The GW space-time method for the self-energy of large systems

Martin M. Rieger[§‡], L. Steinbeck[† *], I. D. White[§], H. N. Rojas[§‖], and R. W. Godby[†¶]
[§] *Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge, CB3 0HE, UK*
[†] *Department of Physics, University of York, Heslington, York YO1 5DD, UK*
(February 1, 2008)

We present a detailed account of the GW space-time method. The method increases the size of systems whose electronic structure can be studied with a computational implementation of Hedin's GW approximation. At the heart of the method is a representation of the Green's function G and the screened Coulomb interaction W in the real-space and imaginary-time domain, which allows a more efficient computation of the self-energy approximation $\Sigma = iGW$. For intermediate steps we freely change between representations in real and reciprocal space on the one hand, and imaginary time and imaginary energy on the other, using fast Fourier transforms. The power of the method is demonstrated using the example of Si with artificially increased unit cell sizes.
keywords: electronic structure, quasiparticle energies, self-energy calculations, GW approximation

71.15.Th,71.20.-b,79.60.Jv

I. INTRODUCTION

Computational electronic structure theory for real materials depends on the use of simplifying approximations for the many-electron problem. Two successful approaches have been the use of density functional theory and many-body perturbation theory. The density functional approach is overwhelmingly dominated by the local-density approximation (LDA) of Kohn and Sham [1], and extensions thereof, such as gradient corrections or self-interaction corrections. The many-electron problem is mapped onto an effective non-interacting electron problem and solved for the ground-state density and energy. The limitation of this approach lies in the fact that in principle it gives no access to the excitation spectrum of the system under study, even if approximations for the exchange and correlation potential are further refined. This limitation is felt particularly severely in semiconductor physics, where many of the phenomena of interest are centered on properties of the excited states. For this class of materials Hybertsen and Louie [2,3] showed that the GW approximation, first proposed by Hedin [4] in 1965, allows computation of band gaps in remarkably good agreement with experiment for a series of semiconducting and insulating materials.

The GW approximation gives a comparatively simple expression for the self-energy operator, which allows the one-particle Green's function of an interacting many-electron system to be described in terms of the Green's function of a hypothetical non-interacting system with an effective potential. The Green's function contains information not only about the ground state density and energy but also about the quasiparticle spectrum. The GW approximation has still proved computationally very expensive and has mainly been used to determine the quasiparticle spectrum of bulk semiconductors and insulators [3,5,6], although progress has also been made in the treatment of systems such as surfaces [7], clusters [8] and simple polymers [9]. The issue of self-consistency has only begun to be addressed fairly recently [10–13]. The situation is similar for total energy calculations, where thorough investigations so far have been made only for the homogeneous electron gas [14]. First GW calculations of the charge density of Si and Ge have been performed recently [15].

The method we describe in this paper substantially reduces the computational effort needed to study larger systems. The core idea was outlined in a Letter by Rojas, Godby, and Needs [16]. Here we give a full account of the method and detailed aspects of our implementation. The improvements in efficiency over traditional implementations of the GW approximation in a reciprocal-space formalism result from choosing the representation most suitable to the computational step being undertaken, either reciprocal space or real space on the one hand and imaginary time or imaginary energy on the other, and switching between representations easily with the help of fast Fourier transforms (FFTs). The choice of representing the time/energy dependence on the imaginary instead of on the real axis allows us to deal with smooth, decaying quantities which give faster convergence. To obtain the self-energy eventually on the real energy axis, we fit a model function to the computed self-energy on the imaginary axis, and continue it analytically to the real axis.

The paper is organised as follows: The second Section reviews briefly the equations needed for implementation of the GW approximation. Section III describes the essential concepts of the GW space-time method. In the fourth Section we outline detailed aspects of the implementation such as mesh discretisation and exploitation of symmetry. In Section V we discuss the analytic continuation procedure for the self-energy, and subsequent determination of the quasiparticle energies. In the sixth Section we present results for bulk silicon and silicon with artificially increased unit cell sizes, and discuss scaling behaviour for these examples. A summary concludes the paper. Atomic units are used throughout.

II. THE *GW* APPROXIMATION

The central idea of Hedin's *GW* approximation is to approximate the self-energy operator Σ by

$$\Sigma(\mathbf{r},\mathbf{r}';\omega) = \frac{i}{2\pi}\int_{-\infty}^{\infty} d\omega'\, W(\mathbf{r},\mathbf{r}';\omega')\, G(\mathbf{r},\mathbf{r}';\omega+\omega')e^{i\omega'\delta}, \tag{2.1}$$

where δ is an infinitesimally small positive time, W is the screened Coulomb interaction,

$$W(\mathbf{r},\mathbf{r}';\omega) = \int d^3\mathbf{r}'' v(\mathbf{r}-\mathbf{r}'')\, \epsilon^{-1}(\mathbf{r}'',\mathbf{r}';\omega) \tag{2.2}$$

where $v(\mathbf{r}-\mathbf{r}'')$ is the Coulomb interaction $1/|\mathbf{r}-\mathbf{r}''|$ and G is the one-particle Green's function. G itself depends on Σ through the Dyson equation and should arguably be determined self-consistently. In practice, however, in almost all calculations for real systems G has been approximated by the non-interacting Green's function at the LDA level, i. e.,

$$G^{LDA}(\mathbf{r},\mathbf{r}';\omega) = \sum_{n\mathbf{k}} \frac{\Psi_{n\mathbf{k}}(\mathbf{r})\, \Psi^*_{n\mathbf{k}}(\mathbf{r}')}{\omega-\epsilon_{n\mathbf{k}}-i\eta} \tag{2.3}$$

where η is a positive (negative) infinitesimal for occupied (unoccupied) one-particle states. The wavefunctions $\Psi_{n\mathbf{k}}$ in this equation are eigenfunctions, with eigenvalues $\epsilon_{n\mathbf{k}}$, determined from a self-consistent LDA calculation for the system under consideration.

For the inverse dielectric function in Eq. (2.2) one has to rely on suitable approximations. We use the random phase approximation (RPA),

$$\epsilon^{RPA}(\mathbf{r},\mathbf{r}';\omega) = \delta(\mathbf{r}-\mathbf{r}') - \int d\mathbf{r}'' v(\mathbf{r}-\mathbf{r}'')\, P^0(\mathbf{r}'',\mathbf{r}';\omega) \tag{2.4}$$

with the irreducible polarization propagator P^0 at RPA level given by

$$P^0(\mathbf{r},\mathbf{r}';\omega) = -\frac{i}{2\pi}\int_{-\infty}^{\infty} d\omega'\, G^{LDA}(\mathbf{r},\mathbf{r}';\omega')\, G^{LDA}(\mathbf{r},\mathbf{r}';\omega'-\omega). \tag{2.5}$$

Part of the efficiency of our method derives from the fact that the convolutions Eqs. (2.1) and (2.5) in the frequency domain become simple multiplications in the time domain (real or imaginary): see next section. For real times the Green's function Eq. (2.3) becomes

$$G^{LDA}(\mathbf{r},\mathbf{r}';\tau) = \begin{cases} i\sum_{n\mathbf{k}}^{occ} \Psi_{n\mathbf{k}}(\mathbf{r})\Psi^*_{n\mathbf{k}}(\mathbf{r}')\exp(-i\epsilon_{n\mathbf{k}}\tau), & \tau<0, \\ -i\sum_{n\mathbf{k}}^{unocc} \Psi_{n\mathbf{k}}(\mathbf{r})\Psi^*_{n\mathbf{k}}(\mathbf{r}')\exp(-i\epsilon_{n\mathbf{k}}\tau), & \tau>0. \end{cases} \tag{2.6}$$

For imaginary times our expression for G^{LDA} (see Eq. (3.3) later) corresponds to analytically continuing the $\tau<0$ form (the retarded Green's function) to the positive imaginary time axis, and the $\tau>0$ form (the advanced Green's function) to the negative imaginary axis.

Once the self-energy operator is known we can employ first-order perturbation theory in $\langle \Sigma - V_{\rm xc}^{\rm LDA}\rangle$ to compute quasiparticle corrections to the LDA eigenenergies (see Section V B).

III. THE *GW* SPACE-TIME METHOD

A. Mathematical formulation

The traditional way to set up and solve the equations for the *GW* approximation has been to express and compute all quantities in the reciprocal-space and energy domain, using a plasmon-pole model for the energy dependence of W [3]. To compute P^0 (Eq. (2.5)) and Σ (Eq. (2.1)) this involves sums scaling with the fourth power of the number of plane waves $N_{\mathbf{G}}$ used to represent the wavefunctions and quadratically with the number of energy points N_ω used to represent the energy dependence, whereas the scaling in the real space and time domain is quadratic in $N_{\mathbf{G}}$ and linear in N_ω, since convolutions in the reciprocal-space and energy domain become simple multiplications in the real-space and time domain. The expressions for ϵ and W, on the other hand, are more efficiently calculated in a reciprocal-space and energy representation. As fast Fourier transforms permit us to switch between representations with a numerical cost that scales like $N\log N$, where N is the number of points involved in the FFT, we can efficiently exploit the advantages offered by the one or other representation.

The time- or energy-dependence of the quantities involved shows a structure that is not easily represented on an equidistant grid suitable for an FFT. However, we can rigorously analytically continue the quantities to the imaginary time or energy axes where the structure is much smoother and therefore amenable to a representation on an equidistant grid. This point is illustrated by Fig. 1 of Ref. [16] which shows the energy dependence on the real and imaginary axis of the imaginary part of the self-energy $\Sigma(k,\omega)$ of jellium with a density parameter of $r_s = 2.0$. The rather ragged shape on the real energy axis contrasts with the smooth shape on the imaginary energy axis. We emphasise that, while the same amount of physical information is contained for Σ or similar functions on the real or imaginary time or energy axis, to obtain a well converged final answer for quantities that go through several Fourier transformations, the important information is more easily represented in imaginary time/energy, in much the same way that the choice of basis set can reduce the effort needed to satisfactorily represent wavefunctions. This is also shown in plots of the self-energy in Section VI, where it can be seen that

the smooth form of the self-energy on the imaginary axis still allows stable and accurate reproduction of the more complicated behaviour on the real axis.

Using the imaginary time/energy representation allows us to explicitly take the time- or energy-dependence into account without having to rely on a plasmon-pole or other model for most of the calculation. Only after the full imaginary-energy dependence of the expectation values of the self-energy operator has been established do we use a fitted model function (whose sophistication may be increased as necessary with negligible expense), which we then analytically continue to the real energy axis to compute the quasiparticle energies.

The Fourier transforms between the complex axes work like their counterparts on the real axes, except that additional factors of $\pm i$ have to be included:

$$F(i\tau) = \frac{i}{2\pi} \int_{-\infty}^{\infty} d\omega F(i\omega) \exp(i\omega\tau), \tag{3.1}$$

$$F(i\omega) = -i \int_{-\infty}^{\infty} d\tau F(i\tau) \exp(-i\omega\tau). \tag{3.2}$$

Mathematically they can be understood as Laplace transforms followed by analytic continuation to the imaginary axis.

The computational steps which are successively undertaken in the GW space-time method are in detail:

1. Construction of the Green's function in real space and imaginary time: [17]

$$G^{LDA}(\mathbf{r}, \mathbf{r}'; i\tau) \tag{3.3}$$

$$= \begin{cases} i \sum_{n\mathbf{k}}^{occ} \Psi_{n\mathbf{k}}(\mathbf{r}) \Psi^*_{n\mathbf{k}}(\mathbf{r}') \exp(\epsilon_{n\mathbf{k}}\tau), & \tau > 0, \\ -i \sum_{n\mathbf{k}}^{unocc} \Psi_{n\mathbf{k}}(\mathbf{r}) \Psi^*_{n\mathbf{k}}(\mathbf{r}') \exp(\epsilon_{n\mathbf{k}}\tau), & \tau < 0, \end{cases}$$

2. formation of the RPA irreducible polarizability in real space and imaginary time:

$$P^0(\mathbf{r}, \mathbf{r}'; i\tau) = -iG^{LDA}(\mathbf{r}, \mathbf{r}'; i\tau) G^{LDA}(\mathbf{r}', \mathbf{r}; -i\tau), \tag{3.4}$$

3. Fourier transformation of P^0 to reciprocal space and imaginary energy and construction of the symmetrised dielectric matrix [18] in reciprocal space,

$$\tilde{\epsilon}(\mathbf{k}, \mathbf{G}, \mathbf{G}'; i\omega) = \delta_{\mathbf{G}\mathbf{G}'} - \frac{4\pi}{|\mathbf{k}+\mathbf{G}|\,|\mathbf{k}+\mathbf{G}'|} P^0(\mathbf{k}, \mathbf{G}, \mathbf{G}'; i\omega), \tag{3.5}$$

4. inversion of the symmetrised dielectric matrix for each $\mathbf{k}$ point and each imaginary energy in reciprocal space,

5. calculation of the screened Coulomb interaction in reciprocal space:

$$W(\mathbf{k}, \mathbf{G}, \mathbf{G}'; i\omega) = \frac{4\pi}{|\mathbf{k}+\mathbf{G}|\,|\mathbf{k}+\mathbf{G}'|} \times \tilde{\epsilon}^{-1}(\mathbf{k}, \mathbf{G}, \mathbf{G}'; i\omega), \tag{3.6}$$

6. Fourier transformation of W to real space and imaginary time,

7. computation of the self-energy operator:

$$\Sigma(\mathbf{r}, \mathbf{r}'; i\tau) = iG(\mathbf{r}, \mathbf{r}'; i\tau) W(\mathbf{r}, \mathbf{r}'; i\tau), \tag{3.7}$$

8. evaluation of the expectation values:

$$\langle \Psi_{n\mathbf{k}} | \Sigma(i\tau) | \Psi_{n\mathbf{k}} \rangle, \tag{3.8}$$

9. Fourier transformation of the expectation values to imaginary energy,

10. fitting of a model function to the expectation values of the self-energy, allowing analytic continuation onto the real energy axis,

11. evaluation of the quasiparticle corrections to the LDA eigenvalues by first-order perturbation theory in $\langle \Sigma - V_{\mathrm{xc}}^{\mathrm{LDA}} \rangle$.

B. Discretisation of the equations

A practical implementation on a computer requires the integrals described in the last section to be discretised and suitably truncated. Exploitation of symmetry can help to keep the computational cost for real materials down. These issues will be addressed in this section.

The quantities we are dealing with in real space, such as G, P, W, and Σ are non-local operators that decay as their two spatial arguments move apart. With the exception of W, which will receive some extra attention, the non-locality is short-ranged, i.e, decaying faster than $|\mathbf{r} - \mathbf{r}'|^{-2}$. This suggests that the distance the two variables are allowed to move apart can be restricted suitably and the functions can be assumed zero beyond this range. We call this range the interaction cell (IC). Furthermore, by the symmetry of the crystal, one of the arguments of any of these operators, which shall be symbolically denoted here by $F(\mathbf{r}, \mathbf{r}')$, can be restricted to the irreducible wedge of one unit cell (IUC). The coarseness of the grid and the size of the interaction cell determine the precision. A shape of interaction cell which is compatible

with the fast Fourier transforms and preserves symmetry at the same time is the Wigner-Seitz cell of a lattice whose defining vectors are multiples of the primitive vectors of the crystal lattice. We call this lattice the IC lattice (ICL) (see Figure 1). We choose equal spacing for the IUC and IC grids. They could be offset from each other which would avoid the singularity of the Coulomb potential in real space but requires the handling of additional phase factors. We therefore normally choose the two grids to have no offset between them and deal with a single real-space grid (RSG), which is defined by vectors which are integer fractions of the primitive lattice vectors. The treatment of the singularity of the Coulomb potential is discussed in sections IV B and IV C below.

To summarise: the crystal is defined by the three primitive vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, the ICL by the three vectors

$$\mathbf{l}_i = N_i^{\mathbf{R}} \mathbf{a}_i, \quad i = 1, 2, 3, \tag{3.9}$$

and the RSG by three vectors

$$\mathbf{s}_i = \mathbf{a}_i / N_i^{\mathbf{r}}, \quad i = 1, 2, 3. \tag{3.10}$$

In the example shown in Fig. 1 the $N_i^{\mathbf{R}}$ are 2 and the $N_i^{\mathbf{r}}$ are 3.

The grid vectors $\mathbf{x}$ in the IC are integer linear combinations of the $\mathbf{s}_i$ and fulfil the Wigner-Seitz condition

$$\mathbf{x} \cdot \mathbf{L} \leq \frac{1}{2} L^2, \tag{3.11}$$

for any vector $\mathbf{L}$ that is an integer linear combination of the vectors $\mathbf{l}$. Only one of possibly several vectors for which the equality in Eq. (3.11) holds and which differ only by a lattice vector of the ICL must be contained in the IC, or, expressed differently, only half the surface of the Wigner-Seitz cell defined by Eq. (3.11) is part of the IC. The integers $N_i^{\mathbf{R}}$ in Eq. (3.9) determine the shape and size of the interaction cell in real space and the k-point grid in reciprocal space, as explained below. If there is no reason to believe that the non-locality of the operators has a very different range in one direction than in another, they should be chosen in such proportion to each other that the shape of the IC is as close to a sphere as possible.

The first argument $\mathbf{r}$ of the functions of type F (see above) is again an integer linear combination of the vectors $\mathbf{s}$ and can be restricted to the irreducible wedge of one unit cell of the crystal lattice. This unit cell can in principle have any shape, but it helps to avoid having to deal with phase factors when applying symmetry operations if this unit cell is also chosen to have Wigner-Seitz shape, i. e.,

$$\mathbf{r} \cdot \mathbf{R} \leq \frac{1}{2} R^2, \tag{3.12}$$

where $\mathbf{R}$ is any crystal lattice vector and of any two $\mathbf{r}$ which are connected by a symmetry operation in the crystal's space group only one is kept.

The choice of lattices in real space uniquely determines conjugate lattices in reciprocal space and vice versa, if one follows the rule that the discrete approximation to the Fourier transform must leave the functions unchanged if applied forwards and backwards in succession. In reciprocal space, functions of type F (see above) are functions of the two variables $\mathbf{G}$ and $\mathbf{K}$, where $\mathbf{G}$ is a reciprocal lattice vector and $\mathbf{K}$ can be written as $\mathbf{K} = \mathbf{k} + \mathbf{G}'$, where $\mathbf{G}'$ is again a reciprocal lattice vector and $\mathbf{k}$ a reciprocal vector in the first Brillouin Zone (BZ). $\mathbf{K}$ is restricted to the Wigner-Seitz cell of the reciprocal lattice of the RSG, which we shall call simply the *reciprocal-space cell* (RC) and $\mathbf{G}$ is restricted to its irreducible wedge (IRC). The spacing of the $\mathbf{k}$ is determined by the primitive vectors of the reciprocal ICL and we shall call the grid that is defined that way the $\mathbf{K}$ grid (KG) [19].

The transformation from real-space to reciprocal-space representation of F is given by

$$F(\mathbf{G}, \mathbf{K}) = \frac{1}{N_{IC}} \sum_{\mathbf{r}}^{IUC} \sum_{\rho}^{star(\mathbf{r})} \sum_{\mathbf{r}'}^{IC(\mathbf{r})} F(\mathbf{r}, \mathbf{r}') e^{-i\rho\mathbf{r}\cdot\mathbf{G}} e^{i\rho(\mathbf{r}'-\mathbf{r})\cdot\mathbf{K}}. \tag{3.13}$$

ρ denotes a point group element (including possibly a non-primitive translation in non-symmorphic symmetry groups) and the sum over ρ runs over all symmetry operations that generate the full star of $\mathbf{r}$, i. e., the set of all RSG points in the Wigner-Seitz cell that are obtained by applying symmetry operations of the crystal to this particular RSG point $\mathbf{r}$ in the IUC. N_{IC} is the number of RSG points contained in the IC. The reverse transformation is given by

$$F(\mathbf{r}, \mathbf{r}') = \frac{1}{N_{UC}} \sum_{\mathbf{G}}^{IRC} \sum_{\rho}^{star(\mathbf{G})} \sum_{\mathbf{K}}^{RC} F(\mathbf{G}, \mathbf{K}) e^{i\rho\mathbf{r}\cdot\mathbf{G}} e^{-i\rho(\mathbf{r}'-\mathbf{r})\cdot\mathbf{K}}. \tag{3.14}$$

The sum over ρ runs here over all symmetry operations that generate the full star of $\mathbf{G}$ and N_{UC} is the number of RSG points contained in one unit cell.

Because of the symmetry of the problem and the fact that $\mathbf{K} = \mathbf{k} + \mathbf{G}'$, F can in reciprocal space be alternatively written as a square matrix $F_{\mathbf{k}}(\mathbf{G}, \mathbf{G}')$ with $\mathbf{k}$ restricted to the irreducible wedge of the Brillouin Zone and $\mathbf{G}$ and $\mathbf{G}'$ given everywhere in the RC. $\mathbf{k}$ is here written as an index to emphasise the matrix nature of F in this representation.

It should be briefly mentioned that the 'natural' cell form for a three-dimensional FFT is a parallelepiped and not a Wigner-Seitz cell such as we are dealing with. However, there is a unique mapping between the two shapes using translations by vectors of the lattice defining the Wigner-Seitz cell in question, because F is invariant under such translations by construction.

Apart from the real-space and reciprocal-space representations we will occasionally use a mixed-space representation. This is defined by

$$F_{\mathbf{k}}(\mathbf{r},\mathbf{r}') = \sum_{\mathbf{R}}^{IC} F_{\mathbf{R}}(\mathbf{r},\mathbf{r}') \exp\left(-i\mathbf{k}\cdot\mathbf{R}\right), \tag{3.15}$$

with the reverse transformation

$$F_{\mathbf{R}}(\mathbf{r},\mathbf{r}') = \frac{1}{N_{\mathbf{R}}} \sum_{\mathbf{k}} F_{\mathbf{k}}(\mathbf{r},\mathbf{r}') \exp\left(i\mathbf{k}\cdot\mathbf{R}\right). \tag{3.16}$$

When we use the mixed-space representation, $\mathbf{r}$ and $\mathbf{r}'$ are always understood to be confined to the UC (or its irreducible wedge, see below). The notation $F_{\mathbf{R}}(\mathbf{r},\mathbf{r}')$ is another way of writing $F(\mathbf{r}+\mathbf{R},\mathbf{r}')$ and has been chosen to emphasise that any point on the RSG in the IC can be written as the sum of a crystal lattice vector $\mathbf{R}$ and a vector in the UC. The sum in Eq. (3.15) runs over all crystal lattice vectors in the IC. The $\mathbf{k}$ are vectors in the first BZ of the crystal and are the conjugate vectors of the $\mathbf{R}$. Symmetry can be exploited to reduce the number of points needed to represent F in mixed space. For example, $\mathbf{r}$ can be restricted to the irreducible wedge of the UC while $\mathbf{r}'$ is given for every point of the RSG in the UC and $\mathbf{k}$ on every point of the KG within the first BZ, or $\mathbf{k}$ can be restricted to the irreducible wedge of the first BZ, with $\mathbf{r}$ then given on every RSG point in the UC.

A mixed-space (MS) formalism was used by Blase et al. [20] for calculating the polarizability and dielectric function of periodic systems. A mixed-space representation was also used and described by Godby, Sham and Schlüter [21] for *GW* calculations, albeit without referring to it by that name. Blase et al. [20] discuss the computational efficiency of the MS scheme in comparison with a direct real-space (RS) approach. In Fig. 2 of Ref. [20] they show how the number of $(\mathbf{r},\mathbf{r}')$ pairs which have to be computed in order to set up the polarizability $P^0_{\mathbf{k}}(\mathbf{r},\mathbf{r}')$ for one $\mathbf{k}$ in the MS method (using a $4\times4\times4$ Monkhorst-Pack $\mathbf{k}$ grid) depends on the range of interaction (non-locality range) R_{max}. They compare this to the number of $(\mathbf{r},\mathbf{r}')$ pairs obtained by confining $|\mathbf{r}-\mathbf{r}'|$ to a sphere of radius R_{max} which is assumed to be the corresponding number of pairs to be computed in a RS method. This is somewhat misleading since – as is discussed above – the size of the IC in real space is determined by the size of the $\mathbf{k}$ grid of the conjugate mixed- or reciprocal-space quantity. Hence it does not make sense to increase the interaction sphere beyond the boundaries of this IC which correspond to a non-locality range of R_{max}=14.5 a.u. for a $4\times4\times4$ $\mathbf{k}$ grid. Furthermore, the real-space quantity $P^0(\mathbf{r},\mathbf{r}')$ contains information on all $\mathbf{k}$ points. So for a proper comparison between the two methods the number of pairs given in Ref. [20] for the MS approach has to be multiplied by the number of special $\mathbf{k}$ points which is 10 for a $4\times4\times4$ Monkhorst-Pack grid. The statement made in Ref. [20] that the calculation of $P^0(\mathbf{r},\mathbf{r}',\omega)$ requires a double BZ summation in the real-space scheme - as opposed to a single BZ summation in the MS scheme - does not hold any more if the Green's function is set up in mixed space and then Fourier transformed to real space, as described in section IV A below.

So far we have established the nature of the grids in real and reciprocal space that are suitable for use with the discrete Fourier transform. These grids fill a Wigner-Seitz-cell shaped volume which in turn can be uniquely mapped onto a parallelepiped, the shape that is required by the discrete Fourier transforms. However, the starting point of the *GW* calculation is the output of a standard LDA calculation using plane waves and pseudopotentials which determines the Fourier coefficients of the wavefunctions $\Psi_{n\mathbf{k}}(\mathbf{G})$ for all reciprocal lattice vectors that lie within a sphere defined by $(\mathbf{k}+\mathbf{G})^2$ less than some cutoff. The RC must therefore be big enough to comprise all of these 'shifted spheres'. The volume of the smallest possible such cell will be typically 2 to 4 times larger than the volume of the individual spheres. Furthermore, to strictly prevent aliasing in the steps where we replace a convolution in one representation by a multiplication in the conjugate representation, we would have to choose the grid for the FFTs twice as large as this minimal cell in every dimension, increasing the volume by a factor of eight. This means that the FFT grid would have to comprise between 16 and 32 times more points than the initial LDA calculation had reciprocal lattice vectors. We have found that in practice aliasing effects are very small once we choose enough plane waves for good overall convergence, and in the limit of an infinite number of plane waves any aliasing effects vanish strictly along with any other error caused by truncation of the set of reciprocal lattice vectors. Therefore it would not be justified to accept the huge overhead imposed by the doubling of the grid size in all dimensions, and we usually restrict our FFT grid to the minimum Wigner-Seitz cell as defined above, leaving us with an FFT grid with between 2 and 4 times as many points as the LDA calculation used reciprocal lattice vectors.

The computational effort can be further reduced by physical considerations. The FFT grid fills a Wigner-Seitz cell shaped volume in either space. However, if we assume in real space that the range of non-locality is uniform in all directions we can neglect all the points in the interaction cell outside the largest possible sphere inscribed into it. Similarly we can usually assume that the Fourier coefficients in reciprocal space fall to zero after an equal length in all directions and thus we can cut back in reciprocal space to the largest possible sphere inscribed into the RC lattice. It has to be kept in mind that for symmetry reasons it is the vectors $\mathbf{k}+\mathbf{G}$ which must fit into the sphere and not merely the vectors $\mathbf{G}$.

IV. NUMERICAL ASPECTS AND SCALING WITH SYSTEM SIZE

As the *GW* space-time method is primarily designed to enable larger systems to be studied within the framework of the *GW* approximation, in this section we study the scaling of the computational cost with system size. To gauge the cost of a calculation, we first look at the real-space, imaginary-time representation. What matters here is the number of points in the irreducible wedge of the unit cell, the number of points in the interaction cell and the number of points on the imaginary time axis. We will discuss the time dependence of the operators in the next section and concentrate here on the spatial dimensions. Since in our setup the grid spacing in the unit cell and in the interaction cell are chosen equal, the three factors determining the cost of the calculation are the size of the IUC, which is a system property, and the size of the interaction cell and the grid spacing, which are convergence parameters.

A. Green's function and polarizability

The setting up of the Green's function is the starting point for the *GW* calculation. We take the Fourier coefficients of the lattice-periodic part of the Bloch functions, $u_{n\mathbf{k}}(\mathbf{G})$, and the eigenvalues $\epsilon_{n\mathbf{k}}$ from a previous standard LDA calculation. The eigenvalues are expressed on an energy scale with zero at the Fermi energy, which for semiconductors and insulators is taken to be in the middle of the band gap. The wavefunctions $u_{n\mathbf{k}}$ are transformed via an FFT to real space and the Green's function initially computed in mixed space:

$$G^{LDA}_{\mathbf{k}}(\mathbf{r},\mathbf{r}';i\tau) \tag{4.1}$$

$$= \begin{cases} i\sum_{n}^{occ} \Psi_{n\mathbf{k}}(\mathbf{r})\Psi^*_{n\mathbf{k}}(\mathbf{r}')\exp(\epsilon_{n\mathbf{k}}\tau), & \tau > 0, \\ -i\sum_{n}^{unocc} \Psi_{n\mathbf{k}}(\mathbf{r})\Psi^*_{n\mathbf{k}}(\mathbf{r}')\exp(\epsilon_{n\mathbf{k}}\tau), & \tau < 0. \end{cases}$$

The number of unoccupied bands included in the Green's function for $\tau < 0$ is a convergence parameter. Because of the rapid decay of the exponentials for higher energies it suffices to set a cutoff energy for bands included in the sum that is considerably smaller than the largest LDA eigenvalue, but this cutoff must be several times larger than the highest quasi-particle energy to be computed. We denote the total number of bands, occupied and unoccupied, included in the sum as N_{bands}.

The Green's function can then be transformed from mixed space to real space using an FFT:

$$G^{LDA}_{\mathbf{R}}(\mathbf{r},\mathbf{r}';i\tau) = \frac{1}{N_{\mathbf{R}}}\sum_{\mathbf{k}} G^{LDA}_{\mathbf{k}}(\mathbf{r},\mathbf{r}';i\tau)\exp(i\mathbf{k}\cdot\mathbf{R}), \tag{4.2}$$

where the sum goes over the $\mathbf{k}$ grid in the Brillouin zone corresponding to the interaction cell (see III(B) earlier).

To set up the function in mixed space (Eq. (4.1)) we need $N_{IUC}\times N_{UC}\times N_{bands}$ multiplications in the spatial dimensions. The transformation to real space brings in another factor of $N_{\mathbf{R}}$, the number of unit cells contained in the interaction cell, (disregarding the additional factor of $\log N_{\mathbf{R}}$ from the FFT), so that the overall scaling is like $N_{IUC}\times N_{IC}\times N_{bands}$. Since the number of bands is determined by a cutoff energy that is constant for a given material, N_{bands} will grow linearly with system size. The computational cost at this stage therefore scales quadratically with system size assuming a fixed size of interaction cell (i. e., fixed range of non-locality). Additional care is necessary once the unit cell outgrows the range of non-locality. In this case there is only a single $\mathbf{k}$ point necessary and the transformation between mixed and real space is redundant. To prevent a crossover to cubic scaling the interaction cell must be kept constant at its converged size and will then not be a multiple of the unit cell size, but smaller than it.

The next stage, where the irreducible polarizability is formed in real space according to Eq. (3.4), scales linearly.

B. Dielectric matrix

The formula for the dielectric function in real space and on the imaginary energy axis reads

$$\epsilon(\mathbf{r},\mathbf{r}';i\omega) = \delta(\mathbf{r}-\mathbf{r}') - \sum_{\mathbf{r}''}^{IC(\mathbf{r}')} v(\mathbf{r}-\mathbf{r}'')P^0(\mathbf{r}'',\mathbf{r}';i\omega). \tag{4.3}$$

Here $\mathbf{r}'$ is restricted to the IUC, $\mathbf{r}$ to within an IC around $\mathbf{r}'$, and $\mathbf{r}''$ in the sum on the right hand side runs over an interaction cell around $\mathbf{r}'$. The right hand side contains a convolution that can be more efficiently handled in reciprocal space where it becomes a simple multiplication. As only one of the two spatial arguments of the dielectric function is involved, we can write

$$\epsilon(\mathbf{K},\mathbf{r}') = v(\mathbf{K})P^0(\mathbf{K},\mathbf{r}'), \tag{4.4}$$

where $\mathbf{K}$ is the conjugate variable of $\mathbf{x}=\mathbf{r}'-\mathbf{r}$ in reciprocal space. (The imaginary-energy argument is suppressed here and until the end of this subsection.) Again we can see that for a given interaction cell size the computational effort to set up the dielectric function scales linearly with the number of points in the IUC.

The long-range behaviour of the dielectric function requires some special consideration. Eq. (4.3) contains the

convolution of the long-ranged Coulomb potential with the short-ranged non-locality of P^0. In fact, for any $\mathbf{r}'$

$$\int d\mathbf{r} P^0(\mathbf{r}, \mathbf{r}') = 0, \tag{4.5}$$

whereas

$$\int d\mathbf{r} v(\mathbf{r}) = \infty. \tag{4.6}$$

The analytic convolution of the two functions over all space yields a function whose integral over all space is finite, meaning that the Fourier coefficient for $\mathbf{K} = 0$ has a finite, non-zero value. To preserve this behaviour, which is important if one is actually interested in computing the value of the dielectric constant but also for fast convergence with respect to interaction cell size of the quasiparticle energies, one has to use a modified Coulomb interaction in Eq. (4.3). By construction, the property expressed by Eq. (4.5) is within our discrete and strictly finite range approximation compressed into the interaction cell:

$$\sum_{\mathbf{r}}^{IC(\mathbf{r}')} P^0(\mathbf{r}, \mathbf{r}') = 0. \tag{4.7}$$

To obtain the right dielectric constant, we have to compress the property expressed by Eq. (4.6) into the finite interaction cell as well. A natural way to do this is to use the reciprocal-space definition of v:

$$v(\mathbf{K}) = \frac{4\pi}{\mathbf{K}^2}. \tag{4.8}$$

Transforming to real space, through a discrete Fourier transform restricted to one interaction cell, yields an expression that contains correction terms to the simple $1/r$ form that will vanish in the limit of infinitesimal grid spacing and infinite IC size. Since, in practice, we evaluate the convolution Eq. (4.3) as a multiplication in reciprocal space, there is no need to explicitly transform Eq. (4.8) to real space.

In practice, to deal with the divergence of the Coulomb potential at zero wavevectors, we follow the procedure described in the literature [22,18], i. e., we employ $\mathbf{k} \cdot \mathbf{p}$ perturbation theory for calculating the head ($\mathbf{k} = \mathbf{G} = \mathbf{G}' = 0$) and wings ($\mathbf{k} = \mathbf{G} = 0$; $\mathbf{G}' \neq 0$) of the symmetrised dielectric matrix Eq. (3.5). The matrix elements of type $\langle \Psi_{n\mathbf{q}} | e^{-i\mathbf{k}\cdot\mathbf{r}} | \Psi_{n,\mathbf{q}-\mathbf{k}} \rangle$ appearing in the expression for the RPA polarizability in reciprocal-space representation (cf. Eq. (18) of Ref. [22]) are evaluated by expanding the wavevector dependence of the wavefunctions and the exponential for small $\mathbf{k}$. This yields the lowest-order (in $\mathbf{k}$) terms for head and wings of the polarizability. The head of the polarizability goes to zero like k^2 thus cancelling the $1/k^2$ divergence of the Coulomb potential at $\mathbf{k} = \mathbf{G} = \mathbf{G}' = 0$, yielding a finite value for the head of vP^0 whereas the lowest-order term for the wings $P^0_{0\mathbf{G}'}(0)$ is proportional to k, cancelling the $1/k$ divergence of $v_{0\mathbf{G}'}(0)$ and yielding a finite $(vP^0)_{0\mathbf{G}'}(0)$. Head, wings and body of the inverse dielectric matrix are then obtained in terms of head, wings and body of the dielectric matrix by block-wise inversion as described by Pick, Cohen and Martin [23].

The inversion of the dielectric matrix can be performed either in mixed space or in reciprocal space. While this is strictly equivalent if the whole FFT grid is used in either space, it is important to look at ways to reduce the grid size without compromising the exactness of the result, as the matrix inversion is in fact the one operation in the computational sequence which scales worst with system size (unless sparseness is exploited, see below) and will therefore be the bottleneck for large systems. In mixed space we have to invert square matrices of dimension N_{UC} for every $\mathbf{k}$ point in the irreducible wedge of the Brillouin zone (IBZ).

First we look at the case where the IC is in no dimension smaller than the UC. We are then dealing with the inversion of fully occupied square matrices, which scales cubicly with the dimension of the matrix, so that the inversions together scale like $N_{IBZ} \times N_{UC}^3$. This number is essentially the same as $N_{IUC} \times N_{IC} \times N_{UC}$, small differences arising possibly because of the coarseness of the grid. This shows that the scaling is quadratic with system size for constant IC size and constant N_{UC}/N_{IUC}. If the increase of the unit cell size means also a lowering of symmetry, an additional factor corresponding to the symmetry reduction comes in. Transforming to reciprocal space and cutting back to a set of reciprocal lattice vectors $\mathbf{G}$ within a sphere as described earlier reduces the dimension of the matrices by a factor of 2 to 4, which speeds up the inversions by a factor of 8 to 64, because of the cubic scaling. This is the procedure we choose routinely in our calculations.

If the unit cell size exceeds the interaction cell size in one or more dimensions the matrix in mixed space becomes sparse, as all those elements for which $\mathbf{r}'$ is outside the IC become zero. This sparseness can be exploited to the effect that the matrix inversion scales only as N_{UC}^2, so that the overall scaling with system size remains the same. However, in the scheme we are employing at the moment this sparseness is not yet explicitly exploited so that we are in fact dealing with a situation where the IC grows with the UC, leading for the matrix inversion to an N_{UC}^3 scaling once the unit cell outgrows the interaction cell.

C. Screened Coulomb interaction

The screened Coulomb interaction is defined in real space as

$$W(\mathbf{r}, \mathbf{r}'; i\omega) = \int d\mathbf{r}'' \epsilon^{-1}(\mathbf{r}, \mathbf{r}''; i\omega) v(\mathbf{r}'' - \mathbf{r}') \tag{4.9}$$

The convolution on the right hand side is again best dealt with as a multiplication in reciprocal space

$$W_{\mathbf{G}\mathbf{G}'}(\mathbf{k};i\omega) = \epsilon_{\mathbf{G}\mathbf{G}'}(\mathbf{k};i\omega)v_{\mathbf{G}\mathbf{G}'}(\mathbf{k}). \quad (4.10)$$

However, in a semiconductor and in metals for frequencies other than zero, W is truly long ranged, and therefore diverges for zero wavevectors. In order to avoid problems (in the $\mathbf{G}/\mathbf{r}$ and $\mathbf{G}'/\mathbf{r}'$ FFT) associated with the resulting long-range tail and (in the $i\tau/i\omega$ FFT) with the asymptotic frequency dependence we separate W into a long-range part which is immediately set up in real space and a short-range part W_s which is first computed in reciprocal space. From the long-range behavior of the dielectric function we find the isotropic part of the screened interaction in the long-range limit

$$\lim_{|\mathbf{r}-\mathbf{r}'|\to\infty} W(\mathbf{r},\mathbf{r}';i\omega) = f(i\omega)v(\mathbf{r}-\mathbf{r}') \quad (4.11)$$

so that we can define a short-ranged part W_s of W

$$W_s(\mathbf{r},\mathbf{r}';i\omega) = \int d\mathbf{r}'' \left(\epsilon^{-1}(\mathbf{r},\mathbf{r}'';i\omega) - f(i\omega)\delta(\mathbf{r}-\mathbf{r}'')\right) \times v(\mathbf{r}''-\mathbf{r}'). \quad (4.12)$$

W_s has well-defined Fourier coefficients for all reciprocal vectors. It can thus be computed in reciprocal space. The long-range part is set up in real space with $f(i\omega)$ determined from the small-wavevector behavior of the inverse dielectric matrix in reciprocal space, i. e., the head of the inverse dielectric matrix ϵ_0^{-1}:

$$f(i\omega) = \epsilon_0^{-1}(i\omega) - 1, \quad (4.13)$$

Even though the interaction is truly long-ranged we only need to set it up inside the IC because we aim to calculate the self-energy whose range is determined by that of the (short-ranged) Green's function G_0. Hence, at this point, it is sufficient to take the contribution of the long-range part of W into account properly *within* the IC. The Coulomb potential at $\mathbf{r} = \mathbf{r}'$ is approximated by the average over a sphere around $\mathbf{r} = \mathbf{r}'$ with the same volume as one real-space grid cell.

D. Self-energy

The self-energy operator is set up in the real-space and imaginary-time domain:

$$\Sigma(\mathbf{r},\mathbf{r}';i\tau) = iG^{LDA}(\mathbf{r},\mathbf{r}';i\tau)W(\mathbf{r},\mathbf{r}';i\tau - i\delta). \quad (4.14)$$

Within the discrete-grid and finite-range approximation a complete description is given if $\mathbf{r}$ is restricted to the IUC and $\mathbf{r}'$ to an IC around $\mathbf{r}$. The scaling is again linear with N_{IUC} for fixed N_{IC}.

E. Expectation values of the self-energy

To compute the expectation values of Σ between wavefunctions at the special $\mathbf{k}$ points, we transform Σ to mixed space and form the matrix elements

$$\langle\Psi_{nk}|\Sigma(i\omega)|\Psi_{nk}\rangle = \sum_{\mathbf{r},\mathbf{r}'}^{UC} \Psi^*_{n\mathbf{k}}(\mathbf{r})\Sigma_{\mathbf{k}}(\mathbf{r},\mathbf{r}';i\omega)\Psi_{n\mathbf{k}}(\mathbf{r}') \quad (4.15)$$

This operation again scales quadratically with the number of points in the UC for any given set of quantum numbers $n\mathbf{k}$. For a general point $\mathbf{q}$ in the first Brillouin zone that is not on our discrete grid, we generate interpolated values through the formula

$$\Sigma_{\mathbf{q}} = \sum_{\mathbf{R}}^{IC} \Sigma_{\mathbf{R}} \exp(-i\mathbf{q}\cdot\mathbf{R}). \quad (4.16)$$

V. QUASIPARTICLE ENERGIES

A. Analytic continuation

In order to calculate the quasiparticle corrections to the LDA eigenvalues we require the self-energy operator as a function of real energy, and thus a key requirement for our imaginary time method is the ability to obtain the expectation values of Σ on the real energy axis accurately from the imaginary energy behaviour. From complex analysis we know that if two functions are equal over any arc in the complex plane then they are equal everywhere in their common region of analyticity. We know from the structure of G and W that $\Sigma(z)$ (z denoting complex energy) has poles in the second and fourth quadrant of the complex plane. If we know the analytic form of the expectation values of the self-energy on the imaginary energy axis we can analytically continue it from the negative imaginary energy axis to the negative real energy axis, and from the positive imaginary to the positive real axis, without crossing any branch cuts.

To obtain such an analytic form, we fit a multipole model function for each pair of quantum numbers $n\mathbf{k}$ to the values $\langle\Psi_{n\mathbf{k}}|\Sigma(i\omega)|\Psi_{n\mathbf{k}}\rangle = \langle\Sigma(i\omega)\rangle$:

$$\langle\Psi_{n\mathbf{k}}|\Sigma(z)|\Psi_{n\mathbf{k}}\rangle \simeq a^0_{n\mathbf{k}} + \sum_{i=1}^{n} \frac{a^i_{n\mathbf{k}}}{z - b^i_{n\mathbf{k}}}, \quad (5.1)$$

where $a^i_{n\mathbf{k}}$ and $b^i_{n\mathbf{k}}$ are complex fit parameters and z is the complex energy. In Fig. 2 the resulting fitted form of the correlation (energy-dependent) part of the self-energy [24] is compared with the calculated matrix elements for bands at the Γ and X points (the fits are plotted with the same line styles as the respective calculated matrix elements, they are on this scale indistinguishable from the

latter). A simple two-pole model ($n = 2$) performs very successfully for Si, with the fitted function reproducing the actual values with an rms error of less than 0.2%. Including several further poles in the fitted form proves to be stable but unnecessary, although more poles are expected to be required for systems with multiple natural energy scales. For the analytic continuation to be valid, the fitted pole positions $b^i_{n\mathbf{k}}$ should lie in the upper half plane when fitting the negative imaginary axis and vice versa, a condition which in practice is obeyed by the optimal parameters. The parameters $a^0_{n\mathbf{k}}$ should in principle be zero, as $\lim_{\omega\to\infty} \langle\Sigma(\omega)\rangle = 0$. Allowing a small finite value for $a^0_{n\mathbf{k}}$ has proved helpful in the fits, though, and as we are interested mainly in energies close to the Fermi energy and not the long-range limit this is perfectly legitimate. Since $\langle\Sigma(i\omega)\rangle = \langle\Sigma(-i\omega)\rangle^*$ it is sufficient to fit only one half-axis.

Convergence of the quasiparticle energies with respect to the parameters of the time-energy transform grid is discussed in the next section, but it is also important to consider as a separate question the stability of the analytic continuation procedure. In particular if we are to achieve smooth convergence it is important that changes in the calculated self-energy on the real axis correspond to genuine changes in the calculated $\langle\Sigma(i\omega)\rangle$, and are not simply resulting from instabilities in the fitting procedure. In Figures 3 and 4 we show the convergence behaviour of a matrix element $\langle\Sigma(i\omega)\rangle$ at Γ with respect to number and spacing of the energy points respectively, and the corresponding form of the analytically continued element. It can be seen that the fitting approach is indeed stable, with the convergence of $\langle\Sigma(\omega)\rangle$ being dictated directly by changes in the calculated $\langle\Sigma(i\omega)\rangle$. Thus the task of converging the quasiparticle energies is equivalent to converging the self-energy on the imaginary axis, with very little comparative loss of accuracy in the analytic continuation procedure itself.

The parameter ω_{max} (describing the energy range used in the calculation) mainly determines the convergence of the large-energy region of $Im\,\langle\Sigma_c(i\omega)\rangle$. The energy grid spacing $\Delta\omega$ predominantly affects the convergence of $\langle\Sigma_c(i\omega)\rangle$ in the region close to ω= 0. The behavior of $\langle\Sigma_c(\omega)\rangle$ (on the whole real energy axis) is more sensitive to the shape of $\langle\Sigma_c(i\omega)\rangle$ in this region of the imaginary axis than to the large-imaginary-energy tail of the latter (see also next paragraph). That is why the convergence with respect to $\Delta\omega$ shown in Fig. 4 appears to be better on the imaginary axis than on the real axis.

Our investigations of the convergence of the calculated matrix elements of the correlation part of the self-energy and the resulting quasiparticle energies with respect to ω_{max} have shown that good convergence can be achieved by keeping the energy range for the fitting procedure fixed rather than fitting the whole range of energies $(-\omega_{max}, +\omega_{max})$ when increasing the latter. The energy range for fitting the matrix elements $\langle\Sigma_c(i\omega)\rangle$ was fixed to (−5 Har, +5 Har) in the present calculation (the results are not sensitive to the exact value). The reason for restricting the fit range for $\langle\Sigma_c(i\omega)\rangle$ is that most information is contained in the form of this function at imaginary energies reasonably close to $i\omega = 0$. Thus, fitting a large range, beyond a certain point, will mean that this part of the shape will be less accurately described due to the loss of weight in this region. This effect prevents $\langle\Sigma_c(\omega)\rangle$ from converging properly unless the fit range is kept fixed (at any reasonably sensible value), as is illustrated by Fig. 5.

B. Quasiparticle corrections

The quasiparticle energies are formally solutions of the equations

$$\left(-\frac{1}{2}\nabla^2 + V_{ext}(\mathbf{r}) + V_H(\mathbf{r})\right)\Psi^{qp}_{n\mathbf{k}}(\mathbf{r}) + \int d\mathbf{r}'\Sigma(\mathbf{r},\mathbf{r}';\epsilon^{qp}_{n\mathbf{k}})\Psi^{qp}_{n\mathbf{k}}(\mathbf{r}') = \epsilon^{qp}_{n\mathbf{k}}\Psi^{qp}_{n\mathbf{k}}(\mathbf{r}). \quad (5.2)$$

Because Σ is in general a non-Hermitian operator, the eigenenergies $\epsilon^{qp}_{n\mathbf{k}}$ are complex, their real part being interpreted as quasiparticle energy and their imaginary part as lifetime. It has been observed [3,21] that the wavefunctions obeying Eq. (5.2) for typical semiconductors and insulators have an almost complete overlap with the LDA eigenfunctions which solve an equation in which the non-local self-energy operator is replaced by a local exchange-correlation potential. This makes it possible to find the quasiparticle energies by computing corrections to the LDA eigenvalues in first-order perturbation theory in most cases. (For some systems, however, the quasiparticle wavefunctions are expected to be qualitatively different from the LDA wavefunctions, such as at a surface. The space-time method allows the full quasiparticle wavefunctions and energies to be calculated where required [25].) Usually the LDA Hamiltonian with its eigenfunctions and eigenvalues is taken as the starting approximation, with $\Sigma(\omega) - V_{xc}$ as a perturbation. Following an idea used by Hedin [4] for the electron gas, we shift the LDA eigenergies by a constant ϵ_s that aligns the Fermi energies of the quasiparticles before and after applying the GW correction. This is intended to simulate to some extent the effect of self-consistency in G and has been shown in a model system [26] to be instrumental in keeping charge conservation violations negligible.

Calculation of the full energy dependence of the self-energy (via the analytic continuation to real energies) allows us to solve the equation

$$\epsilon^{qp}_{n\mathbf{k}} = \epsilon^{LDA}_{n\mathbf{k}} + \langle\Psi_{n\mathbf{k}}|\,\Sigma(\epsilon^{qp}_{n\mathbf{k}}) - V_{xc} - \epsilon_s\,|\Psi_{n\mathbf{k}}\rangle \quad (5.3)$$

for the quasiparticle energies self-consistently. This approach was employed for the calculation of the quasiparticle energies in the present paper. Alternatively, a Taylor expansion of $\Sigma(\omega)$ at $\omega = \epsilon^{LDA}_{n\mathbf{k}}$ can be used:

$$\epsilon^{qp}_{n\mathbf{k}} = \epsilon^{LDA}_{n\mathbf{k}} + \frac{1}{Z_{n\mathbf{k}}} \langle \Psi_{n\mathbf{k}} | \Sigma(\epsilon^{LDA}_{n\mathbf{k}}) - V_{xc} - \epsilon_s | \Psi_{n\mathbf{k}} \rangle , \tag{5.4}$$

$$Z_{n\mathbf{k}} = 1 - \frac{d}{d\omega} \langle \Psi_{n\mathbf{k}} | \Sigma(\omega) | \Psi_{n\mathbf{k}} \rangle \Big|_{\omega=\epsilon^{LDA}_{n\mathbf{k}}} . \tag{5.5}$$

Because of the analytic fit of the expectation values of the self-energy the derivative with respect to the energy argument is readily found in closed form. For ϵ_s a closed form but rather involved expression can also be given. Comparison of the self-consistent solution of Eq. (5.3) and the result of the Taylor expansion Eq. (5.4) shows that the latter is usually a very good approximation, yielding quasiparticle energies differing by only up to 5 meV from the solution of Eq. (5.3) for bulk Si.

VI. RESULTS FOR BULK SILICON AND SILICON SUPERCELLS

In order to test the performance and the scaling properties of the space-time *GW* method we have performed calculations of self-energy and quasiparticle energies for bulk silicon and silicon with artificially increased unit cell sizes. The latter were slab-like supercells containing four and eight atoms, respectively, with tetragonal symmetry, i. e., the unit cells were artificially enlarged in one direction. Supercells of this geometry can be used to model semiconductor superlattices and surfaces.

A. Convergence parameters

Before discussing the performance and scaling issues we briefly summarize the results of convergence tests for bulk Si. In Table I the parameters needed to converge quasiparticle energy differences to an accuracy of 50 meV and 20 meV, respectively, with the present method are gathered. This accuracy is meant to be with respect to each parameter individually. We estimate the overall accuracy of our results to be better than 0.1 eV. The plane-wave (PW) cutoff in Table I is the cutoff of $\frac{1}{2}(\mathbf{k}+\mathbf{G})^2$ in the LDA calculation providing the wavefunctions $\Psi_{n\mathbf{k}}$. All PW components of the LDA wave functions are used in the *GW* calculation. The real-space grids in the unit cell (see Section III B) resulting from PW cutoffs of 13.5 Ry and 16 Ry comprise $9\times9\times9$ and $12\times12\times12$ points, respectively. The band cutoff determines the number of bands included in the band summation in Eq. (3.3) for the calculation of the Green's function. Band cutoffs of 8 and 10 Ry correspond to taking 109 and 145 bands, respectively, into account. For sampling the BZ we employ a regular non-offset grid of $\mathbf{k}$ points. The LDA wavefunctions are calculated at the $\mathbf{k}$ points in the irreducible wedge of the BZ. As described in Section III B the dimensions of the $\mathbf{k}$ grid in the BZ are equal to the dimensions (in multiples of primitive lattice vectors) $N_i^{\mathbf{R}}$ of the IC, which must be large enough to comprise the range of non-locality of the Green's function and the self-energy in Si. The smaller $\mathbf{k}$ grid given in Table I can roughly accomodate a non-locality range of 14.5 a.u. and the larger of 18 a.u. $(-\tau_{max}, +\tau_{max})$ and $\Delta\tau$ stand for the range and spacing of the equidistant grid for sampling the functions $F(\mathbf{r}, \mathbf{r}', i\tau)$ on the imaginary time axis. The parameters ω_{max} and $\Delta\omega$ of the corresponding grid in the imaginary energy domain are related to the time-grid parameters by $\Delta\omega = 2\pi/(2N_\tau - 1)\Delta\tau$ with $\tau_{max} = N_\tau\Delta\tau$.

B. Quasiparticle energies for bulk silicon

The quasiparticle energies for bulk Si obtained from a well converged calculation with the present method are given in Table II [27]. The LDA calculation was done with the PW pseudopotential method employing a pseudopotential constructed according to the prescription of Hamann [28] and using the Ceperley-Alder exchange-correlation potential [29] in the parametrization of Perdew and Zunger [30]. As can be seen from Table II the calculated *GW* quasiparticle energies agree well with quasiparticle excitation energies derived from photoemission, inverse photoemission and optical experiments [31–37], except for the position of the bottom of the valence band which appears to be too high in our calculation. The agreement of the quasiparticle energy differences obtained with the present method (*GW* 1 in Table II) with those of a recent *GW* calculation by Fleszar and Hanke [38] is very good. Like us, these authors did not make use of a plasmon pole model to describe the energy dependence of the dynamically screened Coulomb interaction but directly computed its full energy dependence.

We investigated how the quasiparticle energies change when the Green's function is updated by replacing the LDA eigenvalues by the quasiparticle energies and the self-energy recalculated with the updated Green's function as input, as was done by Hybertsen and Louie [3]. The results (*GW* 2) are also included in Table II. Updating the Green's function increases the gap by about 0.1 eV. Now the conduction band energies agree within better than 0.1 eV with those given in Ref. [3] whereas the valence band energies remain higher by between 0.1 eV and 0.5 eV (relative to the top of the valence band), the difference increasing with the distance of the respective state from the Fermi level.

Indeed, the results of several earlier *GW* calculations [3,39] are somewhat different from those given in Table II (and Ref. [38]). This can be attributed to the following reasons:

(i) a possible lack of convergence with respect to the number of conduction bands taken into account in the earlier calculations, causing the topmost occupied state (which is particularly sensitive to this

parameter) to be about 0.2 eV too high, as discussed in Ref. [38]; and

(ii) the use of plasmon pole models for the energy dependence of the dynamically screened Coulomb interaction in the earlier calculations.

C. Silicon supercells

In Table IV we compare the quasiparticle energies for slab-like Si_4 and Si_8 supercells to the bulk Si results. Shown are the quasiparticle energies for the high-symmetry **k** points of Si (which are not necessarily symmetry points for the supercell geometry) and for those **k** points which are mapped onto these symmetry points when the unit cell size is increased. The calculations were done with a PW cutoff of 13.5 Ry (leading to a real-space grid spacing of $\Delta r = 0.8$ a.u.), a band cutoff of 10 Ry, $\Delta\tau = 0.3$ a.u. and $\tau_{max} = 20$ a.u. The **k** grids have been chosen to make the IC of equal size in all three dimensions for all the calculations, for the reason discussed in Section III B. The grid sizes resulting from these parameters are summarized in Table III. N_{IUC}, N_{UC}, and N_{IC} are the numbers of real-space grid points in the irreducible part of the unit cell, the unit cell and the interaction cell, respectively (see Section IV above). N_k stands for the number of **k** points in the BZ and $N_{t/\omega}$ is the number of positive time/energy points. The supercell results in Table IV agree with the bulk Si results, as expected. The small differences for the highest conduction band given in Table IV are ascribed to the lower symmetry of the supercells leading to different **G** vector sets.

D. Scaling

As outlined in Section IV the computational effort in the present method should scale quadratically with the number of atoms in the unit cell. Fig. 6 shows the CPU times for a full *GW* calculation for Si_n (n = 2, 4, 8) on a Digital Alpha 500/500 workstation. The parameters of these calculations are those given in Table III and the discussion thereof. Si_2 (bulk Si) has a higher symmetry than the tetragonal supercells Si_4 and Si_8. As the symmetry is exploited in most parts of the calculation this saves approximately a factor of two in CPU time (cf. N_{IUC} values in Table III). Taking this into account the overall scaling is indeed quadratic, as expected. The most time consuming parts of the calculation are:

(i) setting up the Green's function according to Eq. (3.3)

(ii) calculating the dynamically screened Coulomb interaction (including the inversion of the dielectric matrix [45], the transformation to and from reciprocal space, and reading of P^0 from and writing of W to disk)

(iii) computation of the matrix elements of the energy dependent self-energy for a given number of **k** points and bands.

Optionally, the Green's function can be recomputed when the self-energy is calculated, rather than storing it on disk when it is first set up and reading it in for the calculation of the self-energy. This procedure reduces the required disk space by a factor of 2/3. It has been used for the Si supercell calculations described here. As can be seen from the breakdown of the total CPU time shown in Fig. 6 all these parts of the calculation scale roughly like N^2. More precisely, the scaling is like $N_{IUC}N_{bands}$ for (i), $N_{IUC}N_{UC}$ for (ii) and N_{UC}^2 for (iii) as was discussed in Section IV.

Comparing the performance of our method with conventional techniques we note that although a number of quasiparticle calculations within the GW approximation have been reported for systems with up to 60 atoms, e. g. for surfaces [40–42], defects [43], and fullerenes [44], several simplifying approximations have been employed in these calculations in order to reduce the computational effort. These authors focus on the calculation of quasiparticle energies and therefore only a number of self-energy matrix elements were computed. This is in contrast to the present method which provides the full self-energy (thus giving acccess to quantities other than the quasiparticle energies, see conclusions section below). In all the calculations mentioned above plasmon-pole models have been employed to approximate the energy dependence of the dynamically screened Coulomb interaction. In some cases [41,44] a model static dielectric matrix has been used as well. To aid comparison we note that although the plane-wave cutoff of the underlying LDA pseudopotential calculation can be (and has been in many GW calculations reported in the literature) considerably reduced in the GW calculation because the quasiparticle corrections usually converge much faster with this parameter than the LDA eigenvalues, this has not been done in the test calculations reported in the present paper. The energy dependence of the screened interaction is fully taken into account in the method proposed in this paper. Hence the dielectric matrix is computed and inverted for about 60 to 100 imaginary energies which is computationally much more demanding than using a plasmon-pole model. We estimate that the crossover to the advantegeous scaling behavior of CPU time in our method occurs (a) already for bulk Si if we compare with a calculation of the full self-energy with a conventional reciprocal-space approach and (b) for systems in the range studied in this paper if we compare with a reciprocal-space method where a plasmon-pole model is used and which is restricted to computing a moderate number of self-energy matrix elements only. Finally, we mention that work, particularly regarding the treatment of the time/energy dependence of the key quantities, is in progress to reduce the prefactor of the scaling of our method.

VII. SUMMARY

We have presented a detailed account of a method for calculating the electron self-energy and related quantities from ab-initio many-body perturbation theory within the *GW* approximation. The method is based on representing the basic quantities (Green's function, dielectric response function, dynamically screened Coulomb interaction and self-energy) on a real-space grid and on the imaginary time axis. In those intermediate steps of the calculation where it is computationally more efficient to work in reciprocal space and imaginary energy we change to the latter representation by means of fast Fourier transforms. Working on the imaginary time/energy axis considerably facilitates the numerical treatment. The matrix elements of the self-energy on the real-energy axis are obtained by an analytic continuation procedure which was shown to be accurate and stable. We have demonstrated the accuracy of the method by calculating quasiparticle excitation energies at high-symmetry points for the prototype semiconductor Si. The computational effort of the method scales quadratically with unit cell size. This was shown explicitly by calculations for Si supercells. The method allows the extension of ab-initio work beyond the calculation of quasiparticle energies and its application to materials with larger basis sets or larger unit cells than were previously feasible. Calculating the full self-energy gives access to quantities like the charge density [15], spectral function [16], momentum distribution and total energy at the *GW* level. It also opens the possibility of calculating the self-energy self-consistently and to provide the *GW* Green's function (after solving the Dyson equation) – essential prerequisites for going beyond the *GW* approximation, e. g., by iterating Hedin's equations. This remains to be investigated in future.

VIII. ACKNOWLEDGMENTS

We acknowledge helpful discussions with R. J. Needs and A. Rubio. This work was supported by the Engineering and Physical Sciences Research Council. M. M. Rieger acknowledges financial support by the European Union under the TMR scheme.

---

‡ Present address: Siemens AG, Semiconductor Group, Balanstraße 73, 81617 München, Germany. Electronic address: rieger@hl.siemens.de

‖ Permanent address: Universidad Privada Boliviana, Casilla 3967, Cochabamba, Bolivia

¶ Electronic address: rwg3@york.ac.uk

[1] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[2] M. S. Hybertsen and S. G. Louie, Phys. Rev. Lett. **55**, 1418 (1985)
[3] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).
[4] L. Hedin, Phys. Rev. **139**, A796 (1965).
[5] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. Lett. **56**, 2415 (1986).
[6] X. Zhu and S. G. Louie, Phys. Rev. B **43**, 14142 (1991).
[7] S. G. Louie, Surf. Sci. **299/300**, 346 (1994).
[8] G. Onida *et al.*, Phys. Rev. Lett. **75**, 818 (1995).
[9] E. C. Ethridge, J. L. Fry, and M. Zaider, Phys. Rev. B **53**, 3662 (1996).
[10] H. J. de Groot, P. A. Bobbert, and W. van Haeringen, Phys. Rev. B **52**, 1100 (1995).
[11] U. von Barth and B. Holm, Phys. Rev. B **54**, 8411 (1996).
[12] E. L. Shirley, Phys. Rev. B **54**, 7758 (1996).
[13] W.-D. Schöne, A. G. Eguiluz and J. A. Gaspar, Bull. Am. Phys. Soc. **43**, 170 (1998).
[14] B. Holm and U. von Barth, Phys. Rev. B **57**, 2108 (1998).
[15] M. M. Rieger and R. W. Godby, Phys. Rev. B **58**, 1343 (1998).
[16] H. N. Rojas, R. W. Godby, and R. J. Needs, Phys. Rev. Lett. **74**, 1827 (1995).
[17] Note that the signs of the exponents and of τ were given incorrectly in Eq. (3) of Ref. [16].
[18] S. Baroni and R. Resta, Phys. Rev. B **33**, 7017 (1986).
[19] The interaction cell used in the earlier paper, Ref. [16], was a sphere which slightly exceeded the Wigner-Seitz cell of the IC lattice corresponding to the k-point grid used. Because the use of a discrete **k**-grid imposes an artificial periodicity on the non-local quantities as a function of $|\mathbf{r}-\mathbf{r}'|$, this is not optimal. The use of the WS cell in the present paper utilises maximum information while avoiding double-counting arising from the periodicity.
[20] X. Blase, A. Rubio, S. G. Louie, and M. L. Cohen, Phys. Rev. B **52**, R2225 (1995).
[21] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **37**, 10159 (1988).
[22] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **35**, 5585 (1987).
[23] R. M. Pick, M. H. Cohen, and R. M. Martin, Phys. Rev. B **1**, 910 (1970).
[24] We define this quantity as the difference between the GW self-energy and the energy-independent bare-exchange part of the self-energy.
[25] I. D. White, R. W. Godby, M. M. Rieger and R. J. Needs, Phys. Rev. Lett. **80**, 4265 (1998).
[26] A. Schindlmayr, Phys. Rev. B **56**, 3528 (1997).
[27] These results are somewhat different from those reported in Ref. [15]. The calculations there still used a spherical cutoff for the interactions as described in Ref. [16], instead of the interaction cell defined by the **k** point grid. Tests indicate that the discrepancy arises from the fact that the spherical cutoff in $|\mathbf{r}-\mathbf{r}'|$ used in Ref. [16] is not quite consistent with the **k**-grid used (see Section III B above).
[28] D. R. Hamann, Phys. Rev. B **40**, 2980 (1989).
[29] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
[30] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048

(1981).
[31] *Numerical Data and Functional Relationships in Science and Technology*, Vol. 17a of *Landolt-Börnstein, New Series, Group III*, edited by K.-H. Hellwege and O. Madelung (Springer, Berlin, 1982).
[32] J. E. Ortega and F. J. Himpsel, Phys. Rev. B **47**, 2130 (1993).
[33] W. E. Spicer and R. C. Eden, in *Proceedings of the 9th Int. Conf. on the Phys. of Semicond.*, edited by S. M. Ryvkin (Nauka, Moskow, 1968), Vol. 1, p. 68.
[34] A. L. Wachs *et al.*, Phys. Rev. B **32**, 2326 (1985).
[35] F. J. Himpsel, P. Heimann, and D. E. Eastman, Phys. Rev. B **24**, 2003 (1981).
[36] R. Hulthen and N. G. Nilsson, Solid State Commun. **18**, 1341 (1976).
[37] D. Straub, L. Ley, and F. J. Himpsel, Phys. Rev. Lett. **54**, 142 (1985).
[38] A. Fleszar and W. Hanke, Phys. Rev. B **56**, 10228 (1997).
[39] M. Rohlfing, P. Krüger, and J. Pollmann, Phys. Rev. B **48**, 17791 (1993).
[40] M. S. Hybertsen and S. G. Louie, Phys. Rev. Lett. **58**, 1551 (1987).
[41] J. E. Northrup, Phys. Rev. B **47**, 10032 (1993).
[42] M. Rohlfing, P. Krüger, and J. Pollmann, Phys. Rev. B **54**, 13759 (1996).
[43] M. P. Surh, H. Chacham, and S. G. Louie, Phys. Rev. B **51**, 7464 (1995).
[44] E. L. Shirley and S. G. Louie, Phys. Rev. Lett. **71**, 133 (1993).
[45] Although (ii) forms a major part of the calculation, the inversion of the dielectric matrix itself accounts for only between 2 % (Si_2) and 4 % (Si_8) of the total CPU time, as a consequence of doing this inversion in reciprocal space as discussed in Section IV B.

FIG. 1. The grids and cells used in the calculation in (a) real and (b) reciprocal space. These are shown here schematically in two dimensions for the case of a 2×2 grid of **k**-points (corresponding to an interaction cell of 2×2 unit cells), and a 3×3 real-space grid in the unit cell (corresponding to a 3×3 grid of reciprocal lattice vectors in reciprocal space). The bare Coulomb interaction $1/|\mathbf{r}-\mathbf{r}'|$ is non-periodic on the interaction cell, and its amplitude at a real-space-grid point is taken to be that at the corresponding point in the Wigner-Seitz cell around **r** of the IC lattice. All other quantities, in both real and reciprocal space, are periodic. In these cases the choice of the primitive (shown) or Wigner-Seitz cell is a matter of computational convenience.

FIG. 2. Fitting of the matrix elements of the correlation self-energy $\langle\Sigma_c(i\omega)\rangle$ of silicon as a function of imaginary energy. The real (top panel) and imaginary (lower panel) parts are both reproduced extremely accurately by the fits (which on this scale are indistinguishable from the calculated values). The bands shown are band 1 at the Γ point (solid line for fitted and calculated function), bands 2-4 at Γ (dotted line), bands 3-4 at X (dashed) and bands 5-6 at X (dot-dashed).

FIG. 3. Convergence of a matrix element (degenerate bands 2-4 at Γ) of the correlation self-energy $\langle\Sigma_c(i\omega)\rangle$ of silicon with respect to ω_{max} with a fixed energy grid spacing of $\Delta\omega$= 0.16 Har. The top panel shows the calculated self-energy on the imaginary axis with the analytically continued dependence on real energy shown in the lower panel. The lines correspond to ω_{max}= 5 Har (dotted), 10 Har (dashed) and 20 Har (solid). Changes in $\langle\Sigma_c(\omega)\rangle$ are directly related to changes in the calculated $\langle\Sigma_c(i\omega)\rangle$, rather than any instabilities in the analytic continuation technique.

FIG. 4. Convergence of $\langle\Sigma_c(i\omega)\rangle$ with respect to energy grid spacing $\Delta\omega$ with fixed ω_{max}= 10 Har. The lines correspond to $\Delta\omega$= 0.24 Har (dotted), 0.16 Har (dashed) and 0.08 Har (solid). As in the previous Figure the analytically continued matrix element converges well, indicating the stability of the fitting procedure. The matrix elements on the real-energy axis (which are obtained in form of fits to a model function, see text) are plotted with a fixed grid spacing of 0.04 Har in order to facilitate comparison between the three curves.

FIG. 5. Real part of a matrix element (degenerate bands 2-4 at Γ) of the correlation self-energy $\langle\Sigma_c(\omega)\rangle$ on the real energy axis, computed with ω_{max} = 5 Har (dot-dashed line), 10 Har (dashed line), and 20 Har (long-dashed line) with fitting $\langle\Sigma_c(i\omega)\rangle$ on the imaginary energy axis using the energy range $(-\omega_{max},+\omega_{max})$ and with ω_{max} = 10 Har (dotted line) and 20 Har (solid line) using a fixed energy range of (-5 Har, $+5$ Har) for the fitting.

FIG. 6. Scaling of the CPU time on a Digital Alpha 500/500 workstation with respect to unit cell size for Si_n (n = 2,4,8). Besides the total CPU time ($\circ$) the times for three major parts of the computation are given: calculation of (i) the Green's function ($\Box$), (ii) the dynamically screened Coulomb interaction ($\diamond$), and (iii) the matrix elements of the self-energy including a recomputation of the Green's function ($\triangle$), see text. Note the quadratic spacing of the abscissa.

TABLE I. Convergence of quasiparticle energies for bulk Si with respect to cutoff and grid parameters in the present method. Shown are the parameters needed for a convergence of 50 meV and 20 meV, respectively.

parameter	50 meV	20 meV
plane-wave cutoff (in Ry)	13.5	16
band cutoff (in Ry)	8	10
size of **k** grid	$4\times4\times4$	$5\times5\times5$
spacing Δt of time grid (in a.u.)	0.3	0.15
range t_{max} of time grid (in a.u.)	13.	20.

TABLE II. Calculated quasiparticle energies at points of high symmetry for Si (in eV). The valence band maximum has been set to zero. The calculation was done with a plane-wave cutoff of 16 Ry, a $5 \times 5 \times 5$ k-point grid, a band cutoff of 10 Ry and a time grid with $\Delta\tau = 0.3$ a.u. and $\tau_{max} = 20$ a.u. For comparison the eigenvalues obtained in the LDA calculation providing the input for the *GW* calculation have been included. *GW* 1 refers to a *GW* calculation using the LDA Green function whereas in *GW* 2 the Green function has been updated by replacing the LDA eigenvalues by the quasiparticle energies obtained in *GW* 1. In the last column some experimental data are given (see text).

band	LDA	*GW* 1	*GW* 2	Experiment
Γ_{1v}	-11.89	-11.57	-11.58	-12.5 ± 0.6[a]
Γ'_{25v}	0.00	0.00	0.00	0.00
Γ'_{15c}	2.58	3.24	3.32	3.40[a], 3.05[b]
Γ'_{2c}	3.28	3.94	4.02	4.23[a], 4.1[b]
X_{1v}	-7.78	-7.67	-7.68	
X_{4v}	-2.82	-2.80	-2.81	-2.9[c], -3.3 ± 0.2[d]
X_{1c}	0.61	1.34	1.42	1.25[b]
X_{4c}	10.11	10.54	10.63	
L'_{2v}	-9.57	-9.39	-9.40	-9.3 ± 0.4[a]
L_{1v}	-6.96	-6.86	-6.88	-6.7 ± 0.2[a]
L'_{3v}	-1.17	-1.17	-1.17	-1.2 ± 0.2[a], -1.5[e]
L_{1c}	1.46	2.14	2.22	2.1[f], 2.4 ± 0.15[g]
L_{3c}	3.33	4.05	4.14	4.15 ± 0.1[g]
L_{2c}	7.71	8.29	8.39	
E_{gap}	0.49	1.20	1.28	1.17[a]

[a]Ref. [31].
[b]Ref. [32].
[c]Ref. [33].
[d]Ref. [34].
[e]Ref. [35].
[f]Ref. [36].
[g]Ref. [37].

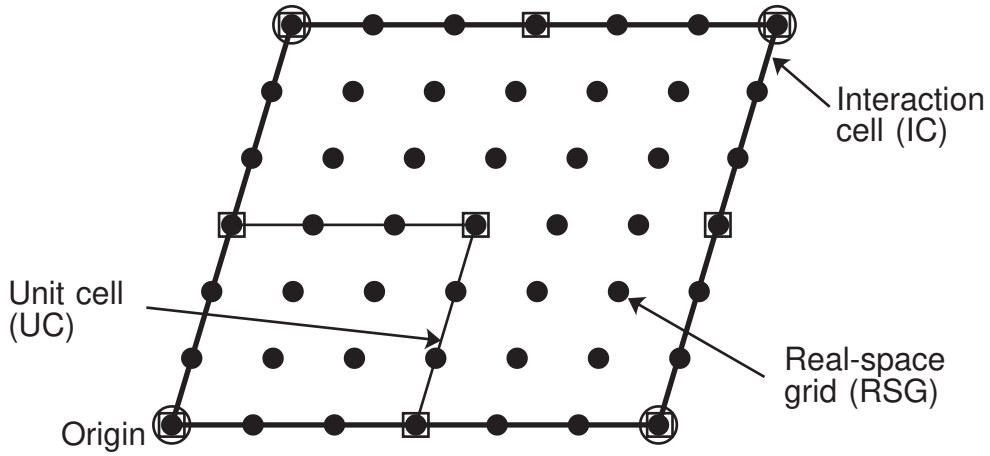
TABLE III. Grid parameters for the Si_n ($n = 2, 4, 8$) calculations (see text).

parameter	Si_2 (bulk Si)	Si_4	Si_8
N_{IUC}	55	230	455
N_{UC}	$9 \times 9 \times 9$	$9 \times 9 \times 18$	$9 \times 9 \times 36$
N_k	$4 \times 4 \times 4$	$4 \times 4 \times 2$	$4 \times 4 \times 1$
$N_{IC} = N_{UC}N_k$	$36 \times 36 \times 36$	$36 \times 36 \times 36$	$36 \times 36 \times 36$
$N_{t/\omega}$	63	63	63

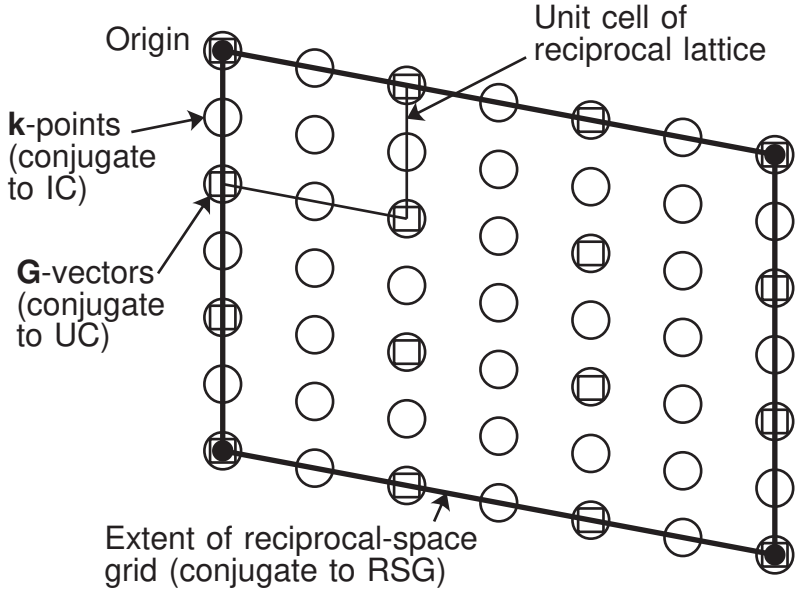
TABLE IV. Comparison of quasiparticle energies (in eV) at high symmetry points for bulk Si with the corresponding values obtained from calculations for Si_4 and Si_8 supercells (see text). All energies refer to the respective valence band top. Shown are the band energies for all k-points mapping onto Γ, X, and L, respectively, when the unit cell size is increased, where those k-points appear in the calculation.

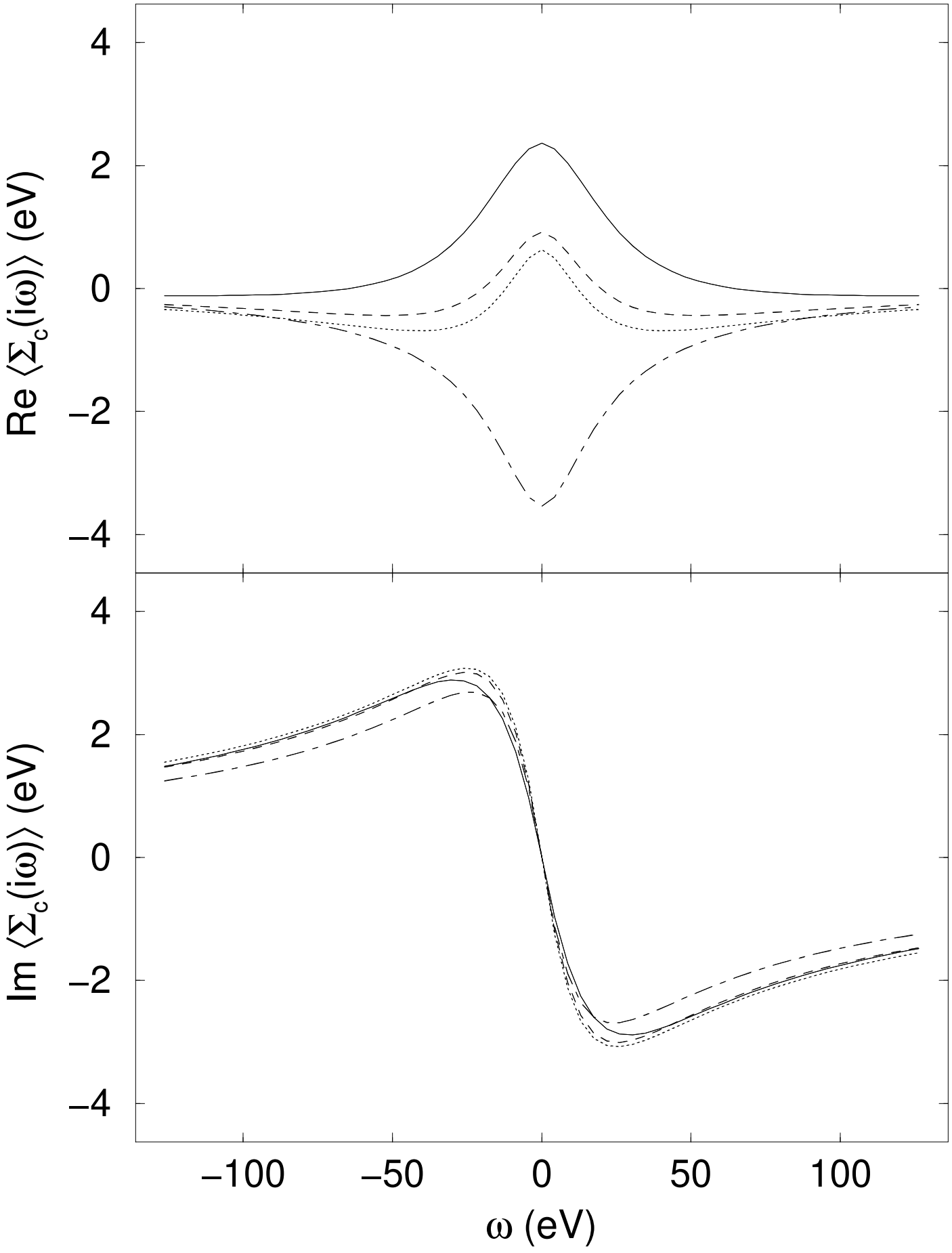
	bulk Si			Si_4		Si_8		
Γ	-11.55	-7.63	-10.53	-11.53	-7.61	-11.55	-7.63	-10.53
	0.00	-2.78	-3.41	0.00	-2.78	0.00	-2.78	-3.41
	3.23	1.35	-1.81	3.22	1.37	3.22	1.37	-1.81
	3.97	10.52	1.77	3.96	10.45	3.96	10.43	1.78
			3.82					3.82
			6.35					6.30
L	-9.35	-9.62		-9.34		-9.36	-9.63	
	-6.83	-5.35		-6.81		-6.83	-5.35	
	-1.16	-3.62		-1.16		-1.16	-3.62	
	2.17	-1.27		2.18		2.18	-1.27	
	4.04	3.13		4.03		4.02	3.12	
	8.28	5.46		8.26		8.25	5.42	
		5.66					5.63	
		6.75					6.70	
X	-7.63	-7.46		-7.61		-7.64	-7.47	
	-2.78	-3.73		-2.78		-2.79	3.73	
	1.35	4.84		1.37		1.36	4.82	
	10.52	5.67		10.44		10.42	5.63	

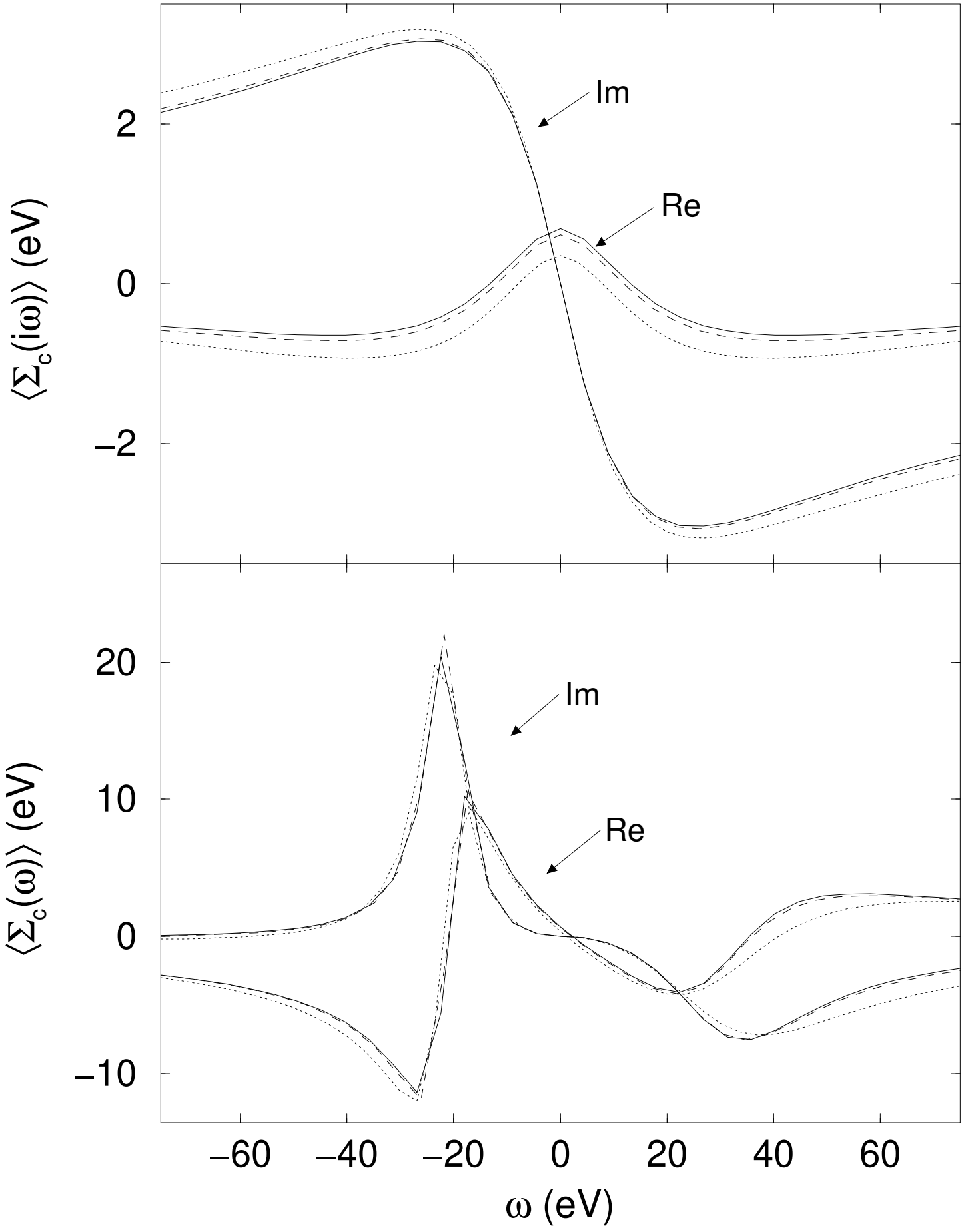
(a) Real Space

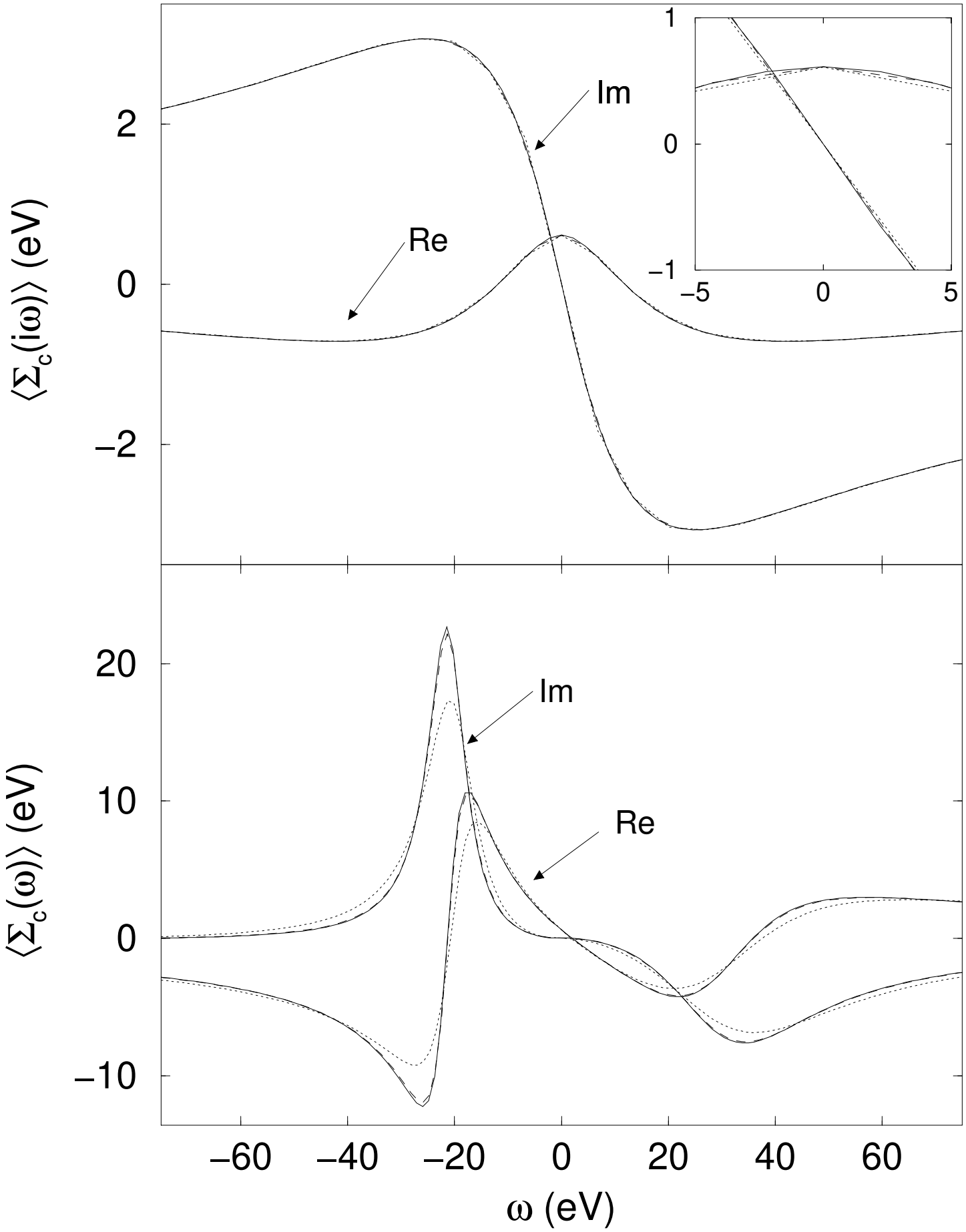


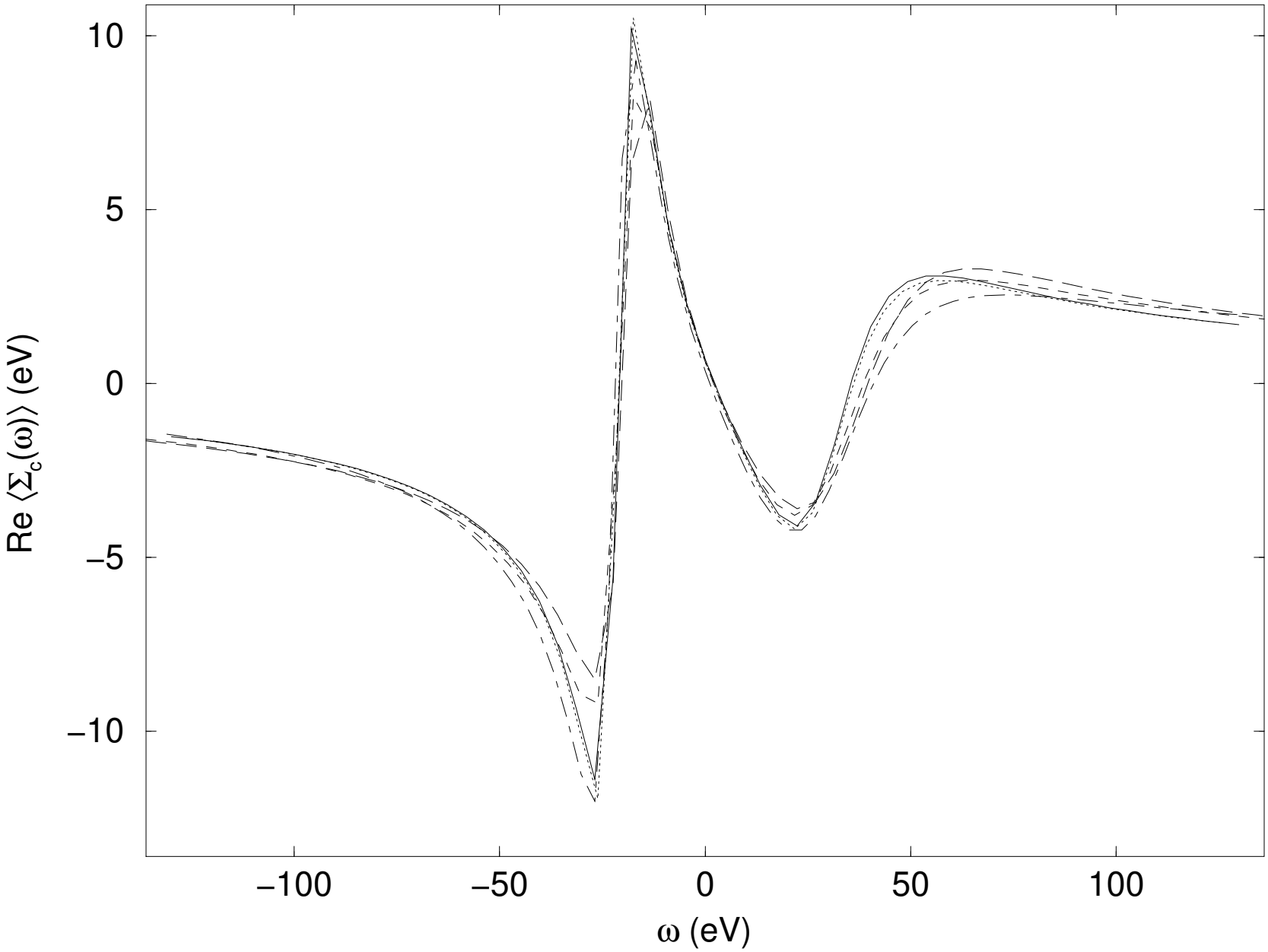
(b) Reciprocal Space

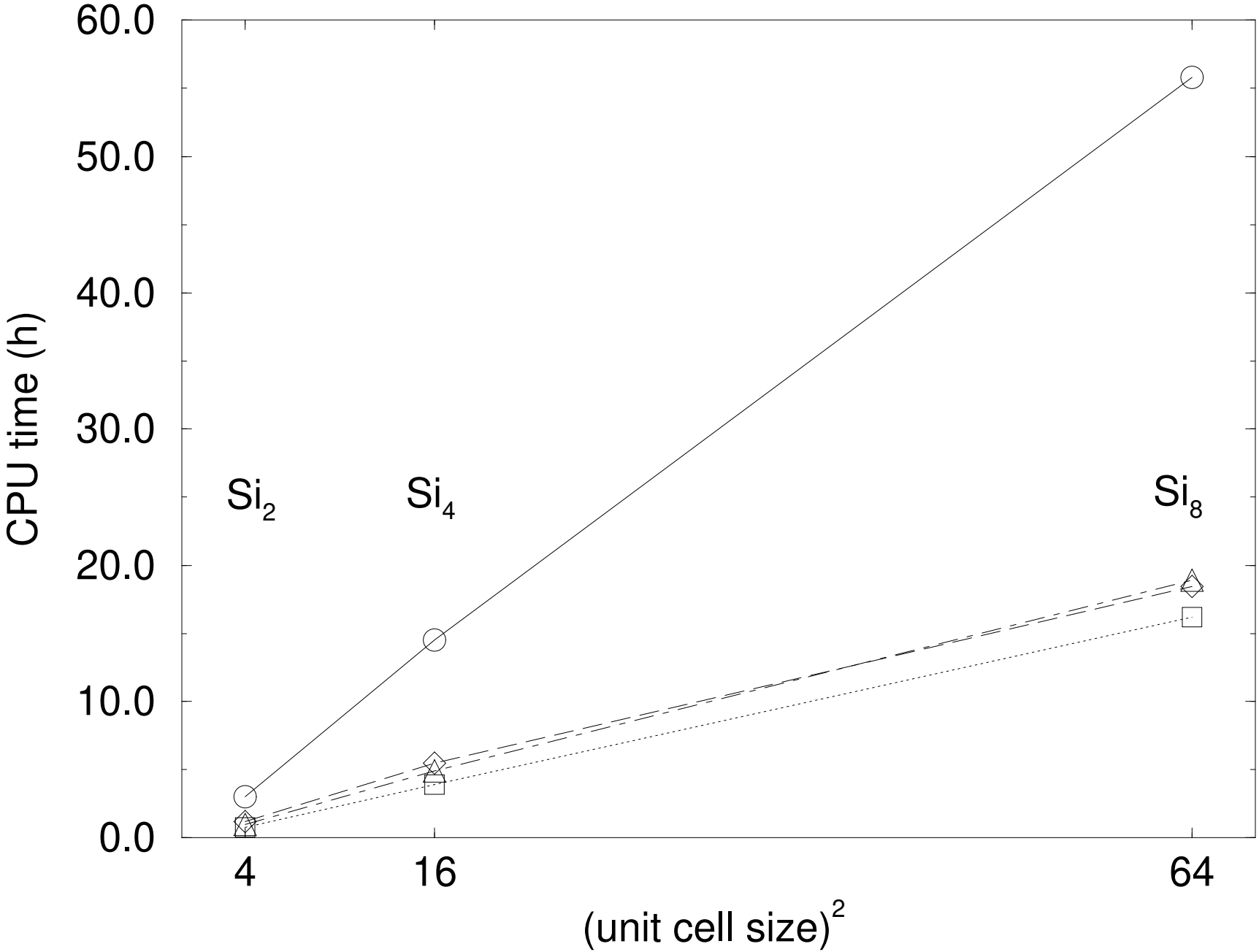



4
2
0
−2
−4
Re ⟨Σc(iω)⟩ (eV)
4
2
0
−2
−4
Im ⟨Σc(iω)⟩ (eV)
−100
−50
0
50
100
ω (eV)


Im
Re
$\langle\Sigma_c(i\omega)\rangle$ (eV)
2
0
−2
Im
Re
$\langle\Sigma_c(\omega)\rangle$ (eV)
20
10
0
−10
−60
−40
−20
0
20
40
60
ω (eV)


Im
Re
$\langle\Sigma_c(i\omega)\rangle$ (eV)
2
0
−2
1
0
−1
−5
0
5
Im
Re
$\langle\Sigma_c(\omega)\rangle$ (eV)
20
10
0
−10
−60
−40
−20
0
20
40
60
ω (eV)


10
5
0
−5
−10
−100
−50
0
50
100
Re ⟨Σc(ω)⟩ (eV)
ω (eV)


60.0
50.0
40.0
30.0
20.0
10.0
0.0
CPU time (h)
Si2
Si4
Si8
4
16
64
(unit cell size)2