experiments of Ref. [29], in which the onset was regarded as single particle excitation from the *d*-bands to states above the Fermi energy. Furthermore, based on that picture, those authors attempted to

incorporate the experimental momentum dependence of the strength of the first features in the spectra (shown in Figure 11) by reducing the "oscillator strength" in the model as momentum increase, while keeping constant the plasmon energy that corresponds to the second feature. This picture did, to a certain degree, reduce the strength of the first feature, with a huge price of unavoidably decreasing the energy of the second feature, which is against the experimental results. It is exactly this failure of the model description that calls for a better theoretical support.

In the present work, it is found that the "oscillator strength" defined in the original experimental analysis [29] depends very weakly on the momentum transfer for energy transfers near threshold. This can be easily concluded by observing the weak *q*-dependence of $\max\{\mathrm{Im}\,\varepsilon\}$ in the energy range corresponding to the threshold excitation in Figure 12. Thus, the original picture is not supported by results of this work. By contrast, the weakening of the onset with increasing momentum transfer is a natural consequence of the mechanism that explained the existence of the sharp onset for small wave vectors. Indeed, as the momentum transfer increases, the nominal plasmon disperses upwards, bringing about an increase in the energy separation between the *d*-threshold and the plasmon. As this energy separation increases, the *d*-electron-induced near-zero of $\mathrm{Re}\,\varepsilon$ quickly disappears — thereby washing up the onset.

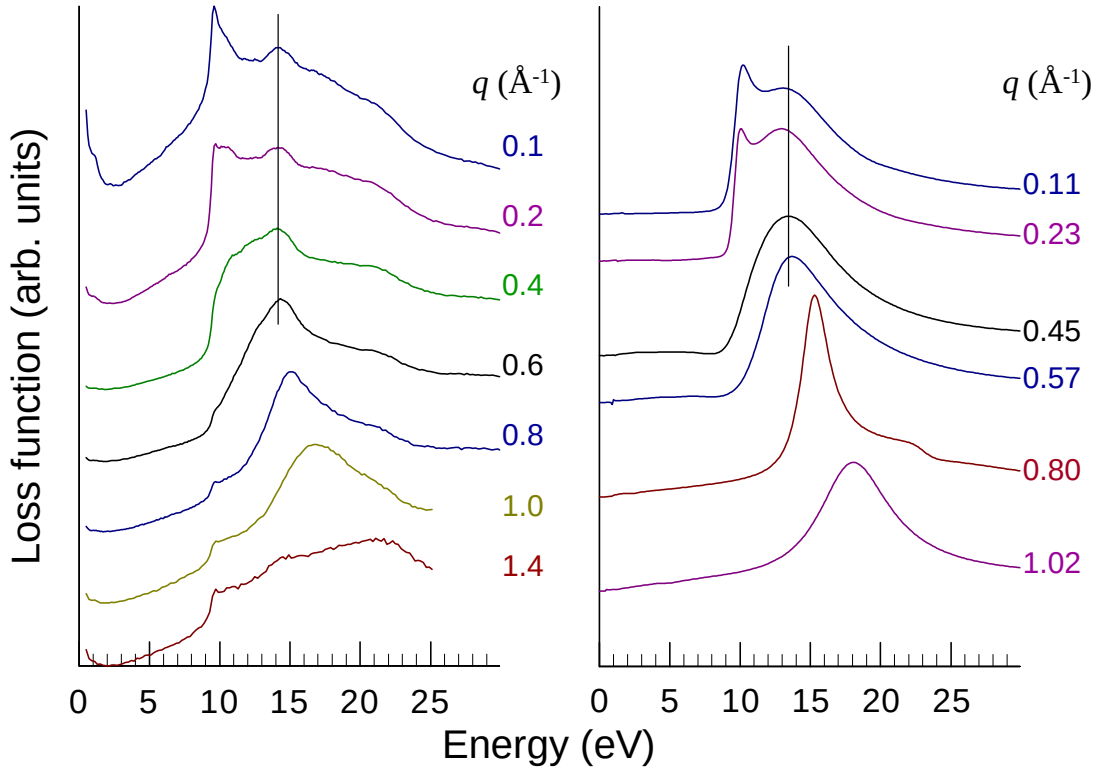


Figure 11 Momentum Dependence of Features in the Loss Functions of Zn

Comparison of experimental (left panel) and theoretical (right panel) loss functions. The strong momentum dependence of the strength of the first feature in the experimental spectra is well reproduced by the phenomenological calculation (in which the "oscillation strength" is found to remains ~ constant. See Figure 12)

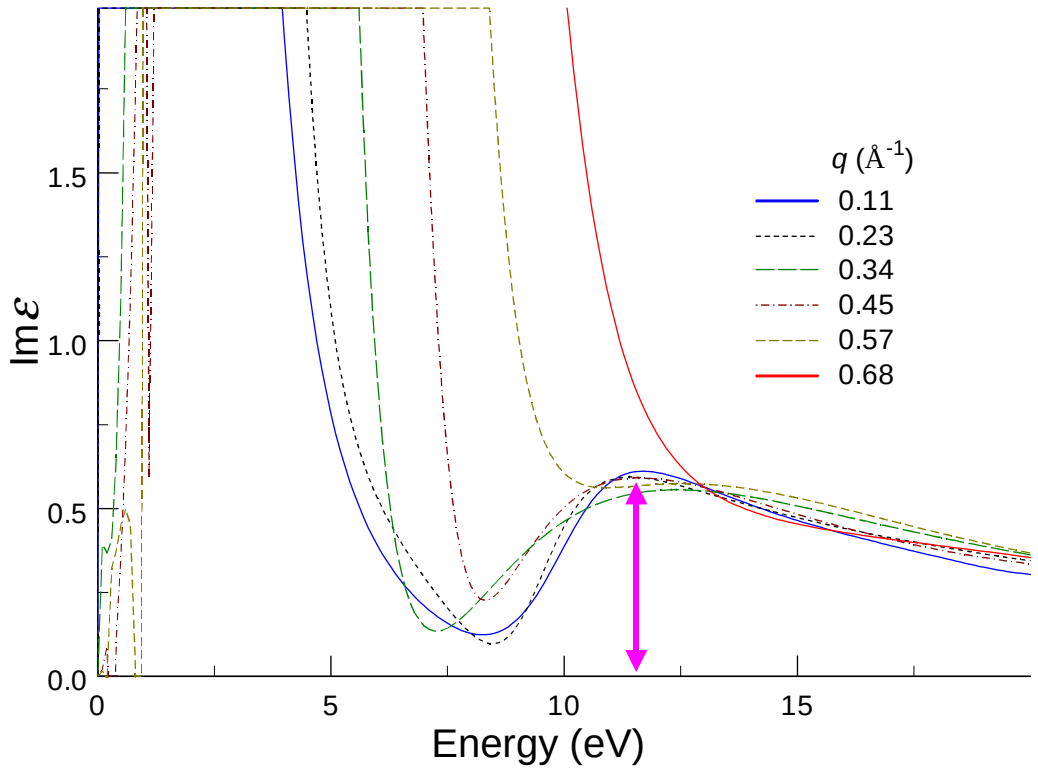


Figure 12 Weak Momentum Dependence of the "Oscillator Strength" of Zn

Comparison of imaginary part of dielectric functions for increasing momentum transfer. The very weak momentum transfer dependence of the "oscillator strength" of the first feature in the calculation can be easily observed through the weak *q*-dependence of the height (indicated by the arrow) of the *d*-threshold structure around 12 eV.

2.3.4 Summary

In summary, analysis of the present work has shown that a direct implementation of the standard LDA fails qualitatively to describe the dynamical response of Zn. This failure is closely related to the improper energy of the shallow Kohn-Sham 3*d* states, partially due to the inappropriate self-interaction build into the employed LDA functional. A phenomenological approach, using Slater's $X\alpha$ treatment to lower the shallow semi-core states, brings a qualitative change to the spectra and recovers the strong onset observed in the experiment. This illustrates the sensitivity of the density response of Zn to the underlying Kohn-Sham band structure.

Based on this analysis, contrary to what was originally proposed [29], it is found that the onset is actually of coherent nature, whose source is the approximate zero in $\mathrm{Re}\,\varepsilon$ due to the *d*-transitions close to the plasmon energy. The finding naturally solves the puzzle posed by the original authors —the decrease of strength of the onset as momentum transfer increases— which is a result of complex interplay between the plasmon mode and the semi-core related collective mode. The underlying "oscillator strength," however, possesses only weak *q*-dependence.

Finally, a more fundamental treatment of a better $V_{xc}[n]$ functional remains a challenge. Such V_{xc} needs at least to be self-interaction free. Furthermore, a proper inclusion in V_{xc} of physics involving the localized Kohn-Sham *d* states is necessary in

order to give accurate binding energy, which is crucial to the dynamical response studied here.

Chapter 3 Quasi-Particle Excitation

Besides the dynamical linear response discussed in the previous chapter, another widely studied electronic excitation is quasi-particle excitation, which is usually measured through angular resolved photoemission spectroscopy (ARPES) or the inverse photoemission experiment (see 3.1). While many-body perturbation theory, in which the concept of quasi-particle (QP) is introduced, has been maturely developed, the actual *ab initio* implementations of the theory to realistic crystal systems are rather limited, mainly due to the limitation in the computation resources.

In this chapter, the recent debate [42,43] around the occupied QP bandwidth will be discussed to illustrate the author's viewpoint on how to properly assign many-body mechanisms. Furthermore, the recently revived concern on the QP band gap of silicon will be focused based on a newly developed implementation of Baym-Kadanoff (BK) scheme of the finite temperature many-body perturbation theory (MBPT). Finally, the QP band gap of germanium will be re-examined with this new implementation, and the widely accepted core polarization potential (CPP) treatment will be questioned.

3.1 Finite Temperature Many-Body Perturbation Theory

The finite temperature many-body perturbation theory has been maturely developed. A complete introduction can be found in Ref. [45-48]. Outlined in this section are only the relevant formulas.

In the framework of MBPT, the properties of the many-particle system are obtained through the dressed one-particle propagator, G, and two-particle propagator, G_2, defined as

$$G(1,1') \equiv -\left\langle \hat{\mathrm{T}}\, \hat{\Psi}(1) \hat{\Psi}^{\dagger}(1') \right\rangle \tag{12}$$

and

$$G_2(1,2;1',2') \equiv -\left\langle \hat{\mathrm{T}}\, \hat{\Psi}(1) \hat{\Psi}(2) \hat{\Psi}^{\dagger}(2') \hat{\Psi}^{\dagger}(1') \right\rangle \quad , \tag{13}$$

where $\hat{\mathrm{T}}$ is the "time-ordering operator", $\hat{\Psi}(1)$ and $\hat{\Psi}^{\dagger}(1')$ are the second quantization annihilation and creation operator respectively, $\langle\ \rangle$ denotes thermal average over the grand canonical ensemble, and 1 is a simplified notation for the four-dimensional variable $(\vec{x}_1, \tau_1)$, where $0 \le \tau_1 \le \beta\hbar$ is imaginary time (called "Matsubara time") with inverse temperature β. The one-particle propagator, $G(1,1')$, represents the probability amplitude of the virtual process of injecting one particle into the system at position $\vec{x}_{1'}$ and imaginary time $\tau_{1'}$, and then removing one particle from the system at position $\vec{x}_1$ and imaginary time τ_1 so that the original thermal state of the system is recovered. Thus, it contains information useful for describing physical processes that involve one particle. Similarly, the two-particle propagator, $G_2(1,2;1',2')$, is related to Physical processes that involve two particles, e.g.: particle-hole, or hole-hole interaction. Note that the introduction of imaginary time comes naturally with the grand canonical ensemble statistics for finite temperature [45-48].

In practice, the ARPES photo current spectra are commonly considered proportional to the spectra function, A, defined as

$$A(\vec{q},\omega) \equiv \frac{1}{\pi} \operatorname{Im} G(\vec{q},\omega) \quad , \tag{14}$$

where $\vec{q}$ is momentum transfer and ω is energy transfer. The derivation of this relation, which involves several approximations, is available in Ref. [49]. The term "quasi-particle" is used when the spectral function presents a well-defined mode whose lifetime is long enough (in energy space, a peak with linewidth $\ll$ its energy.) [50-53]

3.1.1 Equation of Motion: Dyson Equation

Following the derivation found in Ref. [48], the equation of motion for $G(1,1')$ is

$$\begin{aligned} \left(\hbar\frac{\partial}{\partial\tau} + \hat{h}_0\right) G(1,1';U) &= -\hbar\delta(1-1') \\ &+ \left(\mu - U(1,\bar{2})\right) G(\bar{2},1;U) + V(1,\bar{2}) G_2(1,\bar{2};1',\bar{2}^+;U) \quad , \end{aligned} \tag{15}$$

where $\hat{h}_0$ is the "non-interacting" Hamiltonian, μ is the chemical potential, V is the many-body interaction, and U is a non-local external potential, which will be set to 0 for the systems in equilibrium. Here, bars on variables imply integral over the variables. For condensed matter systems,

$$V(1,2) \equiv v(\vec{x}_1,\vec{x}_2)\delta(\tau_1 - \tau_2) \tag{16}$$

is the static Coulomb interaction.

It is convenient to define the "undressed" one-particle propagator, G_0, as

$$\left(\hbar\frac{\partial}{\partial\tau}+\hat{h}_0-\mu_0\right)G_0\left(1,1'\right)=-\hbar\delta\left(1-1'\right)\quad , \tag{17}$$

and its inverse, G_0^{-1}, as

$$G_0^{-1}\left(1,\bar{2}\right)G_0\left(\bar{2},1'\right)=\delta\left(1,1'\right)\quad . \tag{18}$$

Defining the self-energy, Σ, as

$$\Sigma\left(1,1';U\right)\equiv -V\left(1,\bar{2}\right)G_2\left(1,\bar{2};\bar{3},\bar{2}^+;U\right)G^{-1}\left(\bar{3},1';U\right)\quad , \tag{19}$$

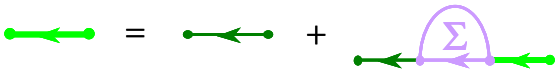
one can then rewrite Eq. (15) as

$$G_0^{-1}\left(1,\bar{2}\right)G\left(\bar{2},1';U\right)=\delta\left(1-1'\right)+\frac{1}{\hbar}\left[\Sigma\left(1,\bar{2}\right)+U\left(1,\bar{2}\right)-\left(\mu-\mu_0\right)\right]G\left(\bar{2},1';U\right) \tag{20}$$

or the famous Dyson equation:

$$G\left(1,1';U\right)=G_0\left(1,1'\right)+\frac{1}{\hbar}G_0\left(1,\bar{2}\right)\left[\Sigma\left(\bar{2},\bar{3}\right)+U\left(\bar{2},\bar{3}\right)-\left(\mu-\mu_0\right)\right]G\left(\bar{3},1';U\right). \tag{21}$$

Expressed in the diagrammatic form, the Dyson equation is



(22)

where the inclusion of $\hbar^{-1}$ factor and addition of $\left[U\left(\bar{2},\bar{3}\right)-\left(\mu-\mu_0\right)\right]$ in Σ is understood.

Notice that G_2 in the definition of self-energy in Eq. (19) can be further related to G, with the help of functional derivative of G respected to U:

$$G_2(1,2;1',2';U) = -G(1,1';U)G(2,2';U) + \hbar \frac{\delta G(1,1';U)}{\delta U(2',2)} \quad . \tag{23}$$

This relation implies that **Σ is a functional of *G***! Therefore, Eq. (19) and (21) form a set of self-consistent equations.

Up to this point, this is an exact formalism, in which the exact self-energy, Σ, should include all "skeleton diagrams" [45-48] corresponding to terms in the perturbation expansion series. In practice, only the dominating terms, corresponding to the main microscopic physical processes, are included. Identifying the important approximate self-energy becomes an important task to properly describe the system and to help gaining physical picture.

3.1.2 Baym-Kadanoff conserving scheme

One important issue was raised by Baym and Kadanoff (BK), that an arbitrary choice of self-energy in Eq. (21) may not guarantee the fulfillment of microscopic conservation laws (including particle number and energy) [54,55]. In order to ensure fulfillment of all the conservation laws, a scheme, referred to as conserving scheme, was proposed. This scheme is summarized as follows:

1. The self-energy, Σ, needs to be chosen as functional derivative of the Luttinger-Ward Φ functional [56], which is a functional of dressed one-particle propagator, G, and the bare Coulomb interaction v.

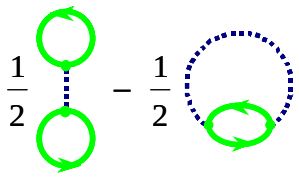
2. The solution of the Dyson equation, G, and the self-energy, Σ, that enters the same equation needs to be solved self-consistently.

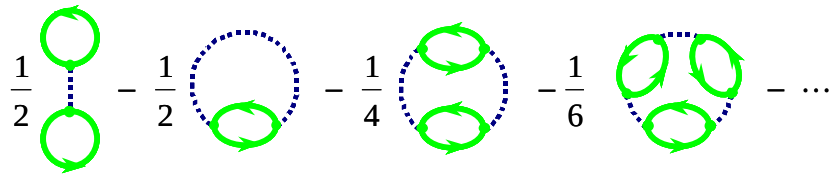
This seriousness of this issue was recently confirmed by Schindlmayr [57], who demonstrated that the popular *GW* approximation (discussed later in this section), when implemented non-self consistently, results in a genuine violation of particle number conservation, for a specially constructed model system that can be solve analytically within the *GW* approximation. Furthermore, this author showed that shifting the QP energies to mimic the effect of self-consistent result significantly reduces this violation. This finding has a large impact, since it implicitly invalidates almost all the existing *GW* calculations, which seem to give "excellent agreement" with the experiment data without further consideration of the lack of self-consistency.

The present work is based on the philosophy that only results respecting all the symmetries of the systems is physical. That is, one first needs to fulfill all the necessary conservation laws before being concerned by the comparison with experiments. In the case when the theoretical results do not agree with experiment, it is the approximate self-energy that needs to be improved, not the conserving requirement of self-consistency. Therefore, a central aspect of this work is to implement MBPT within the BK conserving scheme.

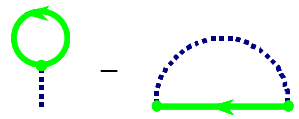
3.1.3 Hartree-Fock, Shielded Interaction, and *GW* Approximation

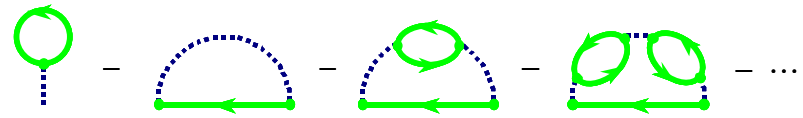
Specifically, two approximate Σ's have been implemented in this work corresponding to "Hartree-Fock" (HF) approximation and the shielded interaction approximation (SIA) [54], whose Φ functionals are expressed in diagrams as follows:

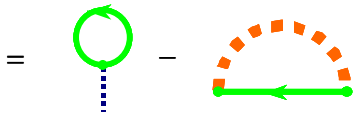
Φ^{HF} :  (24)

Φ^{SIA} :  (25)

The self-energy Σ, obtained as functional derivative of Φ, is then, in each case

Σ^{HF} :  (26)

Σ^{SIA} : 

 (27)

where

W^{SIA} :  (28)

is the shielded interaction and ..... denotes the bare Coulomb interaction. It is apparent that the SIA has the same form as Hedin's *GW* approximation (GWA) [58].

It is important to note that even though the SIA and the GWA have the same form, they are equivalent only when the GWA is solved self-consistently, which is not commonly performed in practice. In fact, almost all the existing so-called "GW calculations" involve evaluating Σ with G^{LDA}. This approach, corresponding to only the first iteration of the calculations performed in the present work, will be referred to as $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ method, non-conserving GWA, or "non-self-consistent GWA", while the fully self-consistent calculation will be referred to as conserving GWA or "self-consistent GWA".

3.1.4 Numerical Implementation

In the actual implementation, the "reference system" is chosen to be the Kohn-Sham ground state. This is made possible by adding $V^{\mathrm{LDA}} = V_H + V_{xc}^{\mathrm{LDA}}$ to both sides of equation (15) and redefining G_0 and Σ as:

$$\left(\hbar\frac{\partial}{\partial\tau} + \hat{h}_0 + V^{LDA}(\vec{x}_1) - \mu_0\right) G^{\mathrm{LDA}}(1,1') = -\hbar\delta(1-1') \quad , \tag{29}$$

and

$$\tilde{\Sigma}(1,1';U) \equiv -V(1,\bar{2})\, G_2(1,\bar{2};\bar{3},\bar{2}^{+};U)\, G^{-1}(\bar{3},1';U) - V^{LDA}(\vec{x}_1) \quad . \tag{30}$$

Thus the formalism remains the same. Solution of Eq. (29) is then given in Bloch representation ($\vec{k}$, j) as:

$$
\begin{aligned}
G^{\mathrm{LDA}}_{\vec{k},j}(\tau) &\equiv \left\langle \vec{k}, j \middle| \overline{x}_1 \right\rangle G^{\mathrm{LDA}}(\overline{x}_1, \overline{x}_{1'}; \tau = \tau_1 - \tau_{1'}) \left\langle \overline{x}_{1'} \middle| \vec{k}, j \right\rangle \\
&= -e^{-\varepsilon_{\vec{k},j}\tau/\hbar}\left(\theta(\tau) - n_F\left(\varepsilon_{\vec{k},j}\right)\right)
\end{aligned}
\tag{31}
$$

where θ is the unit step function, n_F is the Fermi occupation function, and $\varepsilon_{\vec{k},j}$ is the eigenenergy of the LDA Kohn-Sham ground state, corresponding to crystal momentum $\vec{k}$ and band number j.

It is worth noting that, instead of the commonly employed Matsubara frequency, the formalism is numerically implemented in this work on Matsubara time axis, for the first time to this author's knowledge, with newly developed uniform-power mesh, of which a detailed discussion is given in Appendices B.4 and B.5. This novel implementation allows efficient investigation into physics involving highly excited states and semi-core states, lowing the temperature of the system studied, as well as application to more complicated processes beyond GWA. All these features pose severe numerical challenge in Matsubara frequency approaches. In terms of space degree of freedom, Bloch representation is employed for G and Σ, while Fourier representation is chosen for W and v. All the integral equations are solved with "preconditioned bi-conjugate gradient stabilized" algorithm [59] (or "preconditioned conjugate gradient stabilized" algorithm in the case of symmetric operator matrices.)

Furthermore, the widely applied pseudopotential is avoided, in order to give a "clean" all electron calculation that fully represents the realistic systems studied. Conceptually, there is no reason why the use of pseudo wave functions should give the

same matrix elements, involving non-local self-energy operator, as the all-electron one, in calculations of which the famous "transferability" [60] of the unrealistic smooth pseudo wave function does not apply. This consideration turns out to be of great importance in the case of QP band gap discussed in 3.3 and 3.4.

3.2 Occupied Quasi-Particle Bandwidth of Sodium

As an example illustrating the importance of the BK conserving scheme, the recent debate [42,43] about the occupied PQ bandwidth of Na is addressed in this section.

The measured occupied QP bandwidth has been long settled to be narrower than the one in "nearly-free electron model" [61,62]. While theoretical source of this reduction of bandwidth is still under debate [63,64], it is now accepted that this bandwidth narrowing can be, at least qualitatively, described by the $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ method [65] for Na crystal. (Reminder: The $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ method is not conserving.)

In a recent article [42], with a homogeneous electron model, Yasuhara, Yoshinaga, and Higuchi (YYH) stated that the actual occupied QP bandwidth should be larger than the nearly free electron model, contrary to the earlier theories. Furthermore, YYH argued that this band widening is supported by a "correct" re-analysis of the ARPES spectra. Finally, in order to obtain the band widening, certain correlation effect beyond GWA was necessarily included, since GWA gives band narrowing instead. Thus, it was concluded in Ref. [42] that these extra correlation effects play an important role in Na.

These statements in that article are certainly intriguing, since 1) YYH implicitly invalidate almost all the ARPES analysis as they are not performed in the same way as

YYH do, and 2) YYH's results suggest that even for a system as simple as Na, these extra correlation effects are important, since they "correct" the band narrowing (from $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ approach) to band widening. While the correctness of the current connection between theory and experiment is under questioning [43], the reasoning of YYH's statement on the importance of correlation effects will be examined here, based on BK conserving scheme outlined in section 3.1.

In this work, it will be argued that from a conserving point of view, YYH's neglect of self-consistency in solving the Dyson equation brings about a substantial error in the calculation of the bandwidth, contrary to the simplifying assumption made by YYH (and many others); in fact, this error introduces clear ambiguities in the identification of the relevant correlation mechanisms.

As discussed in section 3.1.2, $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ method is not conserving. In fact, it is shown in Ref. [57] that it at least violates the conservation of particle number for a specially constructed model system. From the viewpoint taken in this work, it is necessary to first obtain "Physical" solutions that respect the microscopic conservation laws before considering agreement with experiment data.

YYH's self-energy calculation starts from the premise that the GWA [58] necessarily leads to band narrowing. However, it has been shown recently, with the homogeneous electron model [66] and the K crystal [67], that this is true only if one ignores self-consistency in solving the Dyson equation. In the present work, the calculated non-self-consistent GWA bandwidth for bcc Na reproduces the non-self-

consistent GWA result of YYH (band narrowing). However, the obtained self-consistent GWA Green's function gives a bandwidth of 3.80 eV, which is comparable with the band widening obtained by YYH, who includes certain vertex correction. Obviously, then, not only does the neglect of self-consistency lead to a qualitative error in the bandwidth determination — it also leads to an uncontrolled assignment of correlation effects. Indeed, as just demonstrated, if band widening were a goal, the same can be achieved already within the conserving GWA —if implemented fully, i.e., self-consistently.

The true picture of the impact of vertex corrections, which may emerge from a self-consistent treatment of these effects in the Dyson equation, remains to be determined. Using as a guideline the work of Hanke and Sham [68], it is surmised that the ladders for the electron-hole attraction (a form of vertex correction) should give an effect of the opposite sign as the one obtained in the GWA.

3.2.1 Summary

The recent debate on occupied QP bandwidth serves as an example of significance of BK conserving scheme on assigning relevant microscopic physical mechanisms through the identification of important self-energy diagrams against experiment. While non-conserving GWA is thought, by the community, to give "excellent agreement" in many cases studied, careful re-examination of those cases, based on BK conserving scheme, is necessary due to the lack of self-consistency in those existing calculations. The following work on re-investigation of QP band gap of prototype semiconductors, Si and

Ge, motivated by this viewpoint, clearly reinforces the philosophy and allows new microscopic interpretation and insight into the physics of these systems.

3.3 Quasi-Particle Band Gap of Silicon

The description of the QP band gap of semiconductor plays an important role in the history of the development of *ab initio* implementation of MBPT, motivated by the well-known "failure"[*] of the LDA in producing band structure with proper band gap for Si. In the case of Ge, LDA Kohn-Sham band structure does not even contain a gap. The $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ method was shown [69] to "rescue" the situation by opening the Si gap to the experimental value. Since then, the screened interaction, mechanism included in the GWA diagrams, has been regarded as the dominating processes in the physics of QP band gap of Si, based on the assumption that self-consistency has little effect due to the similarity between the "many-body wave functions" and the Kohn-Shame wave functions [69].

[*] Strictly speaking, this is not really a fair judgment of quality of the LDA, because the DFT band structure, as dispersion of eigenenergy of the Kohn-Sham states, is not meant to represent the QP band structure. This is simply a misconception of mixing two different frameworks. A detailed discussion can be found in Ref. [70]. On the other hand, once the discontinuity in V_{xc}, upon addition of an electron, is somehow incorporated, the resulting Kohn-Sham band structure does seem to give reasonable gap similar to the size of QP band gap [37,71].

This viewpoint was challenged by Schöne and Eguiluz in a recent article [67], in which a fully self-consistent (conserving) GWA result was obtained and self-consistency was found to have substantial (qualitative) impacts on the resulting QP band structure: 1) The corresponding QP band gap overshoots as much as LDA Kohn-Sham band structure underestimates it. 2) The occupied QP bandwidth is wider than LDA one, as opposed to be narrower in the case of non-self-consistent GWA result. Therefore, it was concluded in Ref. [67], from the viewpoint of BK conserving scheme, that the band gap problem is still an open question. It was further suggested that processes beyond GWA need to be included in order to improve the basic understanding of the problem via better agreement with experiment.

It may be relevant to reiterate here the recent debate [72,73] about maintaining self-consistency in the context of QP band gap. Since the discovery of significance of impact from self-consistency [66,67], instead of claiming that self-consistency has no effect, some authors have now taken another viewpoint, that self-consistency is futile or to be avoid [66]. Some others have developed methods involving "partial self-consistency" that update G but fix W within GWA to better fit the experimental data, by arguing that W evaluated from G^{LDA} is "better" than the one from G^{GWA} [72]. From the viewpoint of the BK conserving scheme, only the fully self-consistent solution is physically meaningful due to its fulfillment of all the microscopic conservation laws. Thus, what needs to be improved is the inclusion of proper processes into consideration with diagrams beyond GWA. Only this way, more fundamental understanding could be

achieved, instead of "getting the answer right by intentionally solving the equation incorrectly."

Discussed in this section, before "aiming beyond the GWA" as suggested in Refs. [67] and [73], is a careful examination of the existing GWA results, by performing an exhausting all-electron conserving GWA calculation. Intriguingly, it is found that result from the present work significantly differs from all the previous work (conserving or not). The difference comes from the following important factors: 1) The use of pseudo wave functions generated from the pseudopotential, employed in most of the existing GWA implementations, gives sizable error in the resulting band gap. 2) Approximating the full contribution from the core states in GWA, by its LDA counterpart, also produces an error of similar magnitude. Most interestingly, the present work, while still within GWA, suggests that self-consistency helps to improve the resulting band gap to good satisfaction, contrary to the case in Ref. [67].

3.3.1 Breakdown of Pseudo Wave Functions in Evaluation of Matrix Elements of Self-Energy

Listed in Table 1 is a comparison between non-self-consistent GWA results of band gap and bandwidth from previous works and the present work. It is apparent, through comparison with all-electron results in the last two rows, that all the pseudo potential approaches overshoots the resulting QP band gap and bandwidth (the difference between the last two rows is due to the inclusion of full inclusion of core contribution in

Table 1 Breakdown of Use of Pseudo Wave Functions

QP band gap and bandwidth of Si. The data in 3rd row to 7th row are all based on use of pseudo wave functions. All theoretical results, except LDA one, are within non-conserving GWA. Witness the overestimation on absolute band gap in pseudo wave function approaches. The difference between the last two rows is attributed to the inclusion of GWA from core states in the present work.

	abs. gap	direct gap at Γ point	occupied bandwidth
Landolt-Bornstein (experiment) [74]	1.17	3.40	12.5 ± 0.6
Present work (LDA)	0.52	2.53	12.22
Hybertsen and Louie [69]	1.29	3.35	12.04
Rohlfing, Krüger, and Pollmann [75]	1.17	3.33	12.35
Rojas, Godby, and Needs [76]	1.29	3.29	…
Fleszar and Hanke [77]	1.19	3.23	11.57
Schöne and Eguiluz [67]	1.34	3.27	11.65
Hamada, Hwang, and Freeman (all e^-) [78]	1.01	3.30	12.21
Present work (all e^-)	0.85	3.12	12.15

GWA and will be discussed in the next section.) Two possible sources for this discrepancy are: 1) use of pseudopotential, which affect V^{LDA} in Eq. (30), or 2) difference in the wave functions, used as basis functions in these calculations. From the following discussion, it should be clear that the later is responsible of the discrepancy.

The current well-established theory of the pseudopotential (see Ref. [60] for a detail discussion) allows the constructed pseudopotential to accurately mimic the effect on V^{LDA} from the core states, thus leaving the core states out of the actual calculation. Therefore, the quality of V^{LDA} with the pseudopotential approach should be consistent with the all-electron V^{LDA} in the case of Hamada *et al* or the present work. This conclusion rules out the first possibility.

On the other hand, using pseudo wave functions as basis functions in calculating matrix elements of general operators, e.g.: non-local self-energy, is never justified. In fact, recent study [79] has indicated that the error introduced in using the pseudo wave functions in calculating matrix elements of operator $e^{-i\vec{k}\cdot\hat{\vec{x}}}$ could be substantial. This statement is not surprising, as the "transferability" [60] of the pseudo wave functions only guarantees the similarity between the pseudo wave functions and the all-electron wave functions in the "interstitial" region (outside the "atomic spheres"). In general, since most of the important static properties are determined by the behavior of valence wave functions in the interstitial, the pseudopotential approaches can thus generate results similar to the all-electron ones while reducing the computational expense. However, in the evaluation of the matrix elements, the integral involves the whole real

space, including interstitial region and atomic spheres, so the difference in the atomic spheres between the pseudo wave functions and the all-electron ones will become apparent as the all-electron wave functions poses strong oscillation near the atomic sites, required by orthogonality to the core wave functions, while the pseudo wave function is smooth by construction. Based on the above consideration, it is reasonable to suspect that the error in the pseudo potential results is caused by the improper smoothing of the pseudo wave functions near atomic sites.

3.3.2 Role Played by the Core States

The results in the last two rows of Table 1, both corresponding to all-electron approaches, also disagree with each other. This is because of the full inclusion of the core contribution to the self-energy within GWA in the present work.

In the work of Hamada *et al*. [78] (second rows from the last in Table 1) and other similar work [80], the core contribution to GWA self-energy is approximated by LDA:

$$\Sigma^{\text{from core}} \approx V^{\text{LDA}}\left[n^{\text{valence + core}}\right] - V^{\text{LDA}}\left[n^{\text{valence}}\right] \tag{32}$$

where $n^{\text{valence + core}}$ is the total density and n^{valence} is the valence density. This approximation is also implicitly built into all the pseudopotential approaches, by construction. In the present work, this approximation is abandoned, motivated by BK conserving requirement. Instead, the full GWA self-energy is evaluated with the information of (relativistic) core wave functions. It turns out that this approximation also produces error in the band gap, of similar magnitude to the use of pseudo wave

functions, as can be clearly observed as the difference of the last two rows of Table 1. In deed, the following result quantifies the discrepancy

$$\Delta\left(\Sigma_{\vec{k},j}^{\text{from core}} - \left\langle \vec{k}, j \middle| V^{\text{LDA}}\left[n^{\text{valence+core}}\right] - V^{\text{LDA}}\left[n^{\text{valence}}\right] \middle| \vec{k}, j \right\rangle\right) \sim -0.15 eV \qquad (33)$$

where Δ denotes difference between $\left(\vec{k}, j\right)$ corresponding to the bottom of the conducting bands and the top of the valence bands. Thus, use of approximation with Eq. (32) would exaggerate the band gap, similar to the employment of pseudo wave functions.

Interestingly, it is also found that on the absolute scale, the effect of the core contribution to the size of QP band gap is of order of 1 eV, through the difference in the Fock self-energy (second term in first line of Eq. (27)) of the top valence band and the bottom of the conducting band, as a consequence of different degree of localization of these bands. That is, contrary to the statement given in Ref. [81] and the intuitive estimation one might have, the overlap between core and valence wave functions does contribute to the self-energy via the "core-valence exchange process", which is important enough to qualitatively change the size of QP band gap of Si.

In fact, as shown in Figure 13, the existence of core level is partially responsible for making Si a semiconductor, instead of a metal, at the level of the non-conserving GWA. The left panel shows the full $G^{\text{LDA}}W^{\text{LDA}}$ result that possesses a gap of 0.85 eV. The right panel shows the same calculation excluding contribution from the core states into the self-energy within GWA. It can be clearly observed that the resulting new QP band structure demonstrates no band gap at all, which would have made Si a metal!

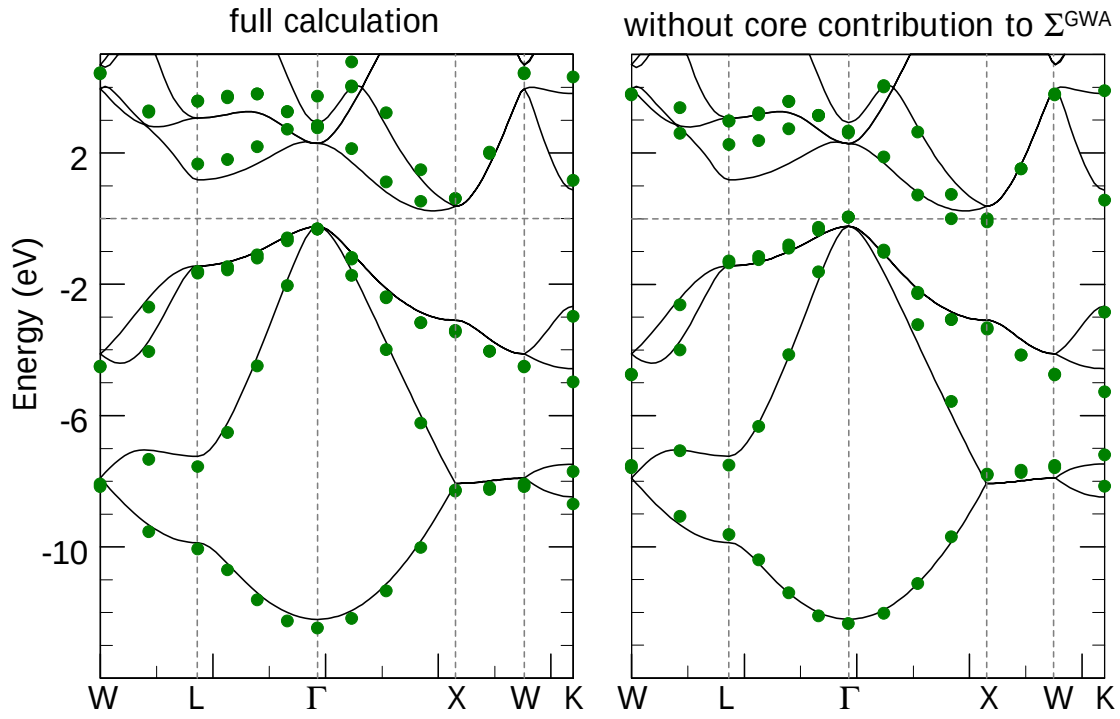


Figure 13 Significance of Core Contribution: QP band structure of Si

The calculated QP band structure (filled circles) with/without (left/right panel) the core contribution to the (non-conserving) $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ self-energy is compared with LDA Kohn-Sam band structure (solid line). The contribution to the size of QP band gap is ~ 0.9 eV. Observe that in a non-conserving GWA calculation, Si would have become a metal in the absence of the contribution!

3.3.3 Effect of Self-Consistency

Illustrated in Table 2 is the effect of self-consistency. As discussed earlier, self-consistency was found in previous work [67] to have considerable effect on the QP band gap of Si, contrary to the assumption of the community. In fact, in that work, self-consistency was found to overshoot the size of band gap by the similar amount of what LDA underestimates it (4^{th} and 6^{th} row in Table 2.) This striking result, as well as the overshooting in QP occupied bandwidth reported in Ref. [66], initiated arguments against performing the self-consistent GWA calculation.

This trend of increasing the size of band gap and occupied bandwidth is consistently reproduced in the present work. However, most interestingly, due to the different magnitude of the enlargement, as well as the different starting non-self-consistent results, the self-consistent result obtained in this work turns out to have an improved band gap, not worse one. (The different starting non-self-consistent results are due to the improvement implemented in the present work discussed in the previous sections: all-electron wave function basis and explicit inclusion of core contribution.) Consider the approximations discussed earlier, it is apparent that the excellent agreement produced by the previous works (3^{rd} - 7^{th} rows in Table 2) is an accident due to cancellation of errors (lack of self-consistency against approximate treatment of core contribution and use of pseudo wave function as basis).

Table 2 Effect of Self-Consistency: QP band gap and bandwidth of Si

The data in 3rd, 4th, and 6th rows are all based on the use of pseudo wave functions. The last two rows are conserving GWA results corresponding to pseudopotential and present all-electron approach. Witness the overestimation on absolute band gap in approaches employing pseudo wave functions. The trend from non-conserving GWA to the conserving one is consistent: enlarging the gap and bandwidth. However, due to difference in magnitude, self-consistency happens to improve the bag gap in present work while "worsen" the one in the pseudo potential case.

	abs. gap	direct gap at Γ point	occupied bandwidth
Landolt-Bornstein (experiment) [74]	1.17	3.40	12.5 ± 0.6
Present work (LDA)	0.52	2.53	12.22
Hybertsen and Louie [69]	1.29	3.35	12.04
Schöne and Eguiluz [67]	1.34	3.27	11.65
Present work (all e^-)	0.85	3.12	12.15
Schöne and Eguiluz (SC) [67]	1.91	4.02	13.10
Present work (SC, all e^-)	1.03	3.48	13.53

It is important to clarify that whether the conserving GWA result give the exact QP band gap should not be regarded as a measure of success or necessity of the self-consistent requirement; instead, it merely reflects the quality of the approximated self-energy process included (GWA in this case). From the point of view of BK conserving scheme, self-consistency is a necessary element to ensure microscopic conservation laws. In the case of QP band gap in Si, the size of the gap seems to be mainly controlled by the process of screened interaction described in conserving GWA. On the other hand, the occupied QP bandwidth is still too large compared to the experimental data and therefore requires inclusion of processes beyond GWA.

3.3.4 Summary

In this section, previous results of QP band gap in Si, within well-known GWA, are re-examined, through a newly developed all-electron implementation of finite temperature MBPT within BS conserving scheme. It is suggested that conserving GWA consists of most of the important mechanisms in the physics of band gap in Si, contrary to suggestion of earlier work in Ref. [67], even though the trend of including self-consistency in both calculations is consistent (substantial increase of the band gap and occupied bandwidth.) Furthermore, it is found that, technically, errors introduced from approximating core contribution with LDA and using pseudo wave functions as basis are both considerable. The excellent agreement achieved in many existing works is therefore an accident due to cancellation of these errors and the lack of self-consistency.

Finally, the core states are found to play a surprisingly essential role in controlling the size of QP band gap in this system.

3.4 Quasi-Particle Band Gap of Germanium

Germanium is interesting because of its similarity to Si in the QP band structure near the chemical potential except for the extra 3*d* shallow core states, which are very localized in space. Furthermore, the Kohn-Sham band structure of Ge within LDA turns out to present no gap at the Γ point, worse than the case of Si. Since the QP band gap was reasonably understood in the study discussed in the previous section, the case of Ge serves as an ideal test case for probing the mechanisms in which these shallow core states participate. That is, one would expect a similar degree of success in using conserving GWA to describe the QP band structure (especially the band gap) of Ge, if effects of these semi-core states were not too different from those of the extended valence states.

The case of Ge also poses a challenge on the theory within GWA, as a straightforward implementation of the non-conserving GWA on Ge without these shallow core states would open the band gap as in the case of Si, but the topology of the resulting QP band structure suffers from the incorrect behavior of possessing direct gap at the Γ point [81], rather than an indirect gap defined by conducting L and valence Γ point (Γ-L gap) as observed in experiment.

To date, for systems with shallow core states, like Ge discussed in this section, it is generally accepted that these core states play an important role in the processes

contained in non-conserving GWA, as the inclusion of these states in the screening process, built into *W*, changes the topology of the QP band structure [81-83]. Specifically, once included through phenomenological core polarization potential (CPP) [81,82] or explicit treating the core level with state-of-the-art pseudopotential [83], the screening from these shallow core states is concluded to be responsible to change the QP band gap of Ge from the direct band gap at Γ point to the indirect Γ-L gap. Furthermore, excellent agreement of the QP band gap with the experiment is achieved within GWA with this screening taken into account.

In this work, the case of Ge will be re-examined in this section with the all-electron approach described above, as the newly implemented conserving method developed in the present work allows a microscopic investigation of the role played by the shallow core states. First, it is found that, surprisingly, the shallow 3*d* core states in Ge participate very weakly in the screening process. Therefore, the suggestion of core polarization is not supported by the *ab intio* result. Second, it is found that without the self-energy contribution from these core states, the resulting QP band structure already possesses the indirect gap, contrary to the existing pseudopotential results in Refs. [81-83]. Finally, the conserving GWA is shown to be insufficient to totally describe the QP band structure of Ge, even though the size of the indirect Γ-L gap is found to be in good agreement with experiment.

3.4.1 Re-Examination of Core Polarization Effect

The idea of Core Polarization Potential (CPP) was introduced [82] in the pseudopotential implementation of QP calculation of Ge within non-conserving GWA in order to improve the incorrect QP band structure near the chemical potential, which still displays a direct gap as in the case of LDA Kohn-Sham band structure. Since energetically these *d* states are only ~ 25 eV below the chemical potential (occupied bandwidth is ~ 13 eV), CPP seems to be a reasonable and physically motivated approach when *ab initio* inclusion of the true screening effect is not available. Furthermore, the use of such potential seems to be supported by recent pseudopotential approach [83], in which the shallow 3*d* core states are explicitly included in the calculation of polarization process.

However, as shown in Figure 14, all-electron evaluation of the dynamical density response function without the 3*d* core contribution (removed from the summation over bands in Eq. (10)) is not very different from the full response function (with inclusion of the 3d core states), which is proportional to W^{LDA}. The main effect of the core states is only to slightly shift the short-lived plasmon mode, as well as to provide a dielectric background beyond 25 eV, which is the binding energy of these core states in LDA. Interestingly, the static limit of the response functions, the most important part of W^{LDA} in GWA calculation of self-energy, displays almost no difference whether the core is included or not. Therefore, it is apparent that the importance of the core polarization

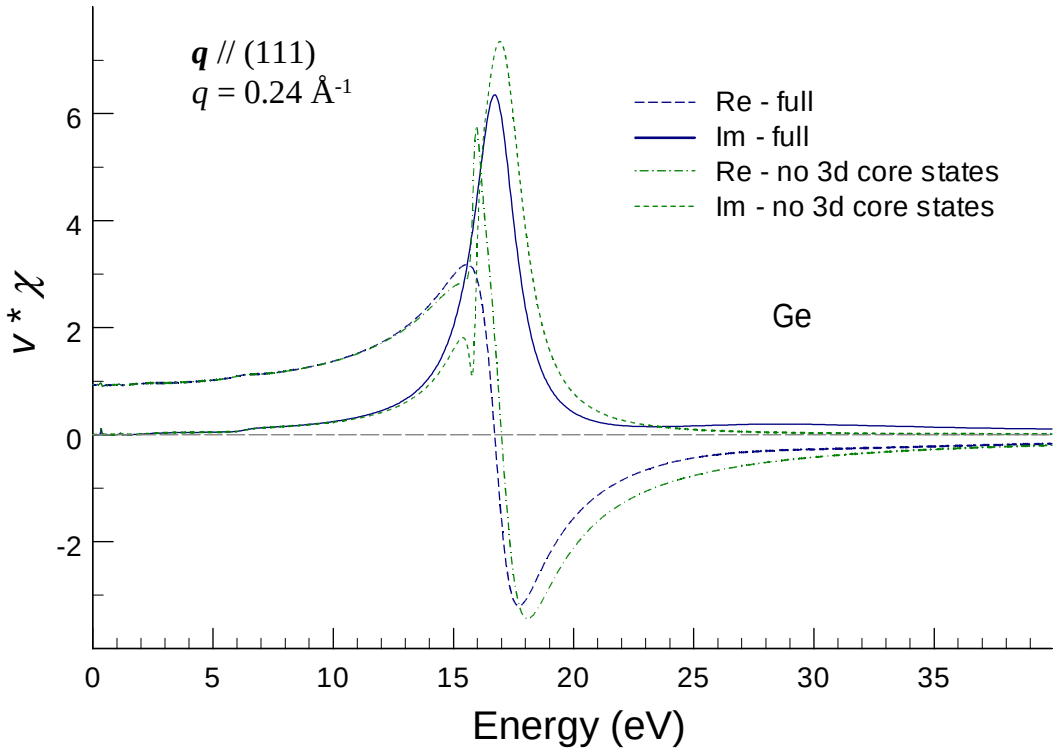


Figure 14 Impact of Core Polarization: Dynamical Response of Ge

The real and imaginary part of the calculated dynamical linear density response function of Ge with and without the core contribution to the polarization process is compared. The effect of 3*d* core states is mainly to slightly shift the peak energy of the short-lived plasmon mode, as well as to provide the small background beyond 25 eV, which is the binding energy of these shallow core states in LDA. Notice that near the static limit ($\omega \sim 0$), the response functions are almost identical to each other.

was exaggerated in the pseudopotential calculations and that the use of CPP to correct earlier GWA result, which does not include core states, is clearly not justified by the *ab initio* result.

In deed, in results of this work, as shown in Figure 15, the effect of the shallow core states, within non-conserving GWA, is found to be very small (except for the minor shift along Γ to X point) when comparing QP band structure from the full all-electron calculation and the one without contribution from 3*d* states. The difference between these two results comes from two effects: 1) the screening process discussed above (the *W* part of GWA) and 2) core propagator contribution to the self-energy (the *G* part of GWA). Based on the comparison of dynamical density response function discussed above, it should not be a surprise now that the first effect is minimal. However, it is intriguing to note that the second effect is also minor, which is probably due to difference in the dominant angular momentum between the 3*d* states and states near the chemical potential (which are mainly of *s* and *p* nature), as well as the rather simple form of GWA which does not strongly account for mechanisms mixing states with different angular momentum.

The results in Figure 15 clearly indicate that the mechanism behind the change from direct gap to indirect one is **not** related to the shallow 3*d* core states, contrary to the current understanding of the community, since the change occurs even without inclusion of these core states in the all-electron calculation. Therefore, it is probably the intrinsic limitation of pseudopotential approach (approximated core self-energy

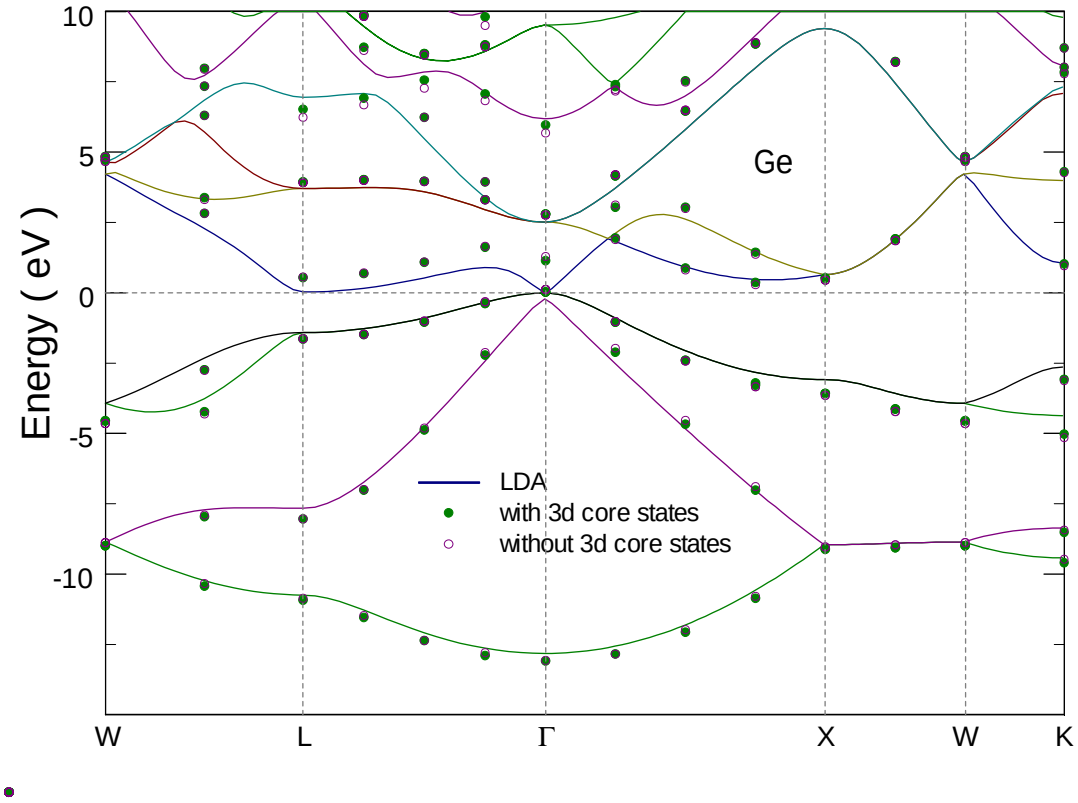


Figure 15 Impact of Core Polarization: QP band structure of Ge

The calculated QP band structure of Ge with and without the core contribution to the polarization process within non-self-consistent GWA ($G^{\mathrm{LDA}}W^{\mathrm{LDA}}$) is compared. The effect of 3*d* core states is very small except near X point. Notice that even without the 3*d* core states, the resulting QP band structure already has the correct topology along Γ to L. That is, the removal of direct gap happens without these 3*d* states.

contribution and lack of realistic wave functions, as discussed above) that produces results in Refs. [81-83] with direct gap, which should have been indirect in the all-electron calculation, in the absence of 3*d* core contribution. In short, from the all-electron *ab intio* results of Ge, the suggestion of core polarization is not justified and 3*d* core states do not strongly contribute to the polarization process or to the change from direct gap to indirect one in the QP band structure.

3.4.2 Effect of Self-Consistency and Insufficiency of GWA

Shown in Table 3 is the self-consistent (conserving) all-electron result of QP band gap and occupied bandwidth, compared with experiment and existing results from other authors. The trend of self-consistency is to increase the QP energy as in the case of Si. Note that the indirect gap at Γ-L turns out to agree with experiment very well. However, careful inspection of other quantities listed in the table clearly exposes the insufficiency of the GWA.

The source of the inadequacy of the GWA can be traced back to the improper strength of screening in GWA as illustrated in Figure 16, in which the band structures of LDA, GWA, and HF are plotted. The first two conducting bands are highlighted so that the trend can be easily observed. The GWA (dynamically) screens out partially the bare Coulomb exchange potential built into the HF case, thus it brings the GWA bands closer to the LDA ones. For these two bands, the GWA screening is so strong near X point that it brings GWA bands nearly back to LDA ones from their HF counterparts.

Table 3 Insufficiency of GWA: QP band gap and bandwidth of Ge

The data in shaded area are both based on pseudo wave functions. The trend of self-consistency is to increase QP energy. While the indirect Γ-L gap is well reproduced, direct gap at Γ, indirect gap at Γ-X and the occupied bandwidth are not in good agreement with the experiment within conserving GWA.

	indirect gap (Γ-L)	direct gap at Γ point	indirect gap (Γ-X)	occupied bandwidth
Landolt-Bornstein (experiment)	0.74	0.89	…	…
Present work (LDA)	0.35	-0.20	0.66	12.82
Shirley, Zhu, and Louie (CPP) [81]	0.73	0.85	1.09	…
Rohfling, Krüger, and Pollmann [83]	0.71	0.81	…	…
Present work (all e^-)	0.51	1.11	0.49	13.12
Present work (SC, all e^-)	0.79	1.51	0.71	14.77

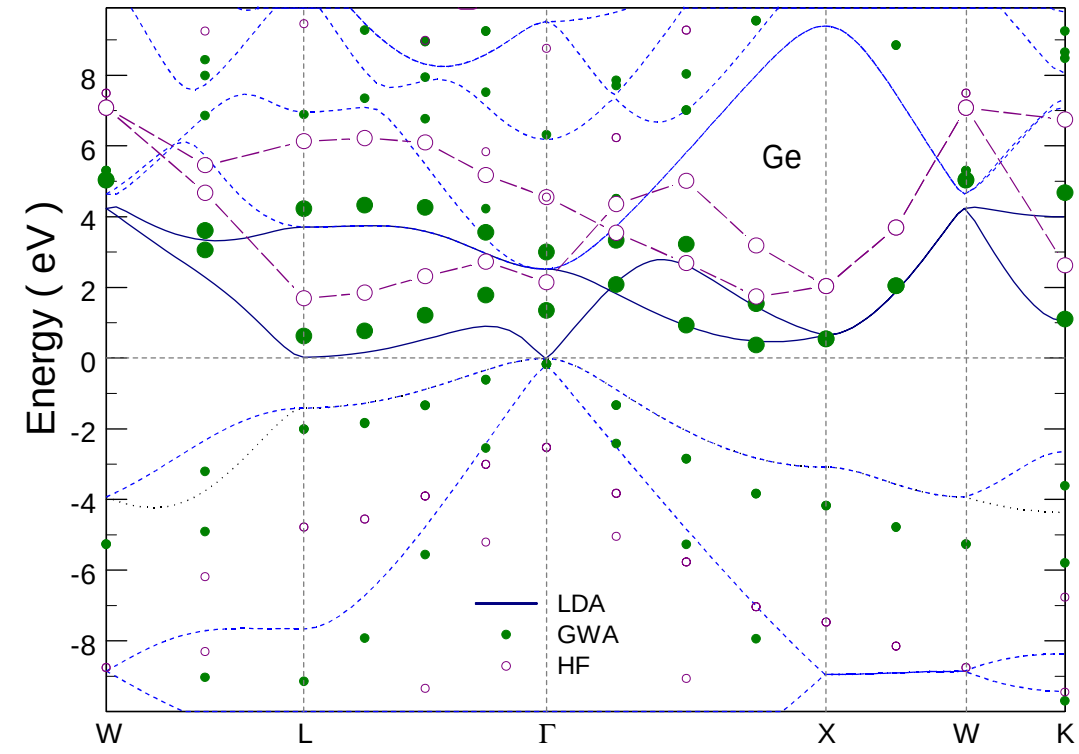


Figure 16 Source of Inadequacy in the GWA: QP band structure of Ge

The calculated QP band structures of Ge within GWA and HF are compared. The first two conducting bands are highlighted by solid lines, large full circles and short dashed lines, corresponding to LDA, GWA, and HF respectively. Notice that the GWA screening is much more efficient near X point for these bands.

It is therefore reasonable to assume that mechanisms that partially compensate the GWA screening (in a non-uniform, angular dependent manner in the phase space) should be included. Guided by the work of Hanke and Sham [68], it seems reasonable to expect improvement through inclusion of "vertex correction", a new set of diagrams beyond GWA.

Finally, it is important to clarify the role of the shallow 3*d* states. As illustrated above, these 3*d* states are found to play insignificant roles in the GWA result. However, this does not rule out their potential to participate in physical processes beyond the GWA. That is, while the core polarization effect within the GWA was questioned above, the true physical impact from these core states remains to be investigated, through self-energy beyond the GWA. For example, a possible important process would be the excitonic excitation involving these localized states. To include such excitation, processes including electron-hole pair interaction (not in GWA) are thus necessary.

3.4.3 Summary

In the sections, the QP band structure of Ge is investigated with a newly developed all-electron conserving GWA implementation. The existing theory of the band gap based on core polarization is found to be unsupported by the *ab initio* result. In fact, the effect of the shallow 3*d* core states on the resulting QP band structure turns out to be small within non-conserving GWA. The change from direct gap to indirect one, which was assumed to be the task of core polarization, occurs even without these core states. This

new result quantitatively differs from published pseudopotential results, the reason of which is attributed to the intrinsic limitation of the pseudopotential approach (detailed discussion in the above sections).

Furthermore, the fully self-consistent (conserving) GWA result in the present work agrees with the experiment for the indirect gap at Γ-L, but fails to describe the direct gap, indirect gap at Γ-X, and the occupied bandwidth. The inadequacy is illustrated by comparing the strength of GWA screening effect on the first two conducting bands. Therefore, processes beyond GWA, that could modify the strength of the screening in a non-uniform (directionally and energetically dependent) manner, are probably necessary be included in the self-energy in order to improve the theoretical description of the system. Finally, as a reminder, the role of the shallow core states in these new processes remains to be demonstrated.

Chapter 4 Conclusions

In the case studies in this work, the many-particle nature of the systems of interest (K, Zn, Na, Si, Ge) was investigated through comparison of rich information of electronic excitations (dynamical linear response and quasi-particle excitation) from experimental and *ab intio* theoretical results. Discovered in the careful numerical analysis were several novel physical effects involving intriguing microscopic processes, based on which new interpretations (on dominating mechanisms) of these systems were specified.

In the case of dynamical linear response of K, the plasmon linewidth dispersion was found not to be due to a dynamical exchange-correlation effect as proposed in the original experimental article [24]. Instead, particle-hole excitation involving transitions into unoccupied Kohn-Sham states of *d* symmetry is the main decay mechanism of the plasmon and thus determines the linewidth dispersion. This investigation was made possible by a newly developed all-electron TDDFT scheme with imaginary frequency sampling technique. Further analysis also demonstrated the non-trivial role of the exchange-correlation effects built into the LDA Kohn-Sham band structure, as they control the energy of the crucial unoccupied bands. Furthermore, the polarization effect of the shallow 2*p* core states also contributes to quantitative description of the lifetime dispersion.

In the case of dynamical linear response of Zn, the *d*-threshold related excitation discovered in EELS experiment was found to be collective in nature, contrary to the

picture of "single-particle transition" employed in the original experimental analysis [24]. The oscillator strength used in the experimental analysis turns out to depend weakly on momentum transfer and thus is not responsible for the decrease of the strength of the "onset" in the EELS spectra. The apparent decrease of the strength of the onset is, in fact, a result of a subtle interplay between this collective mode and the intrinsic plasmon mode through the polarization effect due to their vicinity in energy. Also encountered in this case study is the breakdown of LDA approach in describing the energy of the flat 3*d* semi-core states, which is essential to the existence of the collective mode and thus is capable of dramatically change the corresponding response function. A more fundamental improvement than the employed phenomenological approach remains a theoretical challenge.

In the case of quasi-particle excitation of Na, the occupied QP bandwidth was briefly discussed. This work illustrates the importance of the BK conserving scheme, which is fully employed in the present work. It was shown that the neglect of self-consistency in solving the Dyson equation brings about a substantial error in the calculation of the bandwidth, contrary to the simplifying assumption made by most of authors. In fact, this error introduces clear ambiguities in the identification of the relevant correlation mechanisms.

In the case of QP excitation of Si, the QP band gap was re-examined within the conserving GWA, with a newly developed all-electron finite temperature MBPT implementation of the BK conserving scheme. The resulting indirect fundamental band gap agrees well with the experimental value within GWA. Intriguingly, the deep core

states were found to play an important role in defining the QP band gap of Si, contrary to a intuitive simple picture found in current literature. This non-trivial core contribution can be understood as a consequence of the difference in energy of the exchange processes, involving the deep core states, of the valence states and the conducting states. Furthermore, the nature of strong oscillation near the atomic sites of the wave functions, serving as basis functions in the calculation, was found to also contribute the size of the band gap qualitatively. At last, the results in this work suggest that self-consistency helps to improve the QP band gap, not worsen it.

Conclusions from the study of Si in this work differ qualitatively from those in all the previous works with the pseudopotential approaches. The source of the discrepancy was traced to the approximations employed in the pseudopotential approaches of the QP excitation calculations: 1) use of pseudo wave functions that smooth out the oscillations near atomic sites, and 2) approximate core contribution to the GWA self-energy with its LDA counterpart. The cancellation of errors of these approximations and the lack of self-consistency contributes to the previous successful stories of non-conserving GWA approaches being referred to in the discussion.

At last, in the case of QP excitation of Ge, the QP band gap is investigated, motivated by its similarity to Si with the extra shallow 3*d* core states. Surprisingly, these semi-core states were found to play a minor role in the microscopic processes described by GWA: through the propagator, *G*, in the self-energy, or the polarization processes *W*. Furthermore, even tough the size of the fundamental indirect Γ-L band gap was well reproduced, insufficiency of GWA is more apparent in the case of Ge due

to its poor description on other quantities in the QP band structure, e.g.: direct band gap at Γ point, indirect Γ-X band gap, and occupied bandwidth. Therefore, it seems reasonable to assume that mechanisms beyond those in GWA, that are capable of non-uniformly reducing the GWA screening, in an angular dependent manner, need to be properly (self-consistently) included in order to further our understanding of the system. In those mechanisms, the role played by the shallow core states remains to be demonstrated, as the locality of the core states may enhance their importance.

The *ab initio* results in the study of Ge do not support the account of importance of the core polarization effect in the existing literature. In fact, it was shown that the turnover from direct gap, as in LDA, to indirect gap, as in the experiment, is not due to the shallow 3*d* core states. Thus, the current understanding of the community, that the turnover is due to core polarization, is likely misled by the intrinsic limitation of the existing pseudopotential approaches discussed in the Si section.

Finally, the schemes implemented and novel algorithms developed in this work allow investigation of important physical mechanisms involved in the existing experimental data of electronic excitations. Not only new interpretation of features in experimental data was given in this work, comparison with existing theoretical conclusion also revealed interesting new insight on the microscopic processes involved.

The novel technical advances (discussed in detail in the appendices) in this work include the all-electron approach (of excitation), the imaginary frequency sampling implementation (for linear response calculations), the Matsubara time implementation

(for QP excitation calculations), the uniform-power mesh, the iterative approach of solving the integral equation, and the parallel computing model. Not only they make possible a careful analysis of role played by the deep core states or the shallow semi-core states, but they also allows study of systems with flat *d* bands near the chemical potential, like transition metals or their oxides, while lowering the temperature, enlarging the energy range allowed, and reducing the computation expense.

Further extension of this work may include study of more complicated systems, inclusion of processes beyond those discussed in the work, and investigation of other kind of electronic excitations, e.g.: magnetic excitation or excitonic excitation.

Bibliography

Bibliography

1 W. Schülke, *Handbook on Synchrotron Radiation*, Vol 3, edited by G. Brown and D. E. Moncton (Elsevier Science Publishers, B.V., 1991), pp. 565.

2 E. Runge and E. K. U. Gross, Phys. Rev. Lett. **52**, 2850 (1984).

3 M. Petersilka, U. J. Gossmann, and E. K. U. Gross, Phys. Rev. Lett. **76**, 1212 (1996).

4 R. van Leeuwen, Phys. Rev. Lett. **82**, 3863 (1999).

5 E. K. U. Gross and W. Kohn, Adv. Quant. Chem. **21**, 255 (1990).

6 M. Petersilka, U. J. Gossmann, and E. K. U. Gross, Phys. Rev. Lett. **76**, 1212 (1996).

7 P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964); W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

8 J. P. Perdew and Y. Wang, Phys. Rev. B **45** 13244 (1992)

9 P. Blaha, K. Schwarz, P. Dufek, and R. Augustyn, WIEN95, Technical University of Vienna, 1995 (improved and updated UNIX version of the original copyrighted WIEN code, P. Blaha, K. Schwarz, P. Soratin, and S. B. Trickey, Comp. Phys. Commun. **59**, 399 (1990)).

10 D. J. Singh, *Plane Waves, Pseudopotentials, and the LAPW Method* (Kluwer, Boston, 1994).

11 J. Sullivan, W. Ku, and A. G. Eguiluz, (to be published)

12 A. A. Quong and A. G. Eguiluz, Phys. Rev. Lett **70**, 3955 (1993); N. E. Maddocks, R. W. Godby, and R. J. Needs, Phys. Rev. B **49**, R8502 (1994); N. E. Maddocks, R. W. Godby, and R. J. Needs, Europhys. Lett. **27**, 681 (1994); F. Aryasetiawan and O. Gunnarsson, Phys. Rev. Lett. **74**, 3221 (1995); A. Fleszar, A. A. Quong, and A. G. Eguiluz, Phys. Rev. Lett. **74**, 590 (1995); A. Fleszar, R. Stumpf, and A. G. Eguiluz, Phys. Rev. B **55**, 2068 (1997); M. Ehrnsperger and H. Bross, J. Phys.: Cond. Matt. **9**, 1225 (1997).

13 H. J. Vidberg and J. W. Serene, J. Low. Temp. Phys. **29**, 179 (1977).

14 K. -H. Lee and K. J. Chang, Phys. Rev. B **54**, R8285 (1996).

15 B. Farid, G. E. Engel, R. Daling, and W. van Haeringen, Phil. Mag. B **69**, 901 (1994).

16 Y.-G. Jin and K. J. Chang, Phys. Rev. B **59**, 14841 (1999).

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

18 N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Saunders College Publishing, Fort Worth, 1976).

19 D. F. DuBois and M. G. Kivelson, Phys. Rev. **186,** 409 (1969).

20 T. Kloos, Z. Physik **265**, 225 (1973).

21 P. C. Gibbons, S. E. Schnatterly, J. J. Ritsko, and J. R. Fields, Phys. Rev. B **13** 2451 (1976).

22 K. Sturm, Z. Physik B **25**, 247 (1976).

23 K. Sturm and L. E. Oliveria, Phys. Rev. B **24** 2054 (1981).

24 A. vom Felde, J. Sprösser-Prou, and J. Fink, Phys. Rev. B **40**, 10181 (1989).

25 J. Fink, *Unoccupied Electronic States*, edited by J. C. Fuggle and J. E. Inglesfield (Springer, Berlin, 1992), pp. 203.

26 F. Aryasetiawan and K. Karlsson, Phys. Rev. Lett. **73**, 1679 (1994), and A. Fleszar, R. Stumpf, and A. G. Eguiluz, Phys. Rev. B **55**, 2068 (1997).

27 W. Ku and A. G. Eguiluz, Phys. Rev. Lett. **82**, 2350 (1999).

28 K. Sturm, E. Zaremba, and K. Nuroh, Phys. Rev. B **42**, 6973 (1990).

29 K. Widder, M. Knupfer, O. Knauff, and J. Fink, Phys. Rev. B **56**, 10154 (1997).

30 M. Knupfer, K. Widder, M. Sing, O. Knauff, and J. Fink, Eur. Phys. J. B **6**, 323 (1998).

31 I. Campillo, A. Rubio, and J. M. Pitarke, Phys. Rev. B **59**, 12188 (1999).

32 M. A. Cazadilla, J. S. Dolado, A. Rubio, and P. M. Echenique, cond-mat/9907204.

33 F. Aryasetiawan and O. Gunnarsson, Phys. Rev. B **49**, 7219 (1994).

34 F. Aryasetiawan, O. Gunnarsson, M. Knupfer, and J. Fink, Phys. Rev. B **50**, 7311 (1994).

35 A. Fleszar and W. Hanke, Phys. Rev. B **56**, 12285 (1997); H. Dröge, A. Fleszar, W. Hanke, M. Sing, M. Knupfer, J. Fink, F. Goschenhofer, C. R. Becker, R. Kargenbauer, and H. P. Steinrück, Phys. Rev. B **59**, 5544 (1999).

36 A. Zangwill and P. Soven, Phys. Rev. A **21**, 1561 (1980).

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

38 D. Singh and D. A. Papaconstantopoulos, Phys. Rev. B **42**, 8885 (1990).

39 T. Grabo, T. Kreibich, S. Kurth, and E. K. U. Gross, *Strong Correlations in Electronic Structure: Beyond the Local Density Approximation*, edited by V. Anisimov (Gordon and Breach, 1999).

40 M. Städele, M. Moukara, J. A. Majewski, P. Vogl, and A. Görling, Phys. Rev. B **59**, 10031 (1999).

41 J. C. Slater, Phys. Rev. **81**, 385 (1951).

42 H. Yasuhara, S. Yoshinaga, and M. Higuchi, Phys. Rev. Lett. **83**, 3250 (1999).

43 W. Ku, A. G. Eguiluz, and E. W. Plummer, Phys. Rev. Lett. **85**, 2410 (2000).

44 K. S. D. Beach, R. J. Gooding, and F. Marsiglio, Phys. Rev. **61**, 5147 (2000)

45 G. D. Mahan, *Many-Particle Physics* (Plenum, New York, 1990).

46 J. W. Negele and H. Orland, *Quantum Many-Particle Systems* (Addison-Wedley, Redwood, 1987).

47 R. D. Mattuck, A Guide to Feynman Diagrams in the Many-Body Problem (Dover, New York, 1992).

48 L. P. Kadanoff and G. Baym, *Quantum Statistical Mechanics* (Benjamin, New York, 1962).

49 L. Hedin, J. Phys.: Condens. Matter **11**, R489 (1999).

50 J. J. Quinn and R. A. Ferrel, Phys. Rev. **112**, 812 (1958).

51 J. J. Quinn, Phys. Rev. **126**, 1453 (1962).

52 J. J. Quinn, Appl. Phys. Lett. **2**, 167 (1963).

53 I. Campillo, J. M. Pitarke, A. Rubio, E. Zarate, and P. M. Echenique, Phys. Rev. Lett. **83**, 2230 (1999).

54 G. Baym and L. P. Kadanoff, Phys. Rev. **124**, 287 (1961).

55 G. Baym, Phys. Rev. **127**, 1391 (1962).

56 J. M. Luttinger and J. C. Ward, Phys. Rev. **118**, 1417 (1960).

57 A. Schindlmayr, Phys. Rev. B. **56**, 3528 (1997).

58 L. Hedin, Phys. Rev. **139**, A796 (1965).

59 R. Barrett *et al*., *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, 2nd Edition (SIAM, Philadelphia, 1994); G. H. Golub, C. F. Van Loan, *Matrix computations* (Johns Hopkins University Press, Baltimore, 1996)

60 W. E. Pickett, Comp. Phys. Rep. **9**, 115 (1989)

61 E. Jensen and E. W. Plummer, Phys. Rev. Lett. **55**, 1912 (1985).

62 I.-W. Lyo and W. W. Plummer, Phys. Rev. Lett. **60**, 1558 (1988).

63 K. W.-K. Shung and G. D. Mahan, Phys. Rev. Lett. **57**, 1076 (1986); K. W.-K. Shung, B. E. Sernelius, and G. D. Mahan, Phys. Rev. B **36**, 4499 (1987); K. W.-K. Shung and G. D. Mahan, Phys. Rev. B **38**, 3856 (1988); G. D. Mahan and B. E. Sernelius, Phys. Rev. Lett. **62**, 2718 (1989).

64 X. Zhu and A. W. Overhauser, Phys. Rev. B 33, 925 (1986).

65 J. E. Northrup, M. S. Hybertsen, and S. G. Louie, Phys. Rev. Lett. **59**, 819 (1987); M. P. Surh, J. E. Northrup, and S. G. Louie, Phys. Rev. B **38**, 5976 (1988); J. E. Northrup, M. S. Hybertsen, and S. G. Louie, Phys. Rev. B **39**, 8198 (1989).

66 B. Holm and U. von Barth, Phys. Rev. B 57, 2108 (1998).

67 W.-D. Schöne and A. G. Eguiluz, Phys. Rev. Lett. **81**, 1662 (1998).

68 W. Hanke and L. J. Sham, Phys. Rev. **21**, 4656 (1980).

69 M. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).

70 R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. **56**, 2415 (1986).

71 M. Städele, M. Moukara, J. A. Majewski, P. Vogl, and A. Görling, Phys. Rev. B 59, 10031 (1999).

72 D. Tamme, R. Schepe, and K. Henneberger, Phys. Rev. Lett. **83**, 241 (1999).

73 Adolfo G. Eguiluz, Phys. Rev. Lett. **83**, 242 (1999).

74 *Zahlenwerte und Funktionen aus Naturwissenschaften und Technik*, Landolt-Bornstein, New Series, Vol. III, Pt. 17a (Springer, New York, 1982).

75 M. Rohlfing, P. Krüger, and J. Pollmann, Phys. Rev. B **56**, R7065 (1997); M. Rohlfing, P. Krüger, and J. Pollmann, Phys. Rev. B **57**, 6485 (1998).

76 H. N. Rojas, R. W. Godby, and R. J. Needs, Phys. Rev. Lett. **74**, 1827 (1995).

77 A. Fleszar and W. Hanke, Phys. Rev. B **56**, 10228 (1997).

78 N. Hamada, M. Hwang, and A. J. Freeman, Phys. Rev. B **41**, 3620 (1990).

79 J. M. Sullivan and A. G. Eguiluz, to be published.

80 B. Arnaud and M. Alouani, Phys. Rev. B **62**, 4464 (2000).

81 E. L. Shirley, X. Zhu, and S. G. Louie, Phys. Rev. B **56**, 6648 (1997).

82 E. L. Shirley, X. Zhu, and S. G. Louie, Phys. Rev. Lett. **69**, 2955 (1992).

83 M. Rohlfing, P. Krüger, and J. Pollmann, Phys. Rev. B **57**, 6485 (1998).

84 J. J. Sakurai, *Modern Quantum Mechanics* (Addison-Wesley, Redwood City,1985).

85 J. Schmalian, M. Langer, S. Grabowski, and K. H. Bennemann, Comp. Phys. Comm. **93**, 141 (1996).

86 G. Esirgen and N. E. Bickers, Phys. Rev. B **55**, 2122 (1997).

87 C.-H. Pao and N. E. Bickers, Phys. Rev. B **49**, 1586 (1994).

88 W. Sweldens, P. Schröder, *Wavelets in Computer Graphics* (ACM SIGGRAPH Course notes, 1996)

Appendices

A Derivations of Some Useful Formula

A.1 Symmetry Properties of the Bloch States

The translation symmetry of the Bloch states are well known to conserved the crystal momentum $\vec{k}$:

$$\left\langle \vec{x}+\vec{R}\middle|\vec{k},j\right\rangle = e^{i\vec{k}\cdot\vec{R}}\left\langle \vec{x}\middle|\vec{k},j\right\rangle \quad , \tag{34}$$

where $\vec{R}$ denotes the crystal Bravais lattice vector in real space and *j* is the band index.

The rotation symmetry of the Bloch states can be derived from the symmetry of the Hamiltonian, $\hat{H}$:

$$\left[\hat{S},\hat{H}\right]=0 \quad , \tag{35}$$

where $\hat{S}$ is the rotation operator that belongs to the rotation symmetry group of the lattice, acting on real space as:

$$\hat{S}\left|\vec{x}\right\rangle = \left|\tilde{S}\,\vec{x}+\vec{\tau}\right\rangle \quad , \tag{36}$$

where $\left\{\tilde{S},\vec{\tau}\right\}$ is the pair of rotational matrix and the "partial lattice translation" vector, which corresponds to the rotation symmetry operator $\hat{S}$.

With the he secular equation for the Block states:

$$\left(\hat{H}-E_{\vec{k},j}\right)\left|\vec{k},j\right\rangle = 0 \quad , \tag{37}$$

where $E_{\vec{k},j}$ is the corresponding eigenvalue, Eq. (35) enforces

$$\left(\hat{H}-E_{\vec{k},j}\right)\hat{S}\left|\vec{k},j\right\rangle=\hat{S}\left(\hat{H}-E_{\vec{k},j}\right)\left|\vec{k},j\right\rangle=0 \quad . \tag{38}$$

Therefore, $\hat{S}\left|\vec{k},j\right\rangle$ is also an eigenstate of $\hat{H}$ with the same eigenvalue $E_{\vec{k},j}$. In the special case that $\left|\vec{k},j\right\rangle$ is non-degenerate (only possible when $\vec{k}=0$, at the Γ point, when more than one members exist in the rotation symmetry group),

$$\hat{S}\left|0,j\right\rangle=\left|0,j\right\rangle \quad . \tag{39}$$

As a consequence, the following relation on real space needs to be respected:

$$\left\langle\vec{x}\left|0,j\right\rangle=\left\langle\vec{x}\right|\hat{S}\left|0,j\right\rangle=\left\langle\tilde{S}^{-1}\cdot\left(\vec{x}-\vec{\tau}\right)\right|0,j\right\rangle \quad . \tag{40}$$

A more general rotation symmetry property can be obtained for the case of $\vec{k}\neq 0$. First, defined the periodic portion of the Bloch states $\left|u_{\vec{k},j}\right\rangle$:

$$\left\langle\vec{x}\left|u_{\vec{k},j}\right\rangle\equiv e^{-i\vec{k}\cdot\vec{x}}\left\langle\vec{x}\right|\left|\vec{k},j\right\rangle \quad , \tag{41}$$

which satisfies the periodic condition:

$$\left\langle\vec{x}\left|u_{\vec{k},j}\right\rangle=\left\langle\vec{x}+\vec{R}\right|u_{\vec{k},j}\right\rangle \quad , \tag{42}$$

as a consequence of Eq. (34) above. From Eq. (37), one can then derive the secular equation for $\left|u_{\vec{k},j}\right\rangle$ as

$$\left(\hat{H}_{\vec{k}}-\varepsilon_{\vec{k},j}\right)\left|u_{\vec{k},j}\right\rangle=0 \quad , \tag{43}$$

where

$$\hat{H}_{\vec{k}} \equiv \hat{H} + \frac{\hbar\vec{k}}{m}\cdot\hat{\vec{p}} \quad , \tag{44}$$

and

$$\varepsilon_{\vec{k},j} = E_{\vec{k},j} - \frac{\hbar^2 k^2}{2m} \quad . \tag{45}$$

One important feature of Eq. (43) is that $\vec{k}$ is now treated as an external parameter, not an index for the eigenvalue. From Eq. (35), it is straightforward to show

$$\hat{S}\hat{H}_{\vec{k}}\hat{S}^{-1} = \hat{H}_{\tilde{S}\cdot\vec{k}} \quad . \tag{46}$$

That is, in general $\hat{H}_{\vec{k}}$ does not commute with $\hat{S}$! In the special case when $\tilde{S}\cdot\vec{k} = \vec{k}$ and $\left|u_{\vec{k},j}\right\rangle$ is not degenerate with respect to index *j*,

$$\left\langle \vec{x}\middle|u_{\vec{k},j}\right\rangle = \left\langle \vec{x}\middle|\hat{S}\middle|u_{\vec{k},j}\right\rangle = \left\langle \tilde{S}^{-1}\cdot\left(\vec{x}-\vec{\tau}\right)\middle|u_{\vec{k},j}\right\rangle \quad . \tag{47}$$

Thus,

$$\begin{aligned}\left\langle \vec{x}\middle|\vec{k},j\right\rangle &= e^{i\vec{k}\cdot\vec{x}}\left\langle \vec{x}\middle|u_{\vec{k},j}\right\rangle = e^{i\vec{k}\cdot\left(\tilde{s}^{-1}\cdot\vec{\tau}\right)}e^{i\vec{k}\cdot\left(\tilde{s}^{-1}\cdot\left(\vec{x}-\vec{\tau}\right)\right)}\left\langle \tilde{S}^{-1}\cdot\left(\vec{x}-\vec{\tau}\right)\middle|u_{\vec{k},j}\right\rangle \\ &= e^{i\vec{k}\cdot\vec{\tau}}\left\langle \tilde{S}^{-1}\cdot\left(\vec{x}-\vec{\tau}\right)\middle|\vec{k},j\right\rangle \quad .\end{aligned} \tag{48}$$

In the most general case, Eq. (46) gives

$$\left(\hat{H}_{\tilde{S}\cdot\vec{k}} - \varepsilon_{\tilde{S}\cdot\vec{k},j}\right)\hat{S}\left|u_{\vec{k},j}\right\rangle = \hat{S}\left(\hat{H}_{\vec{k}} - \varepsilon_{\vec{k},j}\right)\left|u_{\vec{k},j}\right\rangle = 0 \quad , \tag{49}$$

where the symmetry requirement $\varepsilon_{\tilde{S}\cdot\vec{k},j} = \varepsilon_{\vec{k},j}$ are employed. Thus, by choosing the phase to be 0, one has

$$\hat{S}\left|u_{\vec{k},j}\right\rangle = \left|u_{\tilde{S}\cdot\vec{k},j}\right\rangle \quad , \tag{50}$$

and derives

$$\left\langle \vec{x} \middle| u_{\tilde{S}\cdot\vec{k},j} \right\rangle = \left\langle \vec{x} \middle| \hat{S} \middle| u_{\vec{k},j} \right\rangle = \left\langle \tilde{S}^{-1}\cdot\left(\vec{x}-\vec{\tau}\right) \middle| u_{\vec{k},j} \right\rangle \quad . \tag{51}$$

Notice that the choice of phase here is to make Eq. (51) consistent with the result of Eq. (47). Finally, one obtains

$$\begin{aligned} \left\langle \vec{x} \middle| \tilde{S}\cdot\vec{k},j \right\rangle &= e^{i\left(\tilde{S}\cdot\vec{k}\right)\cdot\vec{x}} \left\langle \vec{x} \middle| u_{\tilde{S}\cdot\vec{k},j} \right\rangle = e^{i\left(\tilde{S}\cdot\vec{k}\right)\cdot\vec{\tau}} e^{i\vec{k}\cdot\left(\tilde{S}^{-1}\cdot\left(\vec{x}-\vec{\tau}\right)\right)} \left\langle \vec{x} \middle| \hat{S} \middle| u_{\vec{k},j} \right\rangle \\ &= e^{i\left(\tilde{S}\cdot\vec{k}\right)\cdot\vec{\tau}} \left\langle \tilde{S}^{-1}\cdot\left(\vec{x}-\vec{\tau}\right) \middle| \vec{k},j \right\rangle \end{aligned} \tag{52}$$

or,

$$\left\langle \vec{x} \middle| \tilde{S}^{-1}\cdot\vec{k},j \right\rangle = e^{-i\vec{k}\cdot\vec{\tau}} \left\langle \tilde{S}\cdot\vec{x}+\vec{\tau} \middle| \vec{k},j \right\rangle \quad . \tag{53}$$

Now, consider the case when state $\left|\vec{k},j\right\rangle$ is degenerate with $\left|\vec{k},i\right\rangle$ for $i\in D$. Eq. (50) is now

$$\hat{S}\left|u_{\vec{k},j}\right\rangle = \sum_{i\in D} c_{j,i} \left|u_{\tilde{S}\cdot\vec{k},i}\right\rangle \quad , \tag{54}$$

where $c_{j,i}$ is unitary. Following the same derivation, one concludes the relation similar to Eq. (53):

$$\left\langle \vec{x} \middle| \tilde{S}^{-1}\cdot\vec{k},j \right\rangle = e^{-i\vec{k}\cdot\vec{\tau}} \sum_{i\in D} c_{j,i} \left\langle \tilde{S}\cdot\vec{x}+\vec{\tau} \middle| \vec{k},i \right\rangle \quad . \tag{55}$$

In the special case when $\tilde{S}^{-1}\cdot\vec{k}\neq\vec{k}$ and $\vec{k}$ is in the irreducible wedge, redefinition of $\left|\tilde{S}^{-1}\cdot\vec{k},j\right\rangle$ as

$$\left|\tilde{S}^{-1}\cdot\vec{k},j\right\rangle' = \sum_{i\in D} c^{\dagger}_{j,i} \left|\tilde{S}^{-1}\cdot\vec{k},i\right\rangle \quad , \tag{56}$$

gives the same simple relation as Eq. (53).

A.2 Symmetry Properties of *G* in Bloch-τ space

The one-particle propagator, *G*, satisfies the anti-periodic condition in τ-space as a result of Fermi-Dirac statistics [54]:

$$\begin{aligned} G(\vec{x},\tau;\vec{x}',\tau') &= -G(\vec{x},\tau+\beta\hbar;\vec{x}',\tau') \\ &= -G(\vec{x},\tau;\vec{x}',\tau'+\beta\hbar) \quad . \end{aligned} \tag{57}$$

Furthermore, from the definition of *G* in Eq. (12), it is clear that *G* is real in $(\vec{x},\tau)$ space:

$$G(1,1') = G(1,1')^* \quad . \tag{58}$$

Now, since

$$\begin{aligned} \left[\hat{\psi}(\vec{x},\tau)\right]^\dagger &= \left[e^{(\hat{H}-\mu\hat{N})\tau/\hbar}\hat{\psi}(\vec{x})e^{-(\hat{H}-\mu\hat{N})\tau/\hbar}\right]^\dagger \\ &= e^{-(\hat{H}-\mu\hat{N})\tau/\hbar}\hat{\psi}^\dagger(\vec{x})e^{(\hat{H}-\mu\hat{N})\tau/\hbar} \\ &= \hat{\psi}^\dagger(\vec{x},-\tau) \\ &\neq \hat{\psi}^\dagger(\vec{x},\tau) \quad , \end{aligned} \tag{59}$$

it is straightforward to show, from Eq. (12), that

$$G(\vec{x},\tau;\vec{x}',\tau') = G(\vec{x}',-\tau';\vec{x},-\tau)^* = G(\vec{x}',-\tau';\vec{x},-\tau) \quad . \tag{60}$$

Now, represented in Bloch-τ space, *G* is

$$
\begin{aligned}
G_{\vec{k},j;\vec{k}',j'}(\tau,\tau') &= \left\langle \vec{k},j \middle| \bar{\vec{x}} \right\rangle G\left(\bar{\vec{x}},\tau;\bar{\vec{x}}',\tau'\right)\left\langle \bar{\vec{x}}' \middle| \vec{k}',j' \right\rangle \\
&= \left(\left\langle \vec{k}',j' \middle| \bar{\vec{x}}' \right\rangle G\left(\bar{\vec{x}}',-\tau';\bar{\vec{x}},-\tau\right)\left\langle \bar{\vec{x}} \middle| \vec{k},j \right\rangle\right)^* \\
&= \left[G_{\vec{k}',j';\vec{k},j}(-\tau',-\tau)\right]^* \quad .
\end{aligned} \tag{61}
$$

In equilibrium, when external potential is set to 0 and symmetries like crystal momentum and imaginary time invariance are respected,

$$
G_{\vec{k},j;\vec{k}',j'}(\tau,\tau') = G_{\vec{k},j,j'}(\tau-\tau')\,\delta_{\vec{k}'-\vec{k}} \quad , \tag{62}
$$

and thus Eq. (61) gives

$$
G_{\vec{k},j,j'}(\tau) = \left[G_{\vec{k},j',j}(\tau)\right]^* \quad . \tag{63}
$$

In this work, only diagonal elements of $G_{\vec{k},j,j'}(\tau)$ are considered, which are then real quantities due to Eq. (63):

$$
G_{\vec{k},j,j}(\tau) = \left[G_{\vec{k},j,j}(\tau)\right]^* \quad . \tag{64}
$$

It is also necessary to apply the rotation symmetry in order to gain information of *G* corresponding to *k* points outside the irreducible wedge. First, we note that the rotation symmetry is respected by *G* in $(\vec{x},\tau)$ space:

$$
G\left(\tilde{S}\cdot\bar{\vec{x}}+\vec{\tau},\tilde{S}\cdot\bar{\vec{x}}'+\vec{\tau};\tau'\right) = G\left(\bar{\vec{x}},\bar{\vec{x}}';\tau'\right) \quad , \tag{65}
$$

where $\vec{\tau}$ is the partial lattice vector while τ' is the Matsubara time. Thus, with Eq. (53),

$$
\begin{aligned}
G_{\tilde{S}^{-1}\cdot\vec{k},j,j'}(\tau') &= \left\langle \tilde{S}^{-1}\cdot\vec{k},\, j \middle| \bar{\bar{x}} \right\rangle G\left(\bar{\bar{x}},\bar{\bar{x}}';\tau'\right)\left\langle \bar{\bar{x}}' \middle| \tilde{S}^{-1}\cdot\vec{k}',\, j' \right\rangle \\
&= \left\langle \vec{k},\, j \middle| \tilde{S}\cdot\bar{\bar{x}}+\vec{\tau} \right\rangle G\left(\tilde{S}\cdot\bar{\bar{x}}+\vec{\tau},\tilde{S}\cdot\bar{\bar{x}}'+\vec{\tau};\tau'\right)\left\langle \tilde{S}\cdot\bar{\bar{x}}'+\vec{\tau} \middle| \vec{k}',\, j' \right\rangle \\
&= G_{\vec{k},j,j'}(\tau') \quad ,
\end{aligned}
\tag{66}
$$

where $\tilde{S}^{-1}\cdot\vec{k}$ is outside the irreducible wedge and $\vec{k}$ is in the wedge. The phases introduced in Eq. (53) are canceled out due to the identical *k* points considered here.

A.3 Symmetry Properties of *W* in Fourier-ω space

Unlike the one-particle propagator, *G*, discussed above, the screened interaction, *W*, satisfies the periodic condition in τ-space as a result of Bose-Einstein statistics [54]:

$$
\begin{aligned}
W\left(\vec{x},\tau;\vec{x}',\tau'\right) &= W\left(\vec{x},\tau+\beta\hbar;\vec{x}',\tau'\right) \\
&= W\left(\vec{x},\tau;\vec{x}',\tau'+\beta\hbar\right) \quad .
\end{aligned}
\tag{67}
$$

Furthermore, from the definition of *W* in Eq. (28), it is clear that *W* is real in $\left(\vec{x},\tau\right)$ space:

$$
W(1,2) = W(1,2)^{*} \quad . \tag{68}
$$

Now, it is also apparent from the definition of *W* that

$$
W(1,2) = W(2,1) \quad , \tag{69}
$$

since in Eq. (28)

$$
P(1,2) = G(1,2)\cdot G(2,1) = G(2,1)\cdot G(1,2) = P(2,1) \quad . \tag{70}
$$

With Eqs. (69) and (60), it is straight to show that

$$W\left(\vec{x},\tau;\vec{x}',\tau'\right)=W\left(\vec{x},-\tau;\vec{x}',-\tau'\right) \quad . \tag{71}$$

Eqs. (67) and (71) implied the following symmetry in τ-space:

$$\begin{aligned} W\left(\vec{x},\tau;\vec{x}',\tau'\right) &= W\left(\vec{x},\beta\hbar-\tau;\vec{x}',-\tau'\right) \\ &= W\left(\vec{x},-\tau;\vec{x}',\beta\hbar-\tau'\right) \\ &= W\left(\vec{x},\beta\hbar-\tau;\vec{x}',\beta\hbar-\tau'\right) \quad . \end{aligned} \tag{72}$$

Especially, in the presence of imaginary time invariance, the following additional inversion symmetry with respect to $\tau=\beta\hbar/2$ is useful as it reduces the amount of necessary stored information by $1/2$:

$$W\left(\vec{x},\vec{x}';\tau\right)=W\left(\vec{x},\vec{x}';\beta\hbar-\tau'\right) \quad . \tag{73}$$

In Fourier-τ space, W is Hermitian following Eq. (69):

$$\begin{aligned} W_{\vec{q},\vec{Q};\vec{q}',\vec{Q}'}\left(\tau,\tau'\right) &= \left\langle \vec{q}+\vec{Q}\middle|\bar{\vec{x}}\right\rangle W\left(\bar{\vec{x}},\tau;\bar{\vec{x}}',\tau'\right)\left\langle \bar{\vec{x}}'\middle|\vec{q}'+\vec{Q}'\right\rangle \\ &= \left(\left\langle \vec{q}'+\vec{Q}'\middle|\bar{\vec{x}}'\right\rangle W\left(\bar{\vec{x}}',\tau';\bar{\vec{x}},\tau\right)\left\langle \bar{\vec{x}}\middle|\vec{q}+\vec{Q}\right\rangle\right)^{*} \\ &= \left[W_{\vec{q}',\vec{Q}';\vec{q},\vec{Q}}\left(\tau',\tau\right)\right]^{*} \quad . \end{aligned} \tag{74}$$

In equilibrium, when external potential is set to 0 and symmetries like crystal momentum and imaginary time invariance are respected,

$$W_{\vec{q},\vec{Q};\vec{q}',\vec{Q}'}\left(\tau,\tau'\right)=W_{\vec{q},\vec{Q},\vec{Q}'}\left(\tau-\tau'\right)\delta_{\vec{q}-\vec{q}'} \quad , \tag{75}$$

and thus Eq. (74) gives

$$W_{\vec{q},\vec{Q},\vec{Q}'}\left(\tau\right)=\left[W_{\vec{q},\vec{Q}',\vec{Q}}\left(-\tau\right)\right]^{*}=\left[W_{\vec{q},\vec{Q}',\vec{Q}}\left(\tau\right)\right]^{*} \quad . \tag{76}$$

That is, the diagonal elements of *W* are real. Specially, systems studied in this work possess inversion symmetry with respect to the origin, which, with the proper choice of convention in defining the Bloch states, makes all the elements of *W* real. As a consequence, W is symmetric in Fourier space from Eq. (76).

It is also necessary to apply the rotation symmetry in order to gain information of *W* corresponding to *q* points outside the irreducible wedge. First, we note that the rotation symmetry is respected by *W* in $(\vec{x},\tau)$ space:

$$W\left(\tilde{S}\cdot\vec{\bar{x}}+\vec{\tau},\tilde{S}\cdot\vec{\bar{x}}'+\vec{\tau};\tau'\right)=W\left(\vec{\bar{x}},\vec{\bar{x}}';\tau'\right) \quad , \tag{77}$$

where $\vec{\tau}$ is the partial lattice vector while τ' is the Matsubara time. Thus, with Eq. (53),

$$\begin{aligned} W_{\tilde{S}^{-1}\cdot\vec{q},\tilde{S}^{-1}\cdot\vec{Q},\tilde{S}^{-1}\cdot\vec{Q}'}(\tau') &= \left\langle \tilde{S}^{-1}\cdot\left(\vec{q}+\vec{Q}\right)\middle|\vec{\bar{x}}\right\rangle W\left(\vec{\bar{x}},\vec{\bar{x}}';\tau'\right)\left\langle\vec{\bar{x}}'\middle|\tilde{S}^{-1}\cdot\left(\vec{q}+\vec{Q}'\right)\right\rangle \\ &= e^{i\left(\vec{q}+\vec{Q}\right)\cdot\vec{\tau}}\left\langle\vec{q}+\vec{Q}\middle|\tilde{S}\cdot\vec{\bar{x}}+\vec{\tau}\right\rangle W\left(\tilde{S}\cdot\vec{\bar{x}}+\vec{\tau},\tilde{S}\cdot\vec{\bar{x}}'+\vec{\tau};\tau'\right) \\ &\quad\cdot e^{-i\left(\vec{q}+\vec{Q}'\right)\cdot\vec{\tau}}\left\langle\tilde{S}\cdot\vec{\bar{x}}'+\vec{\tau}\middle|\vec{q}+\vec{Q}'\right\rangle \\ &= e^{i\left(\vec{Q}-\vec{Q}'\right)\cdot\vec{\tau}}W_{\vec{q},\vec{Q},\vec{Q}'}(\tau') \quad , \end{aligned} \tag{78}$$

where $\tilde{S}^{-1}\cdot\vec{q}$ is outside the irreducible wedge and $\vec{q}$ is in the wedge. The extra phase introduced in Eq. (78) is important to keep track of as it eventually "reduces" the strength of the self-energy evaluated in Eq. (27) compared to the case in which the phase is omitted.

B Numerical Treatments

B.1 Bloch States in Linearized Argumented Plane Wave (LAPW) basis

In this work, the Bloch states of crystal momentum k and band index j are expended in terms of linearized argumented plan waves (LAPW) and its "local orbital" (LO) counterpart [9,10]:

$$\left|\vec{k}, j\right\rangle = \sum_{\vec{K}} c_{\vec{k},j,\vec{K}} \left|\Phi_{\vec{k},\vec{K}}\right\rangle + \sum_{l,m} c^{LO}_{\vec{k},j,l,m} \left|\Phi^{LO}_{l,m}\right\rangle \tag{79}$$

where $\vec{K}$ denotes the reciprocal Bravais lattice vector, (l, m) denotes the angular momentum quantum numbers, and $\left|\Phi_{\vec{k},\vec{K}}\right\rangle$ and $\left|\Phi^{LO}_{l,m}\right\rangle$ are the LAPW and the LO, respectively, defined as

$$\left|\Phi_{\vec{k},\vec{K}}\right\rangle \equiv \begin{cases} \sum_{l,m} c^{\vec{k},\vec{K},t}_{l,m} \left|t, E_l, l, m\right\rangle + \sum_{l,m} \dot{c}^{\vec{k},\vec{K},t}_{l,m} \left|t, E_l, l, m^{\bullet}\right\rangle & ;\text{in MT\# } t \\ \left|\vec{k}+\vec{K}\right\rangle & ;\text{in interstitial} \end{cases} \tag{80}$$

and

$$\left|\Phi^{LO}_{l,m}\right\rangle \equiv \begin{cases} d^t_{l,m} \left|t, E_l, l, m\right\rangle + \dot{d}^t_{l,m} \left|t, E_l, l, m^{\bullet}\right\rangle + d''^t_{l,m} \left|t, E'_l, l, m\right\rangle & ;\text{in MT\# } t \\ 0 & ;\text{in interstitial} \end{cases} \tag{81}$$

where t is the "site" index, E_l and E_l' are energy parameters for linearization of the basis functions, MT denotes "muffin-tin" sphere centered at each atomic site, whose radius is artificially chosen, $\left|\vec{k}+\vec{K}\right\rangle$ is the normalized plane wave function:

$$\left\langle \vec{x}\left|\vec{k},\vec{K}\right.\right\rangle \equiv \frac{1}{\sqrt{\Omega}} e^{i\left(\vec{k},\vec{K}\right)\cdot\vec{x}} \tag{82}$$

and $\left|t,E_l,l,m\right\rangle$ and $\left|t,E_l,l,m^{\bullet}\right\rangle$ are "atomic" states as solution of Dirac equation, with parameter E_l, and its energy derivative. Here, Ω denotes the volume of the macro-crystal. Represented in real space and spinor notation (assuming for spin-up states):

$$\left\langle \vec{x}\left|t,E_l,l,m\right.\right\rangle = U_{l,m}^{E_l,t}\left(\vec{r}\equiv\vec{x}-\vec{x}_t\right) = \frac{1}{r}\begin{pmatrix} P_l^t\left(\vec{r}\right)\chi_{\uparrow}Y_{l,m}\left(\hat{r}\right) \\ -i\alpha Q_l^t\left(\vec{r}\right)\left(2\vec{s}\cdot\hat{r}\right)\chi_{\uparrow}Y_{l,m}\left(\hat{r}\right)\end{pmatrix} \quad , \tag{83}$$

where $\vec{x}_t$ is the position of atomic site t, α is the fine-structure constant, $\vec{s}$ is the spin operator, $\chi_{\uparrow}$ is the spin-up spinor, and P and Q are referred to as the major and minor component. $\left|t,E_l,l,m^{\bullet}\right\rangle$ has a similar representation with different value of P and Q. The coefficients $c_{l,m}^{\vec{k},\vec{K},t}$, $\dot{c}_{l,m}^{\vec{k},\vec{K},t}$, $d_{l,m}^t$, $\dot{d}_{l,m}^t$, and $d_{l,m}'^t$ are fixed by boundary conditions that ensure the continuity in value and slope of the basis functions in the boundary between the MTs and the interstitial.

Now, inserting the basis functions into Eq. (79) and write the Bloch states in real space, one obtains the following representation, which is the starting point of all the calculation performed in this work:

$$\left\langle \vec{x} \middle| \vec{k}, j \right\rangle = \frac{1}{\sqrt{\#}} \cdot \begin{cases} \sum_{l,m,a} A_{l,m,a}^{\vec{k},j,t} U_{l,m,a}^{t}\left(\vec{x}\right) & ; \text{ in MT\# } t \\ \frac{1}{\sqrt{\Omega_0}} \sum_{\vec{K}} C_{\vec{K}}^{\vec{k},j} e^{i\left(\vec{k}+\vec{K}\right)\cdot\vec{x}} & ; \text{ in interstitial} \end{cases} , \tag{84}$$

where Ω_0 denotes the volume of the unit cell, # denotes the total number of unit cells in the macro-crystal, and *a* is 0, 1, or 2, corresponding to terms of $\left|t, E_l, l, m\right\rangle$, $\left|t, E_l, l, m^{\bullet}\right\rangle$, or $\left|t, E_l', l, m\right\rangle$.

Finally, it is important to note that the LAPW basis set is not orthogonal. That is,

$$\left\langle \Phi_{\vec{k},\vec{K}'} \middle| \Phi_{\vec{k},\vec{K}} \right\rangle \neq \delta_{\vec{K},\vec{K}'} \quad , \tag{85}$$

even though the resulting Bloch states do form a complete and orthonormal set.

B.2 Evaluation of Transition Probability Amplitude

Renormalized transition probability amplitude, $\begin{pmatrix} \left\langle \vec{q}+\vec{Q} \right| \\ \left\langle \vec{k}, j \right| \\ \left| \vec{k}', j' \right\rangle \end{pmatrix}$, defined as

$$\begin{aligned} \begin{pmatrix} \left\langle \vec{q}+\vec{Q} \right| \\ \left\langle \vec{k}, j \right| \\ \left| \vec{k}', j' \right\rangle \end{pmatrix} &\equiv \sqrt{\frac{v}{\Omega}} \left\langle \vec{k}, j \middle| e^{-i\left(\vec{q}+\vec{Q}\right)\cdot\hat{\vec{x}}} \middle| \vec{k}', j' \right\rangle \\ &= \sqrt{\frac{4\pi}{\Omega}} \frac{1}{\left|\vec{q}+\vec{Q}\right|} \int_{\Omega} d^3x \left\langle \vec{k}, j \middle| \vec{x} \right\rangle e^{-i\left(\vec{q}+\vec{Q}\right)\cdot\vec{x}} \left\langle \vec{x} \middle| \vec{k}', j' \right\rangle \quad , \end{aligned} \tag{86}$$

is employed both in the linear response calculation and the quasi-particle excitation calculation, where $\left|\vec{k}, j\right\rangle$ is the Bloch state corresponding to crystal momentum $\vec{k}$ and band index *j*, $\vec{q}$ is the wave vector, $\vec{Q}$ is the reciprocal Bravais lattice vector, and Ω is the volume of the macro-crystal. Due to the translation symmetry of the craytal, only $\vec{k}'$ satisfying

$$\vec{k}' = \vec{k} + \vec{q} - \vec{K} \quad , \tag{87}$$

allows non-zero value for the renormalized transition probability amplitude, where $\vec{K}$ is a reciprocal Bravais lattice vector. In the convention chosen in the current work, $\vec{k}$ and $\vec{k}'$ are chosen to be in the parallelepiped zone, defined as $\left\{\vec{k}; 0 \le k_i < 1\right\}$ in unit of the reciprocal Bravais lattice vector. On the other hand, $\vec{q}$ is chosen to be in the first Brillouin zone. The inclusion of the $\sqrt{v}$ factor is to take into account properly the $\vec{q} + \vec{Q} \to 0$ limit (optical limit), which is discussed in appendix B.2.3.

The evaluation of the renormalized transition probability amplitude can be reduced to only quantities within the unit cell in real space, since

$$
\begin{aligned}
\begin{pmatrix} \left\langle \vec{q}+\vec{Q} \right| \\ \left\langle \vec{k}, j \right| \\ \left| \vec{k}', j' \right\rangle \end{pmatrix} &= \sqrt{\frac{4\pi}{\Omega}} \frac{1}{\left| \vec{q}+\vec{Q} \right|} \int_{\Omega} d^3x \left\langle \vec{k}, j \middle| \vec{x} \right\rangle e^{-i\left(\vec{q}+\vec{Q}\right)\cdot\vec{x}} \left\langle \vec{x} \middle| \vec{k}', j' \right\rangle \\
&= \sqrt{\frac{4\pi}{\Omega}} \frac{1}{\left| \vec{q}+\vec{Q} \right|} \sum_{\vec{R}} \int_{\Omega_0} d^3x \left\langle \vec{k}, j \middle| \vec{x}+\vec{R} \right\rangle e^{-i\left(\vec{q}+\vec{Q}\right)\cdot\left(\vec{x}+\vec{R}\right)} \left\langle \vec{x}+\vec{R} \middle| \vec{k}', j' \right\rangle \\
&= \sqrt{\frac{4\pi}{\Omega}} \frac{1}{\left| \vec{q}+\vec{Q} \right|} \int_{\Omega_0} d^3x \left\langle \vec{k}, j \middle| \vec{x} \right\rangle e^{-i\left(\vec{q}+\vec{Q}\right)\cdot\vec{x}} \left\langle \vec{x} \middle| \vec{k}', j' \right\rangle \sum_{\vec{R}} e^{i\left(\vec{k}'-\vec{k}-\vec{q}-\vec{Q}\right)\cdot\vec{R}} \\
&= \sqrt{\frac{4\pi}{\Omega}} \frac{\#}{\left| \vec{q}+\vec{Q} \right|} \int_{\Omega_0} d^3x \left\langle \vec{k}, j \middle| \vec{x} \right\rangle e^{-i\left(\vec{q}+\vec{Q}\right)\cdot\vec{x}} \left\langle \vec{x} \middle| \vec{k}', j' \right\rangle \quad ,
\end{aligned}
\tag{88}
$$

where # is the number of the unit cells in the macro-crystal. In the last step of Eq. (88), Eq. (87) is employed and the following property is used:

$$
e^{i\vec{K}\cdot\vec{R}} = 1 \quad . \tag{89}
$$

For the calculations performed in this work, only information for either $\vec{k}$ or $\vec{q}$ in the irreducible wedge is necessary. These two different cases are discussed in the following sections.

B.2.1 For *q* in the Irreducible Wedge

For the case that $\vec{q}$ is in the irreducible wedge, all $\vec{k}$ points in the whole zone are needed. The following formula from appendix A.1 is useful to generate Block states outside the irreducible wedge, as only information within the irreducible wedge are available from the ground state calculation:

$$\left\langle \tilde{S}^{-1}\cdot\vec{k},\, j \middle| \vec{x} \right\rangle = e^{i\vec{k}\cdot\vec{\tau}} \left\langle \vec{k},\, j \middle| \tilde{S}\cdot\vec{x}+\vec{\tau} \right\rangle$$
$$\left\langle \vec{x} \middle| \tilde{S}'^{-1}\cdot\vec{k}',\, j' \right\rangle = e^{-i\vec{k}'\cdot\vec{\tau}'} \left\langle \tilde{S}'\cdot\vec{x}+\vec{\tau}' \middle| \vec{k}',\, j' \right\rangle \quad , \tag{90}$$

where

$$\tilde{S}^{-1}\cdot\vec{k}+\Delta\vec{K} = \dot{\vec{k}}$$
$$\tilde{S}'^{-1}\cdot\vec{k}'+\Delta\vec{K}' = \dot{\vec{k}}' = \dot{\vec{k}}+\vec{q}-\Delta\vec{K}'' \quad , \tag{91}$$

while $\dot{\vec{k}}$ and $\dot{\vec{k}}'$ are *k*-points outside the irreducible wedge, which correspond to $\vec{k}$ and $\vec{k}'$ in the irreducible wedge.

Separating the unit cell into "muffin-tin" regions and the interstitial region, as described in appendix B.1, and expressing exponential in the muffin-tin region in spherical harmonics, Y_{lm}, at atomic site *t* located at $\vec{x}_t$:

$$e^{-i\left(\vec{q}+\vec{Q}\right)\cdot\vec{x}} = e^{-i\left(\vec{q}+\vec{Q}\right)\cdot\left(\vec{r}+\vec{x}_t\right)} = e^{-i\left(\vec{q}+\vec{Q}\right)\cdot\vec{x}_t}\left[4\pi\sum_{lm} i^l\, j_l\left(\left|\vec{q}+\vec{Q}\right|\right) Y_{lm}^*\left(\widehat{-\vec{q}-\vec{Q}}\right) Y_{lm}\left(\hat{r}\right)\right] \quad , \tag{92}$$

where $j_l\left(\left|\vec{q}+\vec{Q}\right|\right)$ is the spherical Bessel function, one then obtains the following contributions:

1) in muffin-tin # *t*:

$$\sqrt{\frac{4\pi}{\Omega}}\frac{4\pi}{\left|\vec{q}+\vec{Q}\right|} i^{\bar{l}''} e^{i\left(\vec{k}\cdot\left(\vec{\tau}-\vec{R}_{t_S}\right)-\vec{k}'\cdot\left(\vec{\tau}'-\vec{R}_{t_{S'}}\right)-\left(\vec{q}+\vec{Q}\right)\cdot\vec{x}_t\right)} \cdot Y_{\bar{l}'',\bar{m}''}^*\left(-\tilde{S}\cdot\left(\vec{q}+\vec{Q}\right)\right)\cdot\left[D_{\bar{m}',\bar{m}''}^{\bar{l}'}\left(\tilde{S}\right)\right]^*$$
$$\cdot Gaunt\left(\begin{matrix}\bar{l}' & \bar{l}'' \\ \bar{m}''' & \bar{m}''\end{matrix}\middle|\begin{matrix}\bar{l} \\ \bar{m}\end{matrix}\right)\cdot jru\left(\left|\vec{q}+\vec{Q}\right|, t, \bar{l}''; \begin{matrix}\bar{l} & \bar{l}' \\ \bar{a} & \bar{a}'\end{matrix}\right)\cdot\left(A_{\bar{l},\bar{m},\bar{a}}^{\vec{k},j,t_S}\right)^* A_{\bar{l}',\bar{m}',\bar{a}'}^{\vec{k}',j',t_{S'}} \quad , \tag{93}$$

where bars above variables indicate summing over the variables, $\vec{x}_t$ is the position of *t*-th atomic site, and $\vec{R}_{t_S}$ and $\vec{R}_{t_{S'}}$ are Bravais lattice vectors defined as

$$\begin{aligned} \tilde{S}\cdot\vec{x}_t+\vec{\tau}&=\vec{x}_{t_S}+\vec{R}_{t_S} \\ \tilde{S}'\cdot\vec{x}_t+\vec{\tau}'&=\vec{x}_{t_{S'}}+\vec{R}_{t_{S'}} \quad , \end{aligned} \tag{94}$$

where t_S and $t_{S'}$ are also defined in Eq. (94) as site indices for sites within the unit (parallelepiped) cell in real space. The *jru*() function in Eq. (93) is defined as

$$\begin{aligned} jru\left(\left|\vec{q}+\vec{Q}\right|,t,l'';\begin{matrix}l\ l'\\ a\ a'\end{matrix}\right)&=jru\left(\left|\vec{q}+\vec{Q}\right|,atom,l'';\begin{matrix}l\ l'\\ a\ a'\end{matrix}\right) \\ &\equiv\int_0^{r_t^{MT}}dr\left[P_{l,a}^{atom}(r)P_{l',a'}^{atom}(r)+\alpha^2Q_{l,a}^{atom}(r)Q_{l',a'}^{atom}(r)\right]\cdot j_{l''}\left(\left|\vec{q}+\vec{Q}\right|r\right) \quad , \end{aligned} \tag{95}$$

since the rotation symmetry guarantees the identity of *P* and *Q* function for symmetry-related sites. Here the following identity is utilized to suppress the spinor degree of freedom in the Bloch states:

$$\left(2\vec{s}\cdot\hat{r}\right)^{\dagger}\cdot\left(2\vec{s}\cdot\hat{r}\right)=1 \tag{96}$$

The *Gaunt* coefficients in Eq. (93) are defined as

$$Gaunt\left(\begin{matrix}l'\ \ l''\\ m'''\ m''\end{matrix}\middle|\begin{matrix}l\\ m\end{matrix}\right)\equiv\int d\hat{r}\,Y_{l',m''}(\hat{r})Y_{l'',m''}(\hat{r})Y_{l,m}^{*}(\hat{r}) \tag{97}$$

and $D_{m',m''}^{l'}\left(\tilde{S}\right)$ is the rotational *D*-matrix [84] corresponding to the Euler angles equivalent to the rotation of $\tilde{S}$.

2) in the interstitial region:

$$\sqrt{\frac{4\pi}{\Omega}}\frac{1}{\left|\vec{q}+\vec{Q}\right|}e^{i\left(\vec{k}'\cdot\vec{\tau}'-\vec{k}\cdot\vec{\tau}\right)}\cdot\left(C_{\vec{K}}^{\vec{k},j}\right)^{*}C_{\vec{K}'}^{\vec{k}',j'}$$

$$\cdot\begin{cases}1-\dfrac{4\pi}{3\Omega_0}\sum_t\left(r_t^{MT}\right)^3 & ;\ \text{if}\ \ K''\equiv\left|\vec{K}''\right|=0\\ -\dfrac{4\pi}{\Omega_0}\dfrac{1}{K''^3}\sum_t e^{i\vec{K}''\cdot\vec{x}_t}\left[\sin\left(K''r_t^{MT}\right)-K''r_t^{MT}\cos\left(K''r_t^{MT}\right)\right] & ;\ \text{otherwise}\end{cases} \tag{98}$$

where $\vec{K}''\equiv-\tilde{S}^{-1}\cdot\left(\vec{k}+\vec{K}\right)-\left(\vec{q}+\vec{Q}\right)+\tilde{S}'^{-1}\cdot\left(\vec{k}'+\vec{K}'\right)$.

Finally, adding these contributions together, one obtains the desired renormalized transition probability amplitude.

B.2.2 For *k* in the Irreducible Wedge

The formula for evaluating the renormalized transition probability amplitude in the case when $\vec{k}$ is in the irreducible wedge is very similar to the one discussed above except that the external variable $\vec{q}$ now can be any point in the first Brillouin zone. More importantly there is no need to map $\vec{k}$ back to the irreducible wedge. That is, only the second line of Eq. (90) is necessary, which consequently changes slightly the formula given above in Eqs. (93) and (98).

B.2.3 Special Treatment of $q \rightarrow 0$ Limit

The inclusion of the $\sqrt{v}$ factor in the definition of transition probability amplitude allows proper account of the $\vec{q}+\vec{Q}\rightarrow 0$ limit (optical limit) as the amplitude remains

finite while $\sqrt{v}$ becomes singular. Numerically, the singular behavior needs to be avoided explicitly in the formula employed. Following the secular equation Eq. (43):

$$\left[\hat{H}+\frac{\hbar\vec{k}}{m}\cdot\hat{\vec{p}}+\frac{\hbar^2k^2}{2m}-E_{\vec{k},j}\right]\left|u_{\vec{k},j}\right\rangle=0 \quad , \tag{99}$$

where $\vec{k}$ is considered an external variable, and introducing small perturbation as $\vec{k}\to\vec{k}+\vec{q}$:

$$\begin{aligned} V_{\vec{q}} &= \frac{\hbar\left(\vec{k}+\vec{q}\right)}{m}\cdot\hat{\vec{p}}+\frac{\hbar^2\left(\vec{k}+\vec{q}\right)^2}{2m}-\left(\frac{\hbar\vec{k}}{m}\cdot\hat{\vec{p}}+\frac{\hbar^2k^2}{2m}\right) \\ &= \frac{\hbar\vec{q}}{m}\cdot\left(\hat{\vec{p}}+\hbar\vec{k}\right) \quad , \end{aligned} \tag{100}$$

the perturbation theory for non-degenerate states, with respect to band index j, gives:

$$\begin{aligned} e^{-i\vec{q}\cdot\hat{\vec{x}}}\left|\vec{k}+\vec{q},j\right\rangle &= e^{-i\vec{k}\cdot\hat{\vec{x}}}\left|u_{\vec{k}+\vec{q},j}\right\rangle \\ &= e^{-i\vec{k}\cdot\hat{\vec{x}}}\left[\left|u_{\vec{k},j}\right\rangle+\sum_{j'\neq j}\frac{\left\langle u_{\vec{k},j'}\left|\hat{V}_{\vec{q}}\right|u_{\vec{k},j}\right\rangle}{E_{\vec{k},j}-E_{\vec{k},j'}}\left|u_{\vec{k},j'}\right\rangle\right] \\ &= \left|\vec{k},j\right\rangle+\sum_{j'\neq j}\frac{\left\langle u_{\vec{k},j'}\left|\hat{V}_{\vec{q}}\right|u_{\vec{k},j}\right\rangle}{E_{\vec{k},j}-E_{\vec{k},j'}}\left|\vec{k},j'\right\rangle \\ &= \left|\vec{k},j\right\rangle+\frac{\hbar\vec{q}}{m}\cdot\sum_{j'\neq j}\frac{\left\langle\vec{k},j'\left|\hat{\vec{P}}\right|\vec{k},j\right\rangle}{E_{\vec{k},j}-E_{\vec{k},j'}}\left|\vec{k},j'\right\rangle \quad , \end{aligned} \tag{101}$$

where

$$\left\langle u_{\vec{k},j'}\left|\hat{V}_{\vec{q}}\right|u_{\vec{k},j}\right\rangle=\frac{\hbar\vec{q}}{m}\cdot\left\langle\vec{k},j'\left|\hat{\vec{P}}\right|\vec{k},j\right\rangle \quad , \tag{102}$$

which is straightforward to prove from the definition of $\hat{V}_{\vec{q}}$ in Eq. (100), is employed in the last step. Finally, one obtains the following formula for the $\vec{q} \to 0$ limit:

$$\left\langle \vec{k}, j \left| e^{-i\vec{q}\cdot\hat{\vec{x}}} \right| \vec{k}+\vec{q}, j \right\rangle = \delta_{j,j'} + \left(1-\delta_{j,j'}\right) \frac{\hbar\vec{q}}{m} \cdot \frac{\left\langle \vec{k}, j \left| \hat{\vec{P}} \right| \vec{k}, j' \right\rangle}{E_{\vec{k},j'} - E_{\vec{k},j}} \quad . \tag{103}$$

The extension of Eq. (103) to the case with degenerate states is straightforward:

$$\left\langle \vec{k}, j \left| e^{-i\vec{q}\cdot\hat{\vec{x}}} \right| \vec{k}+\vec{q}, j \right\rangle = \delta_{j,j'} + \left(\text{if } j' \notin D_j\right) \frac{\hbar\vec{q}}{m} \cdot \frac{\left\langle \vec{k}, j \left| \hat{\vec{P}} \right| \vec{k}, j' \right\rangle}{E_{\vec{k},j'} - E_{\vec{k},j}} \quad , \tag{104}$$

where D_j is the group formed by states degenerate to *j*, for a fixed $\vec{k}$. Note that the singularity in the renormalized transition probability amplitude, due to the inclusion of $\sqrt{v}$, is now canceled with the $\frac{\hbar\vec{q}}{m}$ factor in Eqs. (103) and (104).

B.3 Analytic Continuation via Padé Approximants

Analytic continuation via Padé approximants is a well-established technique [13-16]. In the studies of QP excitations of this work, the use of analytic continuation is due to the choice of finite-temperature many-body perturbation theory formalism. In the studies of dynamical linear response of this work, however, it is employed as a numerical technique to better sample the Brillouin zone by avoiding summing over the δ-functions in the imaginary part of Eq. (10).

The basic idea of Padé approximants is to fit the unknown function in the complex plane by a rational function:

$$f(x)=\frac{P_\mu(x)}{Q_\nu(x)}=\frac{p_0+p_1x+\cdots+p_\mu x^\mu}{q_0+q_1x+\cdots+q_\nu x^\nu}=\frac{(x-z_0)(x-z_1)\cdots(x-z_\mu)}{(x-s_0)(x-s_1)\cdots(x-s_\nu)} \quad , \qquad (105)$$

where μ and ν are the number of zero's and singular points of the functions. A efficient recursive algorithm to find the coefficients in the rational function was given in Ref. [13].

It is clear this rational function form is ideal for studying physical features presenting as a peak (on real frequency-axis), as a pole slightly below the real frequency axis would be capable of generating the line shape of a peak on the real frequency. That is certainly the case for the linear response for system with a well-defined plasmon peak, as well as for the "normal" QP excitation as the QP peak approaches to a δ-function close to the chemical potential.

There are several limitations of numerical analytic continuation methods due to the limited number of effective digits. One of the most serious limitations is the sensitivity of the resulting function evaluated at points far away from the imaginary axis, where the original data points reside, to the tiny difference in the original data. This problem has recently been addressed by Beach *et al*. [44] with a symbolic computation algorithm, with which an arbitrary number of effective digits can be achieved. It was shown that if enough effective digits is kept in the calculation, the Padé approximants can faithfully recover the analytical from of the functions of interest.

There are no real rules on how to choose the number of the original data points, nor on how large the imaginary frequency should be in order to cover a certain real frequency range. Clearly, the number of original data controls the number of zero and poles in the resulting rational functions. When the number of the original data points is too small, the resulting function form is not capable of describing detail features other than the main (rough) line shape. On the other hand, when the number of the original data points is too large, the resulting function would simply possess many "redundant" pair of zeros and poles (which can be analytically cancelled); as a result, the line shape on real frequency axis would sometimes present strong oscillation, resulting from a pair of zero and pole very close to each other.

As for the energy range of the original data points, if the original data points correspond to imaginary energy range that is too small (compare to the real frequency of the physical features in the function), they will not be sensitive enough to features at large real frequency. On the other hand, if the original data points contain points corresponding to very large imaginary energy, a small round-off error in those high imaginary frequency points may introduce large error in the resulting functions.

In the actual implementation of this work, usually 801 (or more) original data points on imaginary frequency axis are employed to construct the rational function, before evaluating the resulting function on (slightly higher than) the real frequency axis. The number of original data points is chosen to give as much detail as possible in (only) the energy range of interest, while the number is not too large to give many singular (strongly oscillating) behaviors in the results. As for the choice of imaginary frequency

cutoff, it is found, from experience accumulated, that the energy should be chosen to about 1.5 times larger than the largest real frequency of physical features (peaks), not the frequency of the feature of interest. For example, in the cases of linear response, in which semi-core states contribute to the polarization and thus demonstrate a broad peak, in order to faithfully describe the line shape of the plasmon peak, imaginary frequency cutoff should be chosen as about 1.5 times the semi-core binding energy (not the plasmon energy), even thought the binding energy may be three times larger than the plasmon energy.

A useful trick found in the present work is that since the result of analytic continuation depends on the order of the data points being sent to the recursive algorithm, one may estimate the "stability" of the analytic continuation by comparing line shapes on the real frequency axis from different order of the original data. It is also found that usually an order like 0, N/2, 1, N/2+1, 2, …, N/2-1, N, is very stable.

Another useful trick found in this work is to compare results from Eq. (8) on real axis, calculated with 1) evaluation of Eq. (10) on imaginary axis followed by analytic continuation before solving Eq. (8), and 2) evaluation of Eq. (10) and solving Eq. (8) on imaginary axis followed by analytic continuation at the end. The results from these two approaches should be very similar, if the numerical analytic continuation is stable. Otherwise, adjustment on the imaginary cutoff, number of original data points, or the order of the original data is necessary in order to achieve a reliable result on the real frequency axis.

It is important to note that even if the stability of the analytic continuation is checked with the tricks mentioned above, inaccuracy introduced from the limited number of effective digits in the original data cannot be ruled out. Therefore, features with higher frequency than the first real peak in the resulting line shape need to be considered more carefully, as they may suffer from the "poor" quality of the original data.

B.4 Matsubara Time Approach

The formulation of finite temperature many-body perturbation theory are naturally derived on the imaginary time (Matsubara time) axis and its Fourier transform imaginary frequency (Matsubara frequency) axis [45-48], due to the grand canonical ensemble employed in the formalism. The connection to physical systems is then established through analytical continuation to the real frequency that corresponds to the physical energy for quantities of interest. Traditionally, numerical implementations of the theory are performed only on the Matsubara frequency axis due to its simplicity, explicit respect to energy conservation, and direct connection to physical interpretation. Moreover, several numerical difficulties that will be addressed in this section may have made it less trivial to implement the formalism on the Matsubara time axis.

B.4.1 Motivation for Matsubara Time Approach

One of the main drawbacks of Matsubara frequency approach in actual numerical implementation is the discrete summation over infinite number of Matsubara frequency, ω_n:

$$\omega_n = \begin{cases} \frac{\pi}{\beta\hbar}\cdot(2n+1) & \text{; for fermion operators} \\ \frac{\pi}{\beta\hbar}\cdot 2n & \text{; for boson operators} \end{cases} \tag{106}$$

, where $n = 0,1,\cdots,\infty$. A typical example is the following convolution of *P* as part of evaluation of *W*:

$$P(1,2) \equiv G(1,2)\cdot G(2,1) \tag{107}$$

performed in frequency space (omitting the space variables from this point):

$$P(\omega_n) = \sum_{i=0}^{\infty} G(\omega_n - \omega_i)\cdot G(\omega_i) \quad . \tag{108}$$

Clearly, a direct numerical evaluation of such quantity is impractical. Moreover, due to the slow decay of *G* for large ω, summing only finite number of terms up to a cutoff frequency introduces large error in the resulting *P* [67]. Similar structure of summing over infinite number of terms appears in several other places in the conserving GWA calculation, including the evaluation of the self-energy and the occupation number for the determination of the chemical potential.

Several techniques have been employed by previous authors to avoid this numerical difficulty of infinite sum. These include: 1) contour integral with residues

[85], 2) analytical modeling of the sum beyond a certain cutoff frequency [67,86], and 3) renormalization group with decreasing temperature [87]. For the calculations performed in this work, the first approach is not suitable due to the choice of Bloch representation chosen in this work and the self-consistent requirement. The second method is in disadvantage due to the difficulty in modeling after the first iteration. That is, once entering the second iteration, in which the first dressed *G* is employed, this approach are not easily applicable[*]. The third technique is designed to systematically lower the temperature of the system (for the purpose of obtaining the phase diagram, for example) and thus is expensive for the present work as only results in *one* specific low temperature are of interest.

This numerical difficulty on summing over infinite number of Matsubara frequency is naturally overcome in the Matsubara time approach as 1) Matsubara time is bounded: $\tau \in [0, \beta\hbar]$ and 2) τ is a continuous variable. Therefore, a well-designed mesh that allows fast numerical integral (and fast interpolation, as discussed below) over the finite range is enough to treat the formalism accurately and efficiently.

[*] The situation here is different from the case of the Hubbard model, in which only bands close to the Fermi energy are considered. To obtain an accurate *P*, inclusion of highly excited states and semi-core states is necessary. For these states, the cutoff frequency that allows the analytical modeling needs to be much larger than their QP energy.

Furthermore, all the frequency convolutions involving summing over infinite number of ω_i is transformed into product in τ axis (without summation). For example, the above convolution for *P* becomes

$$P(\tau) = G(\tau) \cdot G(-\tau) \quad . \tag{109}$$

Also, the occupation number is now directly read from *G* as

$$n_{\vec{k},j} = G_{\vec{k},j}\left(\tau = 0^-\right) \quad , \tag{110}$$

which is nothing but the value of *G* at the mesh point corresponding to $\tau = 0^-$.

Finally, within the approximations employed in this work, on Matsubara time axis, all the quantities involved in the calculations are real and thus only real-number algorithms are necessary, as opposed to the complex number algorithms required in the case of Matsubara frequency approaches. This gives about a factor of five times difference in performance of the basic multiplication operations heavily involved in the calculations. The required amount of memory is also reduced by ~1/2 because of the same reason.

B.4.2 Numerical Challenges in Matsubara Time Approach

Even though the numerical difficulties mentioned above is clearly circumvented in the Matsubara time approach, new numerical challenges appear in the τ axis approach. First, as illustrated in Figure 17, functions on τ axis usually present extreme

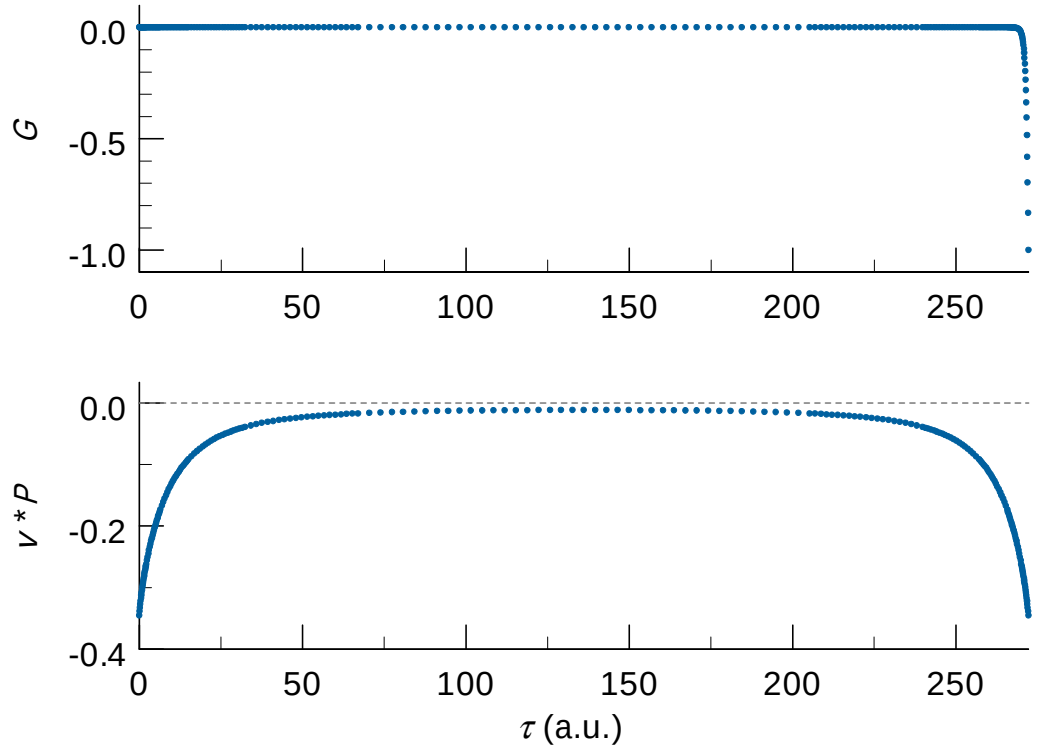


Figure 17 Localization Nature of Functions of τ

Demonstration of extreme localization of functions on τ axis. Examples of G (upper panel) and v*P (lower panel) are plotted along τ axis. Note the extreme localization near $\tau = 0$ and $\tau = \beta\hbar$. Also notice the discontinuity of G at $\tau = 0$ and $\tau = \beta\hbar$, as well as the slope of P at the same limits, knowing that $G(\tau - \beta\hbar) = -G(\tau)$ and $P(\tau - \beta\hbar) = P(\tau)$.

localization near $\tau = 0$ or $\tau = \beta\hbar$. (This is essentially the origin of the slow decay of *G* mentioned above) Therefore, a *non-uniform* sampling mesh would be desired to reduce the numerical expense.

Second, Figure 17 also demonstrates the large discontinuity of *G* and the slope of *P* at $\tau = 0$ and $\tau = \beta\hbar$. As a result, explicit sampling at $\tau = 0^+$ and $\tau = \beta\hbar^-$ (related to $\tau = 0^-$) is necessary. This also renders the use of the fast Fourier transform (FFT) algorithm unreliable, as FFT tends to smooth out the discontinuity and thus changes the value near these end points (e.g.: occupation number).

Third, integrals in the Matsubara time formalisms involve products of functions whose origin in time is shifted: for example, product of *P* and *W* in

$$W(\tau) = v \cdot \delta(\tau) + \int_0^{\beta\hbar} v \cdot P(\tau - \tau') \cdot W(\tau') d\tau' \quad . \tag{111}$$

In order to numerically perform the products, matching mesh points in these functions require the use of *uniform* mesh. This seems incompatible with the first point given above and thus is the center point of developing uniform-power mesh (UPM) discussed in appendix B.5.

Another numerical challenge is to reduce the expense of solving integral equations appearing in the τ-axis formalism, as opposed to the linear matrix equation in ω_n-axis formalism. The iterative linear solver is ideal for this purpose because 1) physically a good initial guess of the solutions is already available, and 2) only the algorithm that evaluates matrix-vector multiplication is necessary, which is equivalent to an integral in

τ-axis, which can be evaluated rather efficiently (and accurately) with high-order Simpson methods. In the actual implementation in this work, the preconditioned Bi-CG_Stab algorithm [59] is employed in the case of non-symmetric "matrix" while the preconditioned CG_Stab algorithm is employed for symmetric "matrix".

B.5 Uniform-Power Mesh (UPM)

As discussed above, the key numerical task is to satisfy both requirements: 1) non-uniform mesh to reduce expense of numerical sampling, and 2) uniform mesh to allow product of functions that have different origin in τ-axis. In this work, a specially mesh, named "uniform-power mesh" (UPM), is developed to satisfy these two requirement by 1) increasing mesh density exponentially to adapt to the extreme localization nature of the functions, and 2) allowing a fast interpolation scheme to build a uniform mesh when it is necessary.

UMP is constructed with a specific number of "power interval", labeled by 'p', in order to "catch" the sharp localized nature of the functions of interest as each interval is twice as large as the previous one (the reason for using 2 as basis will be clear below when fast interpolation is discussed.) After that, each interval is divided into certain number of uniform interval, labeled by 'u', to fill in more sampling points.

Three examples of UPM are illustrated in Figure 18. The first one, p1u10, is a special case identical to a uniform mesh (p = 1 → no "power" effect). The second one, p4u1, is the other extreme case (power mesh only) that has no "uniform" effect (u = 1).

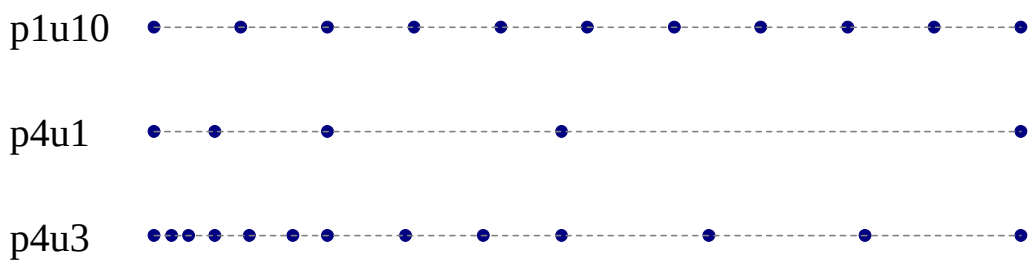


Figure 18 Demonstration of Uniform-Power Mesh

Demonstration of Uniform-Power Mesh. Examples of UPM are shown here. The first one, p1u10, is a special case that is identical to a uniform mesh (no "power" effect). The second one, p4u1, is the other extreme case (power mesh) that has no "uniform" effect. The third one, p4u3, is the general UPM that contains 4 power intervals (same as the second case) with 3 uniform intervals in between.

The third one, p4u3, is the general case of UPM that contains 4 power intervals (same as the second case) with 3 uniform intervals in between. The capability of the UPM should now be apparent, as it is flexible enough to be constructed for extremely localized functions, extremely extended ones, or intermediate ones, by simply choosing the proper parameters of 'p' and 'u'. Thus, use of UPM satisfies the first requirement stated above.

To build a uniform mesh from UPM is straightforward with extension of the fast high-order interpolation technique widely employed in second-generation wavelet transform [88]. The idea of the fast high-order interpolation is to utilize the knowledge that the location of the point being interpolated is fixed with a specific ratio between the nearest neighboring points (1:1, in the case of the UPM implemented). For example, a third order interpolation of $f(0)$ from the information of $f(-3)$, $f(-1)$, $f(1)$, and $f(3)$ is

$$f(0) = \frac{-1}{16} f(-3) + \frac{9}{16} f(-1) + \frac{9}{16} f(1) + \frac{-1}{16} f(3) \quad . \tag{112}$$

That is, only a list of coefficients representing the "weight" for each known neighboring point is necessary. This approach is much cheaper than the general routines that are designed to handle arbitrary point between the neighboring points, especially in the case of higher order, as this approach scales almost linearly to the order of interpolation. Of course, in the case of UPM, more than one list of weight is necessary as several situations involving different distance between neighboring points are encountered.

Now, it should be clear why in the construction of UPM, a factor of 2 increase is chosen in the power as it allows matching mesh points between two functions with shifted origin (see discussion below Eq. (111)) after employing the fast high-order interpolation to achieve a uniform mesh.

It should be noted that due to the efficiency of sampling provided by the UPM, only results of integral (or solutions of the integral equations) that corresponds to the UPM are necessary. The information between these mesh points can then be obtained through the fast high-order interpolation again if necessary. This greatly reduces the expense of the calculation as a small number of sampling points in UPM is usually good enough. For example, a p6u10 UPM uses only 61 points but is equivalent to a 641 uniform mesh in accurately representing the functions in τ-axis. Also notice that due to the high density near the end points in τ-axis, features at high frequency is automatically taken into account. Thus, the difficulty of infinite summation in ω_n-axis is implicitly solved.

Finally, it is worth mentioning several other powerful capabilities of the Matsubara time approach with the help of carefully constructed UPM: 1) to lower temperature, 2) to increase energy range, d 3) to be applied beyond the GWA, and 4) to be potentially more efficient. The more traditional ω_n-axis approaches suffer from the limited number of uniform ω_n mesh, which needs to increase linearly as temperature decreases. More seriously, the number of ω_n needed increase exponentially as the QP energy of the considered bands increases, and thus makes it very demanding to treat highly

excited states or semi-core states. Both above limitations are implicitly circumvented when a dense mesh points are constructed with UPM. Furthermore, UPM is not designed to work specific for the GWA (unlike the large ω_n modeling approach) and therefore should be applicable in calculation involving more complicated self-energy functional (e.g.: GWA with "vertex correction"). Finally, as pointed out in above, better efficiency in performance and memory usage is a direct consequence of the Matsubara time approaches as only real quantities are involved within the approximations employed in this work.

C List of Publications

1. Comment on "Why is the Bandwidth of Sodium Observed to be Narrower in Photoemission Experiments?"

   W. Ku, A. G. Eguiluz, and E. W. Plummer, Phys. Rev. Lett. **85**, 2410 (2000).

2. Plasmon Lifetime in K: A Case Study of Correlated Electrons in Solids Amenable to *Ab Initio* Theory

   W. Ku and A. G. Eguiluz, Phys. Rev. Lett. **82**, 2350 (1999).

3. Dynamical Response of Correlated Electrons in Solids Probed by Inelastic Scattering Experiments: An *Ab Initio* Theoretical Perspective

   A. G. Eguiluz, W. Ku, and J. M. Sullivan, J. of Phys. and Chem. of Solids **61**, 383 (2000).

4. *Ab Initio* Studies of Electronic Excitations in Real Solids

   A. G. Eguiluz and W. Ku, *Electron Correlations and Materials Properties*, edited by A. Gonis, N. Kioussis, and M. Ciftan (Kluwer Academic, New York, 1999), p. 329.

Vita

Wei Ku was born in Taipei, Taiwan, ROC on March 17, 1969. He graduated from Pingtong High School in June 1987. He entered the Department of Physics of the Tamkang University in August 1987 and received the Department Head's Award in the following year. He finished his Bachelor of Science degree in June, 1991.

After serving in the military as a logistic officer for two years, Wei Ku completed his military duty with honors from the commander and the commanding general in June 1993. He then joined the Department of Physics of the Tamkang University as a teaching faculty of general laboratory for one year.

Wei Ku came to US and started his graduate study with teaching assistantship in the Physics Department of Georgia State University in August 1994. He transferred to Department of Physics and Astronomy in University of Tennessee at Knoxville in the next spring. He received the Graduate Teaching Assistantship in August 1995 before joined Dr. Eguiluz's group with Graduate Research Assistantship in January 1996. He received the honor of "Joe Fowler & Jerry Marion Outstanding Graduate Student Award" in 1998. To date, Wei Ku is a coauthor of four published articles (listed in Appendix C) and three articles to be published.