Space-Time Method for *Ab Initio* Calculations of Self-Energies and Dielectric Response Functions of Solids

H. N. Rojas,[*] R. W. Godby, and R. J. Needs

Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE, United Kingdom
(Received 9 August 1994)

We present a new method for efficient, accurate calculations of many-body properties of periodic systems. The main features are (i) use of a real-space/imaginary-time representation, (ii) avoidance of any model form for the screened interaction W, (iii) exact separation of W and the self-energy Σ into short- and long-ranged parts, and (iv) the use of novel analytical continuation techniques in the energy domain. The computer time scales approximately linearly with system size. We give results for jellium and silicon, including the spectral function of silicon obtained from the Dyson equation.

Many-body perturbation theories for condensed-matter physics allow the Green's functions of a system of interacting electrons to be formulated in terms of those of a hypothetical similar system of noninteracting electrons moving in an effective potential. The key quantity that connects the two is the exchange-correlation self-energy operator Σ. From the Green's functions, most properties of the system can be calculated, most notably the quasiparticle energies, the spectral function, the electron density, the momentum distribution, and the total energy. Such theories have usually been developed first for the homogeneous electron gas (jellium), whose translational symmetry makes two-point functions such as the self-energy $\Sigma(\mathbf{r}, \mathbf{r}')$ diagonal in a momentum representation $\Sigma(k)$. Also, because experiments generally focus on energy-dependent measurements, the natural representation of time dependence $\Sigma(\tau = t - t')$ is in the energy (or frequency) domain $\Sigma(\omega)$, yielding the total function $\Sigma(k, \omega)$. In a periodic system the replacement of continuous translational symmetry by discrete translational symmetry turns the functions into matrices $\Sigma_{\mathbf{GG}'}(\mathbf{k}, \omega)$, where $\mathbf{G}$ and $\mathbf{G}'$ are reciprocal lattice vectors, and this is the representation that has been used for practical calculations [1,2].

Four observations form the starting point for our method. First, the commonly used *GW* approximation for the self-energy operator [3] (the first term in an iterative solution of coupled equations relating the many-body quantities) is a computationally expensive multidimensional convolution in a momentum-energy representation, but simply multiplicative when written in terms of space and time,

$$\Sigma(\mathbf{r}, \mathbf{r}', \tau) = iG(\mathbf{r}, \mathbf{r}', \tau)W(\mathbf{r}, \mathbf{r}', \tau), \qquad (1)$$

where G is the one-particle Green's function and W the screened Coulomb interaction. Second, quantities such as Σ, W, and G, which contain much structure along the real energy and time axes, may rigorously be analytically continued to the imaginary energy (e.g., Ref. [2]) or time axis, where all the relevant information is presented in a much smoother form. In particular, oscillatory quantities

in real time become smooth, monotonic functions in imaginary time or energy. Third, fast Fourier transform (FFT) methods allow extremely efficient implementation of a discrete Fourier transform to go between space and momentum, and energy and time. Fourth, the self-energy is known to go to zero as $|\mathbf{r} - \mathbf{r}'| \to \infty$, so that a finite cutoff in $|\mathbf{r} - \mathbf{r}'|$ may be used in numerical work.

The calculation of the screened Coulomb interaction W begins with the expression for the noninteracting density response function

$$\chi_0(\mathbf{r}, \mathbf{r}', \tau) = -iG_0(\mathbf{r}, \mathbf{r}', \tau)G_0(\mathbf{r}', \mathbf{r}, -\tau), \qquad (2)$$

where G_0 is the noninteracting Green's function. In our work, as in most modern *ab initio* applications of many-body perturbation theory, our "zeroth order" noninteracting system is that of the local-density approximation (LDA) to density-functional theory (DFT), where exchange and correlation are described by a local potential. We analytically continue G_0 in its standard eigenfunction expansion from real to imaginary energy, and then take the Fourier transform from imaginary energy to imaginary time, yielding

$$G_0(\mathbf{r}, \mathbf{r}', i\tau) = \begin{cases} i \sum_{n\mathbf{k}}^{\text{occ}} \psi_{n\mathbf{k}}(\mathbf{r})\psi_{n\mathbf{k}}^*(\mathbf{r}') \exp(-\varepsilon_{n\mathbf{k}}\tau), \\ \qquad\qquad\qquad\qquad\qquad \tau < 0, \\ -i \sum_{n\mathbf{k}}^{\text{unocc}} \psi_{n\mathbf{k}}(\mathbf{r})\psi_{n\mathbf{k}}^*(\mathbf{r}') \exp(-\varepsilon_{n\mathbf{k}}\tau), \\ \qquad\qquad\qquad\qquad\qquad \tau > 0, \end{cases}$$
$$(3)$$

where $\psi_{n\mathbf{k}}$ and $\varepsilon_{n\mathbf{k}}$ are the LDA one-electron eigenfunctions and eigenvalues; the zero of energy is taken at the Fermi energy (or in an insulator at the center of the band gap), and the sum is over occupied or unoccupied states depending on the sign of τ. This may be evaluated conveniently by a band summation, and convergence is rapid because of the real, decaying exponentials. After a basic three-dimensional real-space grid has been chosen (with spacing Δr), $\mathbf{r}$ runs over all grid points in the irreducible wedge of the real-space unit cell, and $\mathbf{r}'$ runs over all grid points (offset so as to exclude $\mathbf{r}' = \mathbf{r}$ for numerical reasons) inside a sphere of radius $R_{\max}$ centered on $\mathbf{r}$.

Similarly, the time grid is an equally spaced, offset grid between $\pm\tau_{max}$ with spacing $\Delta\tau$. It is then a simple matter to form χ_0 according to Eq. (2).

In order to calculate the dielectric matrix ε and screened Coulomb interaction W, we take the six-dimensional Fourier transform (6DFFT) [4] from $(\mathbf{r}, \mathbf{r}')$ to $(\mathbf{k}, \mathbf{G}, \mathbf{G}')$, and the one-dimensional Fourier transform (1DFFT) from imaginary time to imaginary energy, and evaluate the expressions

$$\varepsilon(\mathbf{k}, i\omega) = 1 - \nu\chi_0(\mathbf{k}, i\omega), \qquad W(\mathbf{k}, i\omega) = \nu\varepsilon^{-1}(\mathbf{k}, i\omega),$$
$$\tag{4}$$

where matrix multiplication in the subscripts $(\mathbf{G}, \mathbf{G}')$ is implied, and ν is the bare Coulomb interaction $\nu_{\mathbf{G}\mathbf{G}'}(\mathbf{k}) = 4\pi\delta_{\mathbf{G}\mathbf{G}'}/|\mathbf{k} + \mathbf{G}|^2$. (The above expressions correspond to the random-phase approximation (RPA); an equally practical alternative at this stage is the effective screened Coulomb interaction $\tilde{W} = \nu[1 - \chi_0(\nu + K_{xc})]^{-1}$, where K_{xc} is the second derivative of the LDA exchange-correlation energy, which has been argued to be a more consistent choice within the assumptions of the GW approximation [5,6].) After forming $W_{\mathbf{G}\mathbf{G}'}(\mathbf{k}, i\omega)$ we take the 6DFFT back to the real-space representation, and the 1DFFT to the imaginary-time domain. The net computational effort is much reduced in comparison with the usual techniques, primarily because the costly double summation over k points and bands does not appear. Furthermore, the need for plasmon-pole models to represent the ω dependence of W is eliminated: The FFT provides W simultaneously for a large number of imaginary energies or times.

Although we actually require W at the imaginary times that this provides, we note that an alternative at this stage, should W be required at real energies, would be to analytically continue from the imaginary ω axis to the real ω axis. There is already some experience of the reliability of analytic continuation of W in the ω plane [7], and our novel continuation techniques for Σ below are also applicable to W because of the similar analytic structure of W and Σ.

The screened Coulomb interaction is, of course, relatively long ranged: At large distances the limiting behavior of all elements (with the sole exception of the zero-frequency screening in a metal) is proportional to $1/|\mathbf{r} - \mathbf{r}'|$. In order to avoid problems in the 6DFFT associated with the long-range tail and in the 1DFFT associated with the asymptotic frequency dependence, we write (without loss of generality)

$$W(\mathbf{r}, \mathbf{r}', i\tau) = \frac{1}{|\mathbf{r} - \mathbf{r}'|} i\delta(\tau) + g(|\mathbf{r} - \mathbf{r}'|)f(i\tau)$$
$$+ W_S(\mathbf{r}, \mathbf{r}', i\tau), \tag{5}$$

where $g(R)$ is the Yukawa function $[1 - \exp(-R/\lambda)]/R$ with the correct $1/R$ asymptotic dependence (with $\lambda = 1$ a.u.), and f is a function determined by the small-wave-

vector behavior of W in reciprocal space which allows the long-ranged part to be accommodated for each τ, leaving the exact remainder W_S short ranged. The factor of i in the bare interaction simply reflects our Fourier transform convention for imaginary time. The explicit separation of the bare interaction also makes W_S well behaved for small $|\mathbf{r} - \mathbf{r}'|$, where there can be no screening.

We calculate Σ using the GW approximation (1) directly in real space and imaginary time. With W split according to (5), Σ splits into a bare exchange part, a long-ranged term with multiplicative screening, and a short-ranged term [8]. The first two terms are easily evaluated (and, incidentally, account for about 90% of the self-energy). The main computational effort is accounted for by the remaining term.

To calculate quasiparticle energies we form the expectation values $\langle\psi_{n\mathbf{k}}|[\Sigma(i\tau) - V_{xc}]|\psi_{n\mathbf{k}}\rangle$ directly in real space (where V_{xc} is the LDA exchange-correlation potential). Alternatively, if the full self-energy operator is required, we take the 6DFFT to obtain $\Sigma_{\mathbf{G}\mathbf{G}'}(\mathbf{k}, i\tau)$. In either case we next take the Fourier transform from imaginary time to imaginary energy, and may then analytically continue to the real energy axis by first using optimization techniques to fit each element (separately for positive and negative energies) to the multipole form

$$a_0 + \sum_{i=1}^{n} \frac{a_i}{\omega - b_i}, \tag{6}$$

with complex parameters a_i and b_i, which (with $n = 2$) we find to be an extremely stable and accurate fit (rms relative error 0.2%) along the imaginary axis (and, for jellium, in good agreement with directly computed results along the whole of the real axis). The functional form is motivated by the known positions of the branch cuts and the character of the resonant structure in the self-energy. In principle it is necessary to constrain the pole positions b_i to avoid the quadrant of the ω plane through which the function is being continued, but in practice the optimal fit is always found to satisfy the constraint automatically. Sufficient numerical stability and information content are available to allow considerable extension of this functional form should this be desired, but in the present calculations it has not proved necessary.

If the Green's function is required, such as for a calculation of the momentum distribution or charge density, we instead retain Σ in an imaginary-energy representation $\Sigma_{\mathbf{G}\mathbf{G}'}(\mathbf{k}, i\omega)$ without the need for analytic continuation, and solve the matrix Dyson equation

$$G(\mathbf{k}, i\omega) = G_0(\mathbf{k}, i\omega) + G_0(\mathbf{k}, i\omega)[\Sigma(\mathbf{k}, i\omega) - V_{xc}]$$
$$\times G(\mathbf{k}, i\omega), \tag{7}$$

for G using matrix inversion. G may then be integrated as appropriate using contour deformation techniques in the ω plane, for example, to obtain the momentum distribution $n(\mathbf{k})$, or the electron density $n(\mathbf{r})$ (in which case a

6DFFT to real space is taken first). We also obtain the spectral function

$$A_{\mathbf{GG'}}(\mathbf{k}, \omega) = \frac{1}{\pi} |\mathrm{Im} G_{\mathbf{GG'}}(\mathbf{k}, \omega)| \qquad (8)$$

at real energies by analytic continuation from the imaginary-ω axis using the functional form (6), as used above for Σ.

We tested our method for jellium (where exact results at the level of GW are simple to calculate for comparison) and silicon. The jellium calculations show the method to yield the correct results (e.g., for quasiparticle energies, spectral function, and momentum density), and convergence with respect to the key FFT parameters (Δr, $\Delta \tau$, R_{max}, and τ_{max}) is very satisfactory. Typical values are $\Delta r = 0.5$ a.u., $\Delta \tau = 0.3$ a.u., $R_{max} = 18$ a.u., and $\tau_{max} = 10$ a.u., which give convergence of differences of self-energies (and therefore of quasiparticle energy differences) to better than 0.05 eV and absolute values to better than 0.1 eV. In Fig. 1 we give a comparison between the self-energy for real energies and for imaginary energies. As explained, the quantity for imaginary energy shows less structure and is much more convenient to handle numerically. Nevertheless all relevant information is retained, as shown by the fact that our analytic continuation techniques allow the full details of the self-energy for real energies [Fig. 1(b)] to be regained. The full results for jellium, which give further insight into the properties of the analytically continued quantities, will be presented in a future paper [9] together with details of the numerical methods.
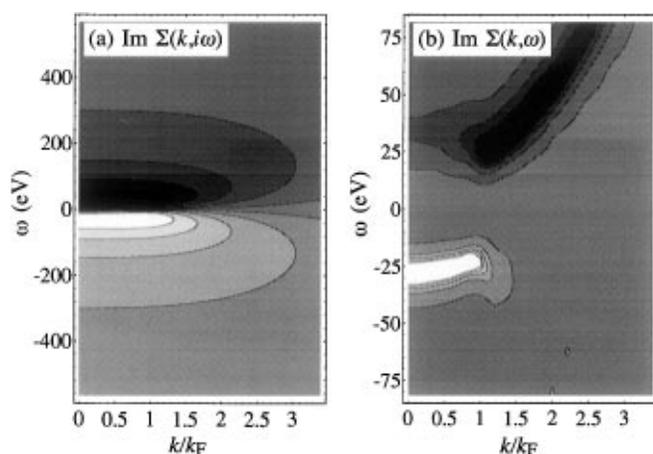
In the other figures we give results for silicon in the diamond structure ($a = 5.432$ Å), where the Green's function and screened Coulomb interaction are constructed from LDA wave functions and eigenvalues calculated using an *ab initio* norm-conserving pseudopotential. Tests show the above values of the FFT parameters to remain appropriate for silicon (the insensitivity of the FFT parameters to the material is discussed further below). Figure 2 shows the matrix element of the self-energy operator of silicon for real energies $\langle \psi_{n\mathbf{k}} | \Sigma(\omega) | \psi_{n\mathbf{k}} \rangle$ continued from the imaginary axis using (6). Plasmon damping at high and low energies is clearly visible. The inset illustrates the excellent quality of the fit (6) along the imaginary axis. In Fig. 3 we show the spectral function of silicon. The sharp peaks, which are automatically properly renormalized, correspond to the quasiparticle energies (and are in excellent agreement with both the quasiparticle energies calculated directly from the self-energy operator, with experiment, and with conventional GW calculations [1,2]) [10], while the remainder of the function corresponds to the well-known spectral background, and includes contributions from plasmons. We are not aware of any comparable calculations of the functions displayed for silicon. The calculations currently take 60 min per imaginary time on a Cray Y-MP for the full $W_{\mathbf{GG'}}(\mathbf{k})$, and 20 min per time for the full $\Sigma_{\mathbf{GG'}}(\mathbf{k})$. The calculation of the full self-energy $\Sigma_{\mathbf{GG'}}(\mathbf{k}, \omega)$ over an energy range of 100 eV takes less than half the time of a conventional technique using a plasmon-pole model for W, which is at best valid only for a much smaller energy range. Full exploitation of the short-ranged character of W_S would, we
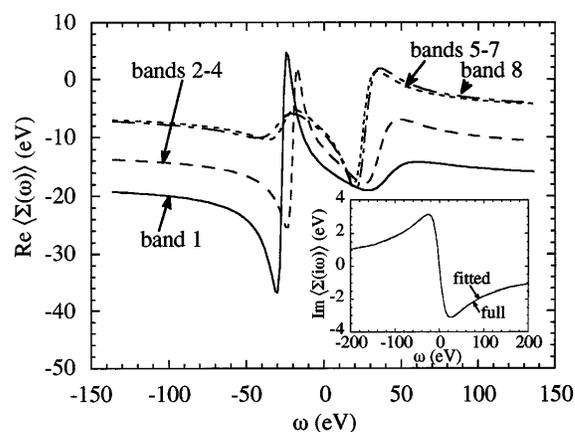

FIG. 1. (a) The GW self-energy of jellium (with density parameter $r_S = 2$) for imaginary energies, calculated using our method [and displayed here as its imaginary part, $\mathrm{Im}\Sigma(k, i\omega)$]. (b) The imaginary part of the same quantity analytically continued to the real-ω axis, $\mathrm{Im}\Sigma(k, \omega)$. The relative numerical simplicity of (a) is evident, but all relevant information is retained, as shown by (b), which is in excellent agreement with direct calculations. In each plot ten contours are used, from -3 (black) to 3 eV (white) in (a), and from -15 to 15 eV in (b).

FIG. 2. The matrix element of the self-energy operator of silicon continued as described onto real energy axis $\mathrm{Re}\langle \psi_{n\mathbf{k}} | \Sigma(\omega) | \psi_{n\mathbf{k}} \rangle$ shown for the first 8 bands at $\mathbf{k} = \mathbf{0}$. The deviation from a linear function is crucial in obtaining accurate quasiparticle energies away from the band gap (whose center is at zero). Inset: The calculated quantity along the imaginary energy axis $\mathrm{Re}\langle \psi_{n\mathbf{k}} | \Sigma(i\omega) | \psi_{n\mathbf{k}} \rangle$ for band 4 (the valence band maximum), together with the form (6) (with two poles) subsequently used to continue to the real axis; the fit is seen to be excellent.
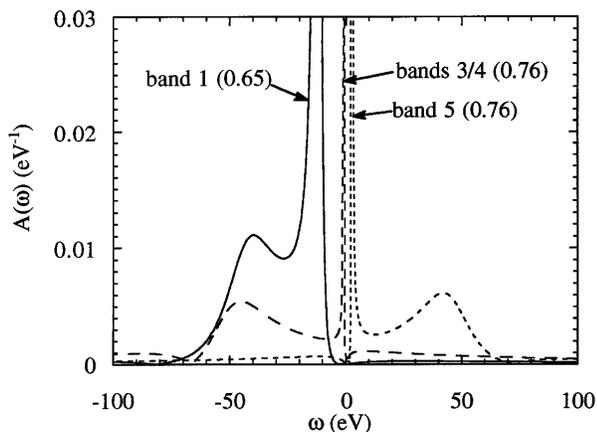
FIG. 3. The spectral function $\langle \psi_{n\mathbf{k}} | A(\mathbf{r}, \mathbf{r}', \omega) | \psi_{n\mathbf{k}} \rangle$ of silicon, for several bands at $\mathbf{k} = (1, 1, 1)\pi/4a$. The renormalized quasiparticle poles are evident as sharp peaks, whose calculated weights (the renormalization factors Z) are given. The spectral background, which includes contributions from plasmons, is also visible.

estimate, yield a further saving of at least 50%, and scaling to larger systems is very favorable as discussed below.

The new method opens new prospects for the applications of *ab initio* many-body perturbation theory. First, the improved computational efficiency of the various parts of the calculation, together with the elimination of the need for plasmon-pole models for W, allow larger and more complex unit cells to be studied. Second, the existence of an efficient way of solving the Dyson equation to obtain the next-level Green's function G [Eq. (7)] raises the possibility of going beyond the GW approximation to treat more strongly correlated electrons by including the vertex function Γ (which is approximated as a delta function at the GW level). One simple way in which this may be done has been mentioned above [5,6]; we are exploring the possibility of a more general Γ in which our real-space grid will be used to exploit its expected short-ranged character.

For large unit cells, the dependence of the total computer time on the number of basis functions N becomes crucial. The most important point is that the range of the nonlocality of W or Σ appears to be approximately the same in all materials [11], and the length scale of the short-range behavior is set by atomiclike quantities, so that R_{max} and Δr need not change with N. Similarly the imaginary-time grid used for silicon is capable of representing much structure, so that $\Delta \tau$ and τ_{max} do not change significantly with N. The information stored is therefore *linear* in N, since the number of $\mathbf{r}$ points (which run over the irreducible wedge of the unit cell, as opposed to the $\mathbf{r}'$ points which run over the sphere of radius R_{max}) is proportional to N. Consequently the time for all the calculations performed using this representation, which are dominated by FFT's, will scale with system size approximately merely as N [12]. If the basis set is changed for a given system size, the time scales as N^2. These scalings

are in marked contrast to conventional techniques, which usually involve N^2 double summations over bands and k points and typically scale as N^4.

In conclusion, we presented a new formulation of *ab initio* computational many-body perturbation theory for solids which gives greatly improved efficiency in the calculation of dielectric matrices, self-energy operators, Green's functions, quasiparticle energies, spectral functions, charge densities, etc. The computational cost of this method scales approximately linearly with the system size, which allows the routine extension of *ab initio* work beyond calculations of quasiparticle energies, and its application to materials requiring larger basis sets or larger unit cells than were previously feasible.

---

[*]Permanent address: Carrera de Fisica, Fac. Sc. y Technologia, Universidad Mayor de San Simon, Cochabamba, Bolivia.

[1] M. S. Hybertsen and S. G. Louie, Phys. Rev. Lett. **55**, 1418 (1985); **58**, 1551 (1987); Phys. Rev. B **34**, 5390 (1986); **38**, 4033 (1988).

[2] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. Lett. **56**, 2415 (1986); Phys. Rev. B **37**, 10 159 (1988); J. P. A. Charlesworth, R. W. Godby, and R. J. Needs, Phys. Rev. Lett. **70**, 1685 (1993).

[3] L. Hedin, Phys. Rev. **139**, A796 (1965).

[4] The quantities are not periodic in the six-dimensional vector $(\mathbf{r}, \mathbf{r}')$, so the transform is not simply a FFT, but the FFT forms the main computational component.

[5] Footnote in the third paper of Ref. [1].

[6] R. Del Sole, L. Reining, and R. W. Godby, Phys. Rev. B **49**, 8024 (1994).

[7] R. Dahling, W. van Haeringen, and B. Farid, Phys. Rev. B **44**, 2952 (1991); N. E. Maddocks, R. W. Godby, and R. J. Needs, Phys. Rev. B **49**, 8502 (1994).

[8] We follow Hedin [3] in introducing a small degree of self-consistency in the Green's function by shifting the Green's function in energy before inserting it into the GW expression.

[9] H. N. Rojas and R. W. Godby (to be published).

[10] For example, our calculated minimum band gap of silicon is 1.29 eV, and the direct gap at Γ is 3.29 eV.

[11] A minor exception is a very large unit cell, larger than the normal R_{max}, in which R_{max} may need to be increased somewhat to allow a correct description of nonlocal screening between all points within the unit cell for certain systems.

[12] The matrix inversions and multiplications, which we perform in the reciprocal-space representation, scale as N^3, but account for a very small fraction of the total time (less than 0.5% in silicon). For extremely large unit cells it would be advantageous to perform the matrix operations in real space.