



The GW approximation: content, successes and limitations

Lucia Reining ^{1,2*}

Many observables such as the density, total energy, or electric current, can be expressed explicitly in terms of the one-body Green's function, which describes electron addition or removal to or from a system. An efficient way to determine such a Green's function is to introduce a self-energy, which is a nonlocal and dynamic effective potential that influences the propagation of particles in an interacting system. The state-of-the-art approximation for the self-energy is the GW approximation, where the system to (or from) which the electron is added (or removed) is described as a polarizable, screening, medium. This is expressed by the name of the approximation: 'GW' stands for the one-body Green's function G and for W , the dynamically screened Coulomb interaction. The GW approximation is very popular for the calculation of band structures in solids, and increasingly used also to describe nanostructures, clusters, and molecules. As compared to static mean-field approximations for the effective potential, the dynamical screening of the Coulomb interaction in GW leads to a renormalization of energies, to broadening and/or to the observation of additional excitations. An analysis of the approximations that lead to the GW self-energy, and of the underlying picture, explains the successes and the limitations of the approach. © 2017 Wiley Periodicals, Inc.

How to cite this article:

WIREs Comput Mol Sci 2018, 8:e1344. doi: 10.1002/wcms.1344

INTRODUCTION

At the quantum level, all properties of materials can be obtained from expectation values of operators. At zero temperature, the value O of an observable is the expectation value of an operator \hat{O} evaluated with the ground-state wavefunction Ψ of the Hamiltonian that describes the system:

$$O = \langle \Psi | \hat{O} | \Psi \rangle \quad (1)$$

In this review, we concentrate on electronic properties of atoms, molecules, nanostructures, or extended materials. We specify the expressions for zero temperature

and a static external potential; however, the discussion is readily generalized, see e.g.,¹ If we neglect the kinetic energy of the nuclei, which are much heavier than the electrons, and work with fixed nuclear positions in the Born–Oppenheimer approximation,² the nonrelativistic Hamiltonian for the electrons reads^a

$$\hat{H} = -\frac{1}{2} \sum_i \nabla_i^2 + \frac{1}{2} \sum_{i \neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,l} \frac{Z_l}{|\mathbf{r}_i - \mathbf{R}_l|} \quad (2)$$

it consists of the operators for the kinetic energy of electrons (indicated by lower case subscripts), their Coulomb interaction energy, and the electron–nuclei interaction, where Z_l is the atomic number and \mathbf{R}_l the position of the nuclei. Other contributions, such as spin–orbit coupling, are neglected here for simplicity, but can also be included in the framework that we will discuss in this review article.

If we could neglect the second term, which contains the two-body Coulomb interaction, the

*Correspondence to: lucia.reining@polytechnique.fr

¹Laboratoire des Solides Irradiés, École Polytechnique, CNRS, CEA, Université Paris-Saclay, Palaiseau, France

²European Theoretical Spectroscopy Facility (ETSF), <http://www.etsf.eu>

Conflict of interest: The author has declared no conflicts of interest for this article.

wavefunction of the many-electron system would simply be a Slater determinant, a product wavefunction that is anti-symmetrized in order to account for the fermionic nature of the electrons, so electrons would be uncorrelated. However, the Coulomb interaction is in general not negligible with respect to the kinetic energy. Therefore, the many-body ground-state wavefunction is a superposition of many Slater determinants. This expresses the fact that electrons are correlated. It is hard to calculate the many-body wavefunction, and therefore to access properties of systems with more than very few electrons, by evaluating expectation values according to Eq. (1). Moreover, the wavefunction is large, often too large to be stored,³ and it contains too much information, much more than what one normally asks for: usually, we are interested only in a few observables.

Can one calculate expectation values *without* knowing the many-body wavefunction? Density functional theory (DFT),³ a today widely used approach in quantum chemistry, tells us that the answer is yes, in principle: the Hohenberg–Kohn theorems state that given the ground-state density, the external potential, and therefore the many-body Hamiltonian, is determined up to a constant. This fixes the many-body ground-state wavefunction, and finally all ground-state expectation values. In other words, all observables are functionals of the density. Note that the spin-resolved density itself is an observable given by the expectation value

$$n(x) = \langle \Psi | \hat{\psi}^\dagger(x, t) \hat{\psi}(x, t) | \Psi \rangle \quad (3)$$

Here, $\hat{\psi}$ are field operators in the Heisenberg picture; $\hat{\psi}^\dagger(x, t)$ ($\hat{\psi}(x, t)$) creates (annihilates) an electron at time t and space-spin-coordinate^b $x \equiv (\mathbf{r}, \sigma)$. Because we deal here with time-independent external potentials, the field operators $\hat{\psi}(x, t)$ in the Heisenberg picture are linked to the field operators $\hat{\psi}(x)$ in the Schrödinger picture by.

$$\hat{\psi}(x, t) = e^{it\hat{H}} \hat{\psi}(x) e^{-it\hat{H}} \quad (4)$$

and as expected the equilibrium density does not depend on the time t .

The theorems are not immediately useful in practice, as the explicit expressions of almost all observables as functionals of the density are unknown. One problem is the kinetic energy $E_{\text{kin}} = -(1/2) \sum_{i=1}^N \int d\mathbf{x}_1 \dots d\mathbf{x}_N \Psi^*(x_1, \dots, x_N) \nabla_i^2 \Psi(x_1, \dots, x_N)$: it is an explicit expression in terms of the many-body wavefunction, but how could one write this in terms of the density? It would be much easier

to express E_{kin} in terms of the (spin-resolved) one-body density *matrix*, which reads

$$\rho(x, x') \equiv N \int d\mathbf{x}_2 \dots d\mathbf{x}_N \Psi^*(x, x_2, x_3, \dots, x_N) \times \Psi(x', x_2, x_3, \dots, x_N) = \langle \Psi | \hat{\psi}^\dagger(x, t) \hat{\psi}(x', t) | \Psi \rangle \quad (5)$$

In that case, the kinetic energy is simply $E_{\text{kin}} = -(1/2) \int d\mathbf{x} [\nabla_{\mathbf{r}}^2 \rho(x, x')]_{\sigma' = \sigma, \mathbf{r}' = \mathbf{r}}$.

By using the nonlocal density matrix instead of the density, one also has an explicit expression for the exact nonlocal exchange potential

$$\Sigma_{\text{x}}(x, x') = -v_{\text{c}}(\mathbf{r} - \mathbf{r}') \rho(x, x') \quad (6)$$

which appears, for example, in the Hartree–Fock (HF) approximation. Here, v_{c} is the Coulomb interaction

$$v_{\text{c}}(\mathbf{r} - \mathbf{r}') \equiv \frac{1}{|\mathbf{r} - \mathbf{r}'|} \quad (7)$$

For other observables, in particular spectra, explicit expressions are not known, be it in terms of the density or the density matrix. Spectra are frequency resolved observables, which means that *time* differences appear. This information is instead encoded in *Green's functions*.

The time-ordered equilibrium one-body Green's function at zero temperature is defined as⁴

$$G(x, t; x', t') = -i \langle \Psi | T [\hat{\psi}(x, t) \hat{\psi}^\dagger(x', t')] | \Psi \rangle \quad (8)$$

where T denotes the time-ordering of operators defined in Eq. (65). The Green's function is a natural generalization of the density and of the one-body density matrix: starting from the density in Eq. (3), it adds space-spin nonlocality ($\rho(x, x')$ in Eq. (5)) and nonlocality in time (Eq. (8)). It is the probability amplitude for inserting an electron in (x', t') and finding an extra electron at (x, t) . Similarly, it is the probability amplitude to find an extra hole (i.e., a missing electron) at (x', t') when an electron is removed in (x, t) . In other words, it describes the *propagation* of electrons and holes. In equilibrium, G depends only on the time difference $(t - t')$, and its Fourier transform $G(x, x'; \omega)$, on one frequency. As we will see later, many observables can be expressed explicitly in terms of G .

Similarly to DFT, where it is important to determine the ground-state density with good precision, the question is now how to obtain the Green's function. DFT is often solved in the Kohn–Sham (KS) approach⁵: one builds an auxiliary system that is

noninteracting, but that has an effective potential consisting of the original external potential $v_{\text{ext}}(\mathbf{r})$, and of the Hartree contribution $v_{\text{H}}[n](\mathbf{r})$ plus a local exchange–correlation potential $v_{\text{xc}}[n](\mathbf{r})$, which are themselves functional of the density. The self-consistent solution of this independent-particle problem yields in principle the exact density, and in practice good approximations to it if a good approximation for v_{xc} is used. Similarly, one can say that the Green's function is determined by an auxiliary system with an effective 'potential.' In addition to $v_{\text{ext}} + v_{\text{H}}$, it consists of an exchange–correlation term $\Sigma_{\text{xc}}(x, t; x', t')$. This exchange–correlation *self-energy* is *nonlocal in space, spin, and time*. It is a functional of the Green's function, and it is known as a perturbative expansion in terms of the Coulomb interaction.⁴ However, there is no closed expression for it.

The lowest order of the perturbation expansion is the Fock term Σ_{x} of Eq. (6); it corresponds to an instantaneous self-energy, which can also be written as

$$\Sigma_{\text{x}}(x, t; x', t') = i\delta(t^+ - t')v_{\text{c}}(\mathbf{r} - \mathbf{r}')G(x, t, x', t') \quad (9)$$

where $t'^+ \equiv \lim_{\eta \rightarrow 0^+} t' + \eta$ as defined in Eq. (59) guarantees the fact that t' tends to t from above. However, HF eigenvalues correspond to total energy differences for a system that remains frozen when an electron is added or removed: these eigenvalues contain no information about the fact that the system would react to the addition of a charge. One can express this by saying that the extra charge is not *screened*.

Screening is a key concept in condensed matter physics and chemistry. A summary of important aspects can be found in Box 1. While screening is completely neglected in HF, it is instead included in the GW approximation (Box 2): in simple words, the GWA is an approximation to the exchange–correlation self-energy, which contains the bare Fock exchange, and some correlation in the form of screening. While HF is the lowest order self-energy in terms of the bare Coulomb interaction v_{c} , the GW self-energy is the lowest order in terms of the *screened* Coulomb interaction W , which is the reason for its name 'GW.'

This review article gives an overview of the essence and effects of screening in electronic spectra,

BOX 1

Screening is one of the key concepts used in this article. It refers mostly to linear response: an external perturbation $u(\mathbf{r}', t')$ in a place \mathbf{r}' at a time t' changes the charge density of the

system at time t and position \mathbf{r} by an amount n_{ind} , the induced charge

$$n_{\text{ind}}(\mathbf{r}, t) = \int d\mathbf{r}' dt' \chi(\mathbf{r}, t; \mathbf{r}', t') u(\mathbf{r}', t').$$

Here, χ is the density–density response function that does not depend on the external perturbation. In equilibrium it depends on the time difference $t' - t$ and its Fourier transform, on one frequency ω . Its poles are the excitation energies of the system for neutral (i.e. that do not change particle number N) excitations λ from the ground state, $E_{N,\lambda} - E_N$. Such excitations correspond to the creation of electron–hole pairs. From the response function, one obtains the inverse test-charge–test-charge dielectric function

$$\epsilon^{-1}(\mathbf{r}, t; \mathbf{r}', t') = \delta(\mathbf{r} - \mathbf{r}')\delta(t - t') + \int d\mathbf{r}'' v_{\text{c}}(\mathbf{r} - \mathbf{r}'')\chi(\mathbf{r}, t; \mathbf{r}'', t')$$

which gives the relation between an external perturbation u and the resulting classical (external plus induced Hartree) potential u_{res} as

$$u_{\text{res}}(\mathbf{r}, t) = \int d\mathbf{r}' dt' \epsilon^{-1}(\mathbf{r}, t; \mathbf{r}', t') u(\mathbf{r}', t')$$

The inverse dielectric function screens the Coulomb interaction: the frequency Fourier transform of the screened Coulomb interaction W is

$$W(\mathbf{r}, \mathbf{r}'; \omega) = \int d\mathbf{r}'' \epsilon^{-1}(\mathbf{r}, \mathbf{r}''; \omega) v_{\text{c}}(\mathbf{r}'' - \mathbf{r}')$$

Because $\chi(\omega)$ displays the excitations of the system, so does $W(\omega)$. The spectrum of excitations is directly seen in its imaginary part. $W(\omega = 0)$ is in general (though not always) weaker than the bare Coulomb interaction v_{c} . It contains the information about the adiabatic relaxation of the system that one also finds in Δ -SCF (see *Screening in Finite and Infinite Systems*). The frequency dependence is sometimes highlighted by speaking about *dynamical screening*.

BOX 2

The acronym 'GW' stands for an approximation to the self-energy, a nonlocal and frequency-dependent effective potential for the propagation of electrons in an interacting system. It contains the exact, nonlocal, Fock exchange, and correlation in the form of dynamical screening. Koopmans' theorem tells us that the energy eigenvalues of the HF Hamiltonian correspond

to electron addition and removal, but with all other system electrons frozen. Instead, in the GW approximation the system *responds* to the perturbation induced by electron addition or removal, through electronic excitations. The name 'GW' expresses this physics: 'G' is the one-body Green's function, which describes the propagation of a particle in an interacting system, and 'W' is the linear response dynamically screened Coulomb interaction. In the GWA, the self-energy is a product of G and W. This means that the additional particle (electron or hole) does not 'see' individual system particles, but a charge density, and only the linear response of this charge density is considered. It is most often calculated in the random-phase approximation, but sometimes other approaches, for example, time-dependent density functional theory, are used to determine W.

One may illustrate this situation using the picture of an object (the electron or hole) moving on water (the system). In HF, the water is frozen and simply yields a fixed potential: this corresponds to a person skating on ice (left panel). In the GWA, the water reacts by creating waves (these could, e.g., be plasmon excitations of the system), which in turn act back on the propagation of the object, e.g., a boat (right panel). The approximation consists in the fact that the propagating object does not meet individual water molecules, but a continuous fluid. Usually additional approximations are made concerning the response of the 'water' (e.g., the RPA for W).

how GW can be derived and how it treats this phenomenon, what can be achieved, and what are the limitations of this approach.

FROM WAVEFUNCTIONS TO GREEN'S FUNCTIONS

Before delving into the GW approximation for the self-energy, let us have a closer look to the problem and the concepts introduced above.

Observables

Our aim is to calculate observables, defined as the expectation value Eq. (1). This equation shows that all observables are simple functionals of the many-body ground-state wavefunction—but our aim is to avoid calculating this wavefunction,

which is huge, and which contains much more information than the few observables we are usually interested in. As outlined in the introduction, one could in principle calculate observables as functionals of the density $n(\mathbf{r})$, which is a much simpler quantity than the many-body wavefunction. The price to pay is that the expressions for observables as functional of n are exceedingly more complicated, they can be nonanalytic, and they are in general unknown. In order to have a practical scheme, many electronic structure calculations for realistic systems work in the framework of DFT and the KS approach.⁵ In KS, the density is derived from a fictitious noninteracting system with an effective Hartree plus exchange–correlation potential $v_{\text{H}}[n](\mathbf{r}) + v_{\text{xc}}[n](\mathbf{r})$ that is itself a functional of the density and that is added to the external potential v_{ext} . Therefore, the effective KS electrons are governed by the KS equations, which are single-particle Schrödinger equations reading

$$\left\{ -\frac{\nabla^2}{2} + v_{\text{ext}}(\mathbf{r}) + v_{\text{H}}(\mathbf{r}) + v_{\text{xc}}(\mathbf{r}) \right\} \varphi_i(\mathbf{r}) = \varepsilon_i \varphi_i(\mathbf{r}) \quad (10)$$

where ε_i and $\varphi_i(\mathbf{r})$ are KS eigenvalues and orbitals, respectively. Usually, one makes two *approximations*:

1. the exchange–correlation potential $v_{\text{xc}}[n](\mathbf{r})$ is approximated, for example, in the local density approximation (LDA).⁵ This yields approximate independent-particle KS orbitals and energies, and an approximate density;
2. because we do not know how to express observables as functionals of the density, they are calculated using Eq. (1), but the many-body wavefunction is approximated by the KS ground-state Slater determinant. In this approximation, for example, the band structure is simply given by KS eigenvalues. Depending on the observable and the system, replacing the unknown functional of the density by an expectation value calculated with KS Slater determinants can be a very rough approximation, although there are arguments that suggest physical meaning for the use of KS eigenvalues (see, e.g.,^{6,7}).

Both issues impact the result. A typical example is the band structure of solids: usually the fundamental band gap (the equivalent of the gap between the highest occupied molecular orbital [HOMO] and the lowest unoccupied molecular orbital [LUMO] in molecules) calculated in KS-DFT–LDA is much too

small compared to experiment. This is illustrated in Figure 1 for the case of bulk silicon: the occupied valence bands are quite well described, but there is an essentially constant relative shift of the unoccupied bands, such that the band gap is smaller than the measured one by about 50%. It has been discussed⁹ that both the use of an approximation for v_{xc} and the fact that electron addition or removal energies are replaced with KS eigenvalues^c contribute to this discrepancy. In the following, we concentrate on the second issue.

Why is it so complicated to express observables in terms of the density, instead of the many-body wavefunction?

Let us look at spectra. For example in the case of electron removal, one wishes to calculate the energies and intensities given by matrix elements of the spectral function

$$A_{ii}(\omega) = \sum_{\lambda} m_i^{\lambda} \delta(\omega - E_{\lambda}) = \frac{1}{2\pi} \left| \int dx dx' \varphi_i^*(x) \varphi_i(x') \right|$$

$$\text{Im} \int_0^{\infty} d\tau e^{i\omega\tau} \langle \Psi | \hat{\psi}^{\dagger}(x, \tau) \hat{\psi}(x', 0) | \Psi \rangle \quad (11)$$

This quantity has peaks at the electron removal energies of the many-body system $E_{\lambda} = E_N - E_{N-1, \lambda}$, where E_N is the N -electron ground-state energy, and

$E_{N-1, \lambda}$ is the energy of the $N - 1$ -electron system in a state λ . The spectrum is weighted by matrix elements m_i^{λ} that are electron removal probability amplitudes projected onto some basis φ_i .

This exact expression for the spectral function is given in terms of an expectation value of two field operators in the Heisenberg picture, which appear at two different points in space-spin and at two different times. However, it is very difficult to link the events at different places and times by working with the instantaneous density or density matrix. It is much simpler to adapt to this situation, and to keep the nonlocality in space and time explicitly in the formulation. This is achieved by working, instead of the density, with a nonlocal object: the one-body Green's function.

Green's Functions

The one-body Green's function G is defined in Eq. (8). Its time Fourier transform (Eq. (63)) allows one to express the spectral function (Eq. (11)) as

$$A_{ii}(\omega) = \frac{1}{2\pi} \left| \text{Im} \int dx dx' \varphi_i^*(x) \varphi_i(x') G(x, x'; \omega) \right| \quad (12)$$

Moreover, all expectation values of a one-body operator are obtained in a simple way:

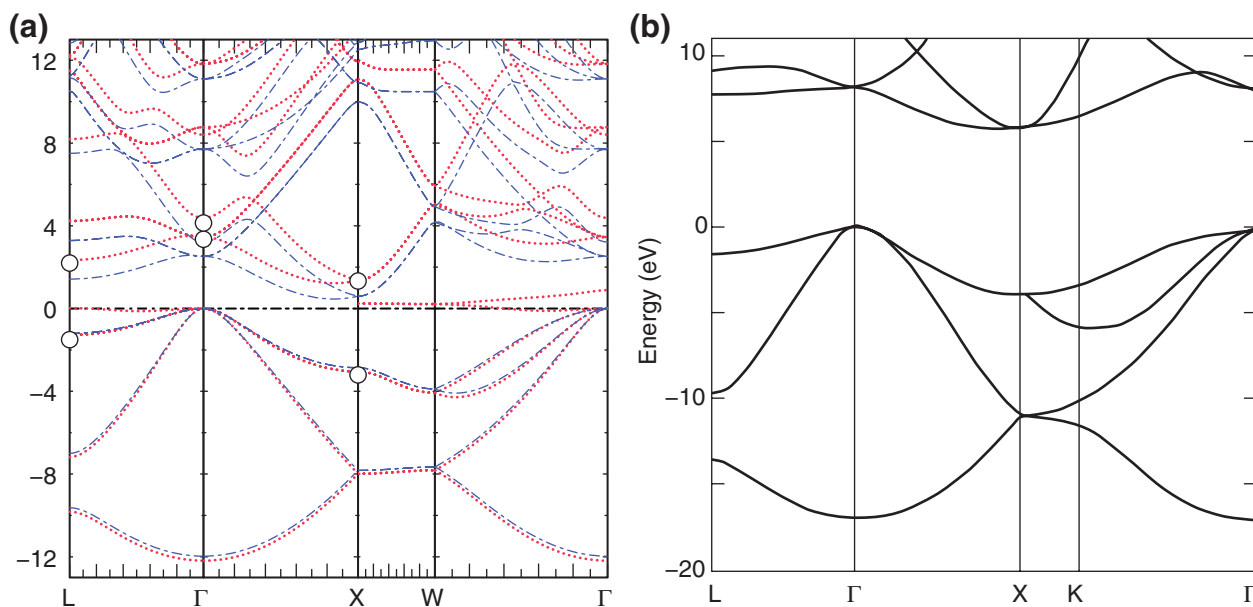


FIGURE 1 | Band structure of bulk silicon. Left panel: Courtesy of Mark van Schilfgaarde. Blue dot-dashed lines are local density approximation results, red dotted lines are results of a GW calculation using the quasi-particle self-consistent GW approximation (see *The GW Approximation in Practice*). Circles are a collection of experimental data. Right panel: Hartree-Fock band structure (Reproduced with permission from Ref 8 Copyright 1985 IOP Publishing).

$$\langle \Psi | \hat{O} | \Psi \rangle = -i \int dx_1 dx_2 O(x_1, x_2) G(x_1, t; x_2, t^+) \quad (13)$$

because in the Heisenberg representation $\hat{O}_1(t) = \int dx_1 dx_2 O(x_1, x_2) \hat{\psi}^\dagger(x_1, t) \hat{\psi}(x_2, t)$.

Many important observables are expectation values of one-body operators and can therefore be expressed explicitly in terms of the one-body Green's function. If one is interested in observables related to two-body operators, expectation values of four field operators have to be calculated, and one is back to the problem. However, the probably most important two-body expectation value, the two-body interaction energy, *can* be expressed in terms of G . Therefore the total energy is given by the Galitskii–Migdal expression¹⁰

$$E = \frac{1}{2} \int dx \lim_{t' \rightarrow t^+} \lim_{x' \rightarrow x} \left[\frac{\partial}{\partial t} - i h(\mathbf{r}) \right] G(x, t, x', t') \quad (14)$$

where $h(\mathbf{r})$ is the noninteracting part of the Hamiltonian.

In view of this, working with functionals of the one-body Green's function is an excellent compromise: the object itself is nonlocal in space-spin and time and therefore more complicated than the density, but still much simpler than the $3N$ -dimensional many-body wavefunction. At the same time, it allows us to access directly and in an principle exact way many more observables than what is possible in practice with density functionals.

In this review, we will focus on aspects of Green's functions and many-body perturbation theory that are inspired by their origin in the electron gas and extended systems. Green's functions are widespread also in quantum chemistry,^{11,12} for example, to deal with problems of open shell systems and strong correlation (see, e.g., Refs 13–16), or to analyze total energies (see, e.g., Ref 17), but this would be a topic that is complementary to the present discussion (for a recent review, see, e.g., Ref 18).

Spectral Functions

Let us concentrate for a moment on the electron removal spectral function given by Eq. (11). As explained above, it exhibits peaks at total energy differences, between the system in its N -electron ground state and in some excited state of the $N - 1$ -electron system. An analogous expression describes electron addition, with energies $E_{N+1,\lambda} - E_N$. The first removal peak is therefore at the ionization energy, and the first addition peak gives the electron affinity.

For more than one electron there are more total energy differences than independent-particle energies, because there are more Slater determinants $|S_\alpha\rangle$ than single-particle orbitals. Matrix elements $\langle S_\alpha | \hat{c}_i | S_{\alpha'} \rangle$ of the electron removal operator \hat{c}_i from a single-particle orbital φ_i are zero unless $|S_{\alpha'}\rangle = \hat{c}_i^\dagger |S_\alpha\rangle$. Therefore, in a noninteracting system most of the possible excitations cannot be seen in the spectral function. In an interacting system, instead, matrix elements are taken between many-body states that are superpositions of several Slater determinants. For this reason, the spectral function of an interacting system displays more peaks than that of an independent-particle system with the same number of electrons; in particular, one can detect multiple excitations in the interacting spectral function. More precisely, when the matrix elements $\int dx \varphi_k^*(x) A(x, x'; \omega) \varphi_k(x)$ of the noninteracting spectral function are calculated with the same orbitals φ_k that form the Slater determinants, they are diagonal, and they consist of single peaks given by δ -functions. In the interacting case, each such matrix element has several peaks, and the number of peaks increases with the number of electrons. In extended systems such as solids the energy spectrum of each matrix element becomes continuous. An example for a typical matrix element of a spectral function in an extended system is shown in Figure 2.

Usually one can still identify a main peak in the interacting spectrum that is continuously linked to an independent-particle peak when the interaction is turned off: this is called a quasi-particle. It is the dominant peak in the example of Figure 2. With respect to the independent-particle peak (vertical line indicating a sharp δ -function peak) it is usually shifted and it is broadened, because the quasi-particle has a finite lifetime and can decay into many close lying excitations.^d The position of the quasi-particle peaks as a function of the momentum quantum number yields the band structure in solids, like the one shown in Figure 1. Moreover, the interacting spectral function can exhibit other structures, called satellites or side-bands in extended systems. Similar to the multiple excitations in finite systems, they highlight the fact that excitations in an interacting system are coupled. For example, electron removal in a metal can lead to the excitation of plasmons, long-range charge oscillations of the electron gas.^e Because these oscillations carry energy, the removal satellites are typically further away from the Fermi level than the quasi-particle peaks (and similarly for the electron addition spectrum, which is found at positive energies). The spectral function is a reasonable quantity to describe and interpret an

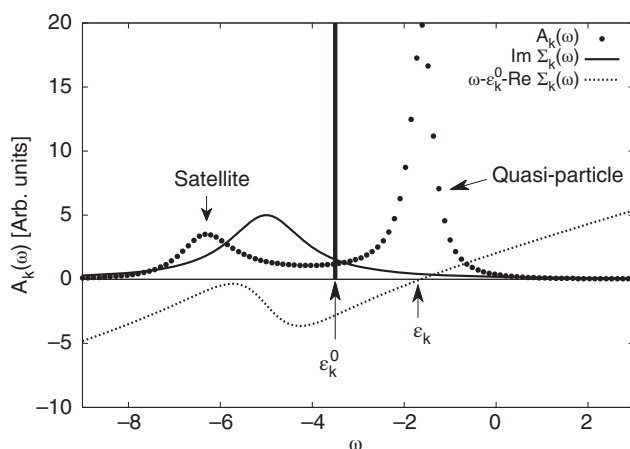


FIGURE 2 | Typical electron removal spectral function matrix element in an extended system. The vertical line indicates an independent-particle result such as HF: it consists of one sharp peak. The interacting spectral function, given by the fat dots, exhibits a broad quasi-particle peak and a satellite due to excitations of the many-body system. Shown are also the imaginary (continuous line) and shifted real (dotted line) parts of the self-energy. The quasi-particle peak appears where the shifted real part crosses zero. Relative shifts would be different if another independent-particle system was chosen.

experimental photo emission or inverse photo emission spectrum.

The above discussion tells us how to obtain observables from the Green's function, but we do not know yet how to calculate the Green's function itself, unless we go back to its definition, Eq. (8). In the absence of interaction, when the ground state $|\Psi\rangle$ is a single Slater determinant, Eq. (8) can be evaluated immediately, yielding simply

$$G_0(x_1, t_1; x_2, t_2) = -i \sum_i \varphi_i(x_1) \varphi_i^*(x_2) e^{-i\epsilon_i^0(t_1 - t_2)} [\Theta(t_1 - t_2) \Theta(\epsilon_i^0 - \mu) - \Theta(t_2 - t_1) \Theta(\mu - \epsilon_i^0)] \quad (15)$$

where the single-particle eigenvalues ϵ_i^0 and eigenfunctions $\varphi_i(x)$ appear, and $\Theta(\tau)$ is the Heaviside step function, $\Theta(\tau) = 0$ for $\tau < 0$ and $\Theta(\tau) = 1$ for $\tau > 0$. There are separate contributions for occupied and empty states (eigenvalues below and above the chemical potential μ). The frequency Fourier transform exhibits poles at the single-particle energies:

$$G_0(x_1, x_2, \omega) = \lim_{\eta \rightarrow 0^+} \sum_i \frac{\varphi_i(x_1) \varphi_i^*(x_2)}{\omega - \epsilon_i^0 + i\eta \operatorname{sgn}(\epsilon_i^0 - \mu)} \quad (16)$$

where sgn is the sign function. Using Eq. (11) the noninteracting spectral function in the basis of the single-particle orbitals becomes

$$A_{kk}^0(\omega) = \delta(\omega - \epsilon_k^0) : \quad (17)$$

this corresponds to the vertical bar in Figure 2.

At this point, we have motivated the use of Green's functions: the one-body Green's function G gives direct access to expectation values of one-body operators, to the total energy and to electron addition and removal spectra. The spectral function of electron addition and removal is proportional to the imaginary part of $G(\omega)$. In a noninteracting system its matrix elements consist of sharp peaks at the single-particle eigenvalues. In an interacting system, additional peaks appear that merge to form a continuous spectrum in infinite systems, exhibiting broadened quasi-particle peaks and satellite structures. We now have to find a way to calculate the Green's function for the general case of an interacting system.

The Self-Energy

For the interacting many-body system, we cannot get the Green's function from Eq. (8), as we do not know the many-body state $|\Psi\rangle$. This rises the question '*how can we capture in an efficient way the effect of the Coulomb interaction, which modifies the propagation of a particle in a material (described by G) with respect to the propagation of an independent particle (described by G_0)?*' Let us try to cast this effect into the form of an effective potential. Some contributions are easy to guess. First, there is the effect of the classical electrostatic potential due to all electrons: this is the Hartree potential

$$v_H(\mathbf{r}) = \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (18)$$

Second, because electrons are fermions we must have an exchange term, the Fock contribution Σ_x of Eq. (6).

Let us pause a second at the level of this well-known HF approximation. Most often one would evaluate the solution by solving the HF equations, a set of effective independent-particle Schrödinger equations. They are similar to the KS Eq. (10). The Hartree potential appears in the same way in both HF and KS equations, but HF replaces the multiplicative action of the local KS potential^f $v_{xc}(x)\varphi_i(x)$ by the nonlocal $\int dx' \Sigma_x(x, x')\varphi_i(x')$. As the HF particles are effective independent particles, one can construct the Green's function G_{HF} from the eigenfunctions and eigenvalues of the HF equations using Eq. (15).

Alternatively, one can directly write the link between G_{HF} and the noninteracting G_0 as

$$G_{\text{HF}}(x, x', \omega) = G_0(x, x', \omega) + \int dx_1 dx_2 G_0(x, x_1, \omega) [\nu_{\text{H}} + \Sigma_{\text{x}}](x_1, x_2) G_{\text{HF}}(x_2, x', \omega) \quad (19)$$

This link between an interacting and a noninteracting Green's function is a *Dyson equation*, and the *kernel* (which is here $\nu_{\text{H}}(\mathbf{r})\delta(x_1 - x_2) + \Sigma_{\text{x}}(x_1, x_2)$) that contains all the interaction effects is called *self-energy*. Note that also KS-DFT can be expressed as a Dyson equation

$$G_{\text{KS}}(x, x', \omega) = G_0(x, x', \omega) + \int dx_1 G_0(x, x_1, \omega) [\nu_{\text{H}}(x_1) + \nu_{\text{xc}}(x_1)] G_{\text{KS}}(x_1, x', \omega) \quad (20)$$

where G_{KS} is the independent-particle Green's function built with KS eigenfunctions and eigenvalues according to Eq. (15).

The HF approximation is a useful and physically meaningful starting point, as it contains the important electrostatic (Hartree) effects, and the information that electrons are indistinguishable. Note that today the Hartree potential is commonly defined as the electrostatic potential due to *all* electrons, including the electron itself on which it acts: it carries a self-interaction problem.¹⁹ The Fock term exactly cancels the self-interaction contribution: contrary to the 'Hartree'-approximation as it is defined above, HF yields the correct total energy for a single electron. However, it completely neglects correlation effects,⁸ which implies that in the HF equations the electrons do not react to the propagation of an electron or hole that is added to the system: all the other electrons are frozen. Therefore, HF eigenvalues correspond to too large energy differences for electron addition or removal. In particular in metals and semiconductors, the consequences are striking. The right panel in Figure 1 shows the band structure of HF eigenvalues for bulk silicon, taken from Ref 8. In comparison to the KS-LDA result in the left panel, the bands are wider and, most importantly, the band gap is drastically increased. It is much larger than the measured band gap that is given by the experimental results in the left panel: the sign of the error has changed with respect to KS-LDA, but its magnitude is rather increased.

The reaction of the electrons as a response to a perturbation—here given by an extra charge—is

summarized in the concept of *screening*, which is further detailed in Box 1. We can say that the fact that HF misses screening spoils its eigenvalue spectrum, and it is in particular detrimental for the band structure of solids.

The Dyson Eq. (19) suggests a direction for further improvement: because all interaction effects are contained in the self-energy, the strategy is to design a better self-energy, rather than to work on the Green's function itself. This is analogous to KS-DFT, where a major effort goes into the design of improved exchange–correlation potentials.

In order to illustrate the power of Dyson equations, let us take Eq. (19) and suppose for simplicity that the HF and noninteracting Green's functions are diagonal in the same basis. This means that HF and noninteracting orbitals are the same, and that HF eigenvalues are shifted with respect to noninteracting ones by the corresponding matrix element of $\nu_{\text{H}} + \Sigma_{\text{x}}$. A given matrix element of G_{HF} reads

$$G_{\text{HF}} = \frac{1}{\omega - \epsilon^0 - \nu_{\text{H}} - \Sigma_{\text{x}}} \quad (21)$$

the interaction potentials shift the poles of the Green's function, with respect to the noninteracting case. Now imagine to iterate the Dyson equation by starting from $G_{\text{HF}} = G_0 = 1/(\omega - \epsilon^0)$. This corresponds to an expansion of G_{HF} in the interaction,

$$G_{\text{H}} = \frac{1}{\omega - \epsilon^0} + \frac{(\nu_{\text{H}} + \Sigma_{\text{x}})}{(\omega - \epsilon^0)^2} + \frac{(\nu_{\text{H}} + \Sigma_{\text{x}})^2}{(\omega - \epsilon^0)^3} + \dots \quad (22)$$

All contributions have a pole at the same energy, although they are of increasing order. In order to approximate the shift of pole energy that appears in the exact solution Eq. (21), one needs a huge number of terms. By solving the Dyson equation one obtains directly Eq. (21), which is an expression for the Green's function that contains contributions of all orders in the Coulomb interaction. This is true even when the self-energy itself is of low order, such as in the case of HF. This shows that in order to obtain correct addition and removal spectra, perturbation theory for the self-energy might be much more efficient than perturbation theory for the Green's function itself.

As already the simple HF shows, the self-energy is nonlocal in space. Beyond HF, as discussed earlier it should also contain the reaction of the system to the propagation of an extra charge, which is in general not instantaneous. Therefore, the self-energy is also nonlocal in time, or frequency-dependent.

The main message of this subsection is the idea to introduce a space-spin and time- nonlocal effective potential called *self-energy*. This allows one to describe interaction effects in the Green's function by solving an integral equation called *Dyson equation*. This equation creates contributions to the Green's function to all orders in the interaction, even when a low-order approximation for the self-energy is used. Therefore, we can expect that it will be easier to approximate the self-energy than to approximate directly the Green's function itself.

THE GW APPROXIMATION: THEORY

In order to find a good approximation for the self-energy, let us start from HF, where $\Sigma = \nu_H + \Sigma_x$. This is probably the most well-known approximation for the self-energy, but as we have seen, it suffers from the complete neglect of screening. Our aim is now to go beyond this approximation. There are many ways to achieve this goal. Often the problem is looked at in terms of perturbation theory,⁴ where the noninteracting $\Sigma = 0$ is the zero-order term, and one expands the self-energy in orders of the Coulomb interaction. This leads to many-body perturbation theory and diagrammatic expansions,²⁰ with established rules for creating higher order terms, and with a graphical representation (called *diagrams*) of the relatively complex multi-dimensional integrals that have to be evaluated. For example, the HF self-energy can be represented as shown in Figure 3, where the continuous line with an arrow represents a Green's function and the dashed line the bare Coulomb interaction.

If one continues the expansion in the Coulomb interaction one finds an increasing number of increasingly complicated diagrams, representing the various processes that can occur during the propagation of an electron or hole, for example, electron–electron scattering. Usually one has to make a choice and concentrate on certain processes that are supposed to be dominant. To obtain the GW approximation, one supposes that the most important contribution beyond HF is the creation of electron–hole pairs as a



FIGURE 3 | Feynman diagrams for the Hartree–Fock self-energy. The dashed line represents the bare Coulomb interaction, and the arrow a one-body Green's function. The arrow closed as a circle is the density Eq. (3) (the diagonal of the Green's function Eq. (8)). The first contribution is the Hartree potential, and the second contribution is the exchange self-energy.

response to the propagation of the extra charge. These electron–hole pairs are responsible for the screening, which distinguishes GW from HF. The explicit link between the creation of electron–hole pairs and screening is summarized in Box 1.

In order to put the GW approximation in a more general context, and to have a clearer idea about what is included and what is neglected, the following subsection contains a formal derivation based on functional derivatives. A reader not interested in this derivation may stick to the insights gained from the equations, which are highlighted in several places. The presentation of the derivation is compact, and some details can be found, for example, in Refs 21,22.

Derivation of the GW Approximation from the Equation of Motion

We start with the equation of motion for the one-body Green's function, which describes the change of G with respect to one of its time arguments. Using Eq. (8) one finds

$$\left[i \frac{\partial}{\partial t_1} - h(1) - \nu_H(1) \right] G(1, 1') - i \int dx_2 \nu_c(\mathbf{r}_1, \mathbf{r}_2) L(1, 2, 1', 2^+) |_{t_2=t_1^+} = \delta(1, 1') \quad (23)$$

where the compact notation $(1) \equiv (x_1, t_1) \equiv (\mathbf{r}_1, \sigma_1, t_1)$ has been adopted. Here, $L(1, 2, 1', 2') = -G_2(1, 2, 1', 2') + G(1, 1')G(2, 2')$ is called the two-body correlation function, since it excludes the uncorrelated contribution $G(1, 1')G(2, 2')$ from the time-ordered two-body Green's function G_2 defined as

$$G_2(1, 2, 1', 2') \equiv (-i)^2 \langle \Psi | T [\hat{\psi}(1) \hat{\psi}(2) \hat{\psi}^\dagger(2') \hat{\psi}^\dagger(1')] | \Psi \rangle \quad (24)$$

The spin-resolved density-density response function^b χ is related to the two-body correlation function by

$$\chi(12) = -iL(12, 1^+; 2^+) \quad (25)$$

and the equilibrium charge–charge response function in frequency space is obtained as a sum in a spin basis

$$\chi(\mathbf{r}, \mathbf{r}'; \omega) = \sum_{\sigma, \sigma'} \int d\tau e^{i\omega\tau} \chi_{\sigma, \sigma'}(\mathbf{r}, \tau; \mathbf{r}', 0) \quad (26)$$

One may call L a generalized response function, since it can be expressed as the variation of the one-body Green's function with respect to an external potential u :

$$\frac{\delta G(2,1')}{\delta u(3)} \Big|_{u=0} = L(2,3,1',3^+) \quad (27)$$

Since with Eqs (3) and (8) the density for a system with spin-independent interaction, where G is spin-diagonal, is

$$n(\mathbf{r}) = -i \sum_{\sigma} G_{\sigma}(\mathbf{r}, t, \mathbf{r}, t^+) \quad (28)$$

this contains the widely used linear response relation

$$\frac{\delta n(\mathbf{r}, t)}{\delta u(\mathbf{r}', t')} \Big|_{u=0} = \chi(\mathbf{r}, t; \mathbf{r}', t') \quad (29)$$

If the interaction is set to zero in Eq. 23, one finds that

$$\left[i \frac{\partial}{\partial t_1} - h(1) \right] G_0(1, 1') = \delta(1, 1') : \quad (30)$$

indeed, the noninteracting Green's function G_0 (Eq. (15)) is solution of this differential equation in equilibrium, i.e., when $h(1)$ does not depend on time. More generally, the same form (Eq. (15)) for the equilibrium solution is obtained when one adds any instantaneous static potential to h , such as the Hartree potential, the HF self-energy, or a KS potential.

Using Eq. (30), one can reformulate Eq. (23) as

$$G(1, 1') = G_0(1, 1') + G_0(1, \bar{2}) \nu_H(\bar{2}) G(\bar{2}, 1') \\ + i G_0(1, \bar{2}) \nu_c(\bar{2}, \bar{3}) L(\bar{2}, \bar{3}^+, 1', \bar{3}^{++}) \quad (31)$$

where arguments with a bar are integrated, $f(\bar{1})g(\bar{1}) \equiv \int d1 f(1)g(1)$, and the Coulomb interaction ν_c carries a δ -function in time, since it is instantaneous, $\nu_c(1, 2) = \delta(t_1 - t_2) \nu_c(\mathbf{r}_1 - \mathbf{r}_2)$. Therefore, three equal times appear in L , and the infinitesimal time differences defined in *Notation and Fourier Transforms* are needed to guarantee the correct order of the field operators.

Equation (31) is the equation of motion for the one-body Green's function. It highlights the essential physics:

The propagation of a particle in an interacting system (given by G) equals its independent-particle expression G_0 plus a correction due to the Hartree

potential, and a term that involves two-particle correlations and contains all exchange-correlation corrections.

If one neglects the last term in Eq. (31), one finds automatically the Dyson equation that yields the Hartree approximation. It is not obvious how to go beyond, since in order to calculate G_2 from Eq. (24), and therefore L , one would again need the unknown many-body state $|\Psi\rangle$. To obtain a closed equation for G , one can use the linear response relation Eq. (27), and rewrite Eq. (31) as

$$G_u(1, 1') = G_0(1, 1') + G_0(1, \bar{2}) \left\{ [u(\bar{2}) + \nu_H(\bar{2})] G_u(\bar{2}, 1') + i \nu_c(\bar{2}, \bar{3}) \frac{\delta G_u(\bar{2}, 1')}{\delta u(\bar{3}^+)} \right\} \quad (32)$$

Here, we suppose that the actual external potential of the system is already contained in G_0 , and u merely represents an additional potential that is added for the sake of the derivation, and that is set to zero at the end of the calculation: in that case, the solution $G_u \rightarrow 0$ should be the physical equilibrium Green's function of the material. Before this last step, all quantities depend on u , which is highlighted here by the subscript u .

On the upside, we have now a closed differential equation for G . If we were able to solve it, we would have solved the many-body problem: Knowing G_u as a functional of u would mean that we could take all derivatives with respect to u , and in this way obtain all correlation functions; Eq. (27) is an example for the first, linear, order. This, in turn, would allow us to calculate all possible expectation values.

On the downside, we do not know how to solve this nonlinear (note that the Hartree potential depends on G itself through the density!), multi-dimensional, functional integro-differential equation. Even worse, such an equation can have many solutions, and it is not obvious how to select the physical one (for a discussion, see, e.g., Ref 23).

To make the problem tractable, let us concentrate on the term $\delta G_u / \delta u$ in Eq. (32). We do not know in general how G_u depends on u . In order to transform the equation to a more useful form, we first insert a chain rule, $\delta G_u(2, 1) / \delta u(3) = [\delta G(2, 1) / \delta u_{\text{tot}}(\bar{5}, \bar{6})] [\delta u_{\text{tot}}(\bar{5}, \bar{6}) / \delta u(3)]$ with respect to some still to be defined, possibly nonlocal, potential u_{tot} , and then use the fact that $\delta G(2, 1) / \delta u_{\text{tot}}(\bar{5}, \bar{6}) = -G(2, \bar{4}) \delta G^{-1}(\bar{4}, \bar{8}) / \delta u_{\text{tot}}(\bar{5}, \bar{6}) G(\bar{8}, 1)$. In this way, the equation of motion becomes a Dyson equation $G = G_0 + G_0 \Sigma G$ with the self-energy given by

$$\Sigma(2,8) = \delta(2,8)v_H(2) + i\nu_c(2,\bar{3})G(2,\bar{4})$$

$$\frac{\delta G^{-1}(\bar{4},8)\delta u_{\text{tot}}(\bar{5},\bar{6})}{\delta u_{\text{tot}}(\bar{5},\bar{6})\delta u(3)} \quad (33)$$

If we chose u_{tot} to be the total potential

$$u_{\text{tot}}(5,6) \equiv \delta(5,6)u(5) + \Sigma(5,6) \quad (34)$$

this equation simplifies to

$$\Sigma(2,8) = \delta(2,8)v_H(2) + iG(2,\bar{4})W^{\text{tot}}(2^+,\bar{4},8) \quad (35)$$

with

$$W_u^{\text{tot}}(2,4,8) \equiv \nu_c(2,\bar{3})\frac{\delta u_{\text{tot}}(4,8)}{\delta u(\bar{3})} \quad (36)$$

Once the self-energy is determined, the Dyson equation can be solved in the limit $u \rightarrow 0$ to obtain equilibrium properties.

Let us compare the exact expression (35) to the HF approximation Eq. (9). The difference lies in the interaction: instead of the bare interaction ν_c , the exchange–correlation contribution Σ_{xc} to the exact self-energy contains the effective interaction W^{tot} ; this interaction is modified with respect to the bare one by the response of the Hartree- and exchange–correlation potentials. The variation of the system-internal potentials due to a perturbation is the *screening* that is missing in HF. To summarize,

The exact exchange–correlation self-energy $\Sigma_{\text{xc}}(1,2)$ can be expressed as the integral of an interacting Green's function $G(1,\bar{4})$ and an effective interaction $W^{\text{tot}}(1,\bar{4},2)$. If the effective interaction is replaced by the bare Coulomb interaction $\nu_c(12)\delta(2,\bar{4})$ the HF approximation is obtained. Beyond HF, the variation of the system-internal potentials modifies the effective interaction: it is *screened*.

Of course, we do not know the effective interaction W^{tot} , since its calculation would require the knowledge of the self-energy contained in u_{tot} according to Eq. (34), so we are creeping around the problem. However, at least some parts of the self-energy are already known, so one can hope to treat the problem iteratively.

Let us start with the simplest approximation, where the self-energy has no exchange–correlation contribution, namely the Hartree approximation

$\Sigma_u(2,3) \approx \delta(2,3)v_{Hu}(2)$. Then, Eqs (35) and (36) yield the approximate self-energy

$$\Sigma(1,2) \approx \delta(1,2)v_H(1) + \Sigma_{\text{xc}}(1,2) = \delta(1,2)v_H(1) + iG(1,2)W(1^+,2) \quad (37)$$

with the effective Coulomb interaction

$$W(1,2) = \epsilon^{-1}(1,\bar{3})\nu_c(\bar{3},2) \quad (38)$$

which is screened by the inverse dielectric function defined as (see also Box 1)

$$\epsilon^{-1}(1,2) \equiv \frac{\delta(u(1) + v_H(1))}{\delta u(2)} = \delta(1,2) + \nu_c(1,\bar{3})\chi(\bar{3},2) \quad (39)$$

This is the exact so-called *test charge–test charge* inverse dielectric function, which is a measurable quantity: for example, inelastic X-ray scattering or electron energy loss spectroscopy experiments give access to the imaginary part of ϵ^{-1} , and, via the Kramers–Kronig relations, also to its real part.⁷ The effective interaction W is called the screened Coulomb interaction, and the resulting approximate self-energy²⁴ is Hedin's GW approximation (GWA).²⁵

With respect to HF, the GWA exchange–correlation contribution to the self-energy replaces the bare interaction $\nu_c(1,2)$ by a screened interaction $W(1,2)$. This is not the full effective interaction W^{tot} , but the bare Coulomb interaction screened by the classical charge response of the interacting quantum system to a classical perturbation.

Note that the Hartree potential remains unscreened. Figure 4 shows the diagrammatic representation of the exchange–correlation contribution to the GW self-energy.

Physics of the GW Approximation

Let us now analyze the GWA self-energy. Used in the Dyson equation, it allows one to access the interacting Green's function G , which *directly yields electron addition and removal spectra*. Even though GW is an approximation, this is a major difference to any KS result. Another difference to KS is the appearance of the nonlocal Fock exchange in the self-energy. It implies that the *self-interaction contained in the Hartree potential is canceled*, contrary to most approximate exchange–correlation functionals in the DFT framework.^k

Most importantly, Eq. (38) expresses the fact that with respect to the HF self-energy the GWA one contains *screening*. The bare Coulomb interaction does not depend on spin; therefore, the screened Coulomb interaction W does not depend on spin either, and the screening stems from the spin-averaged response function $\chi(\mathbf{r}, \mathbf{r}'; \omega)$ defined in Eq. (26).

Let us for the moment neglect the time dependence of the functions. If W were instantaneous, $W(\mathbf{r}, \mathbf{r}')\delta(t^+ - t')$, the exchange–correlation contribution to the GW self-energy would read

$$\Sigma_{xc}(x, x') = -\rho(x, x') W(\mathbf{r}, \mathbf{r}') \quad (40)$$

This expression is nothing else but the statically screened Fock operator (Eq. (6)). Even in this simple approximation,^{26,27} the effect of screening is dramatic, especially in extended systems. This becomes clear in the homogeneous electron gas, where in the case of a spin-diagonal density matrix Eq. (40) reads in reciprocal space:

$$\Sigma_{xc\sigma}(\mathbf{k}) = -\frac{1}{(2\pi)^3} \int d\mathbf{q} \rho_{\sigma}(\mathbf{q}) W(\mathbf{k} - \mathbf{q}) \quad (41)$$

In HF, $W(\mathbf{k} - \mathbf{q}) \rightarrow v_c(\mathbf{k} - \mathbf{q}) = 4\pi/|\mathbf{k} - \mathbf{q}|^2$. As v_c is long-range, it diverges for $\mathbf{q} \simeq \mathbf{k}$, and the Fock self-energy displays a singular variation when \mathbf{k} passes through the Fermi surface, which leads to a pathological density of states.²⁸ A screened interaction, instead, has the form $W(\mathbf{k} - \mathbf{q}; \omega = 0) = 4\pi/(|\mathbf{k} - \mathbf{q}|^2 + \lambda^2)$, where λ is an inverse screening length. There is no divergence, and one obtains a well-behaved self-energy.^{25,29,30} Hybrid functionals, which are widely used in chemistry and physics, are generalized KS exchange–correlation potentials²⁶ that contain a fraction of exact nonlocal Fock exchange. They can be regarded as approximations to the screened exchange approximation contained in the GWA.³¹ In extended systems range separation³² can be used to suppress the long-range (small- $\mathbf{k} - \mathbf{q}$) Fock contribution; this is a way to simulate the effect of screening and obtain well-behaved results; see, e.g. Refs 33–37.

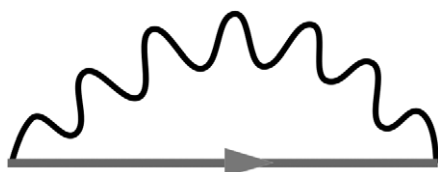


FIGURE 4 | Exchange–correlation self-energy in the GWA. The test-charge–test-charge screened interaction is given by a wiggly line.

The density–density response function, which is responsible for the screening through Eq. (39), links the external potential and the induced density at different times: χ depends on two times or, in equilibrium, on one time difference. Therefore the Fourier transform of W , and by consequence the self-energy, are frequency dependent. The screened Coulomb interaction can be written in a spectral representation:

$$W(\mathbf{r}, \mathbf{r}'; \omega) = v_c(\mathbf{r}, \mathbf{r}') + \sum_{\lambda \neq 0} \frac{2\omega_{\lambda} W_{\lambda}^p(\mathbf{r}, \mathbf{r}')}{\omega^2 - \omega_{\lambda}^2} \quad (42)$$

where $\lambda \neq 0$ labels excited states of the many-body system with energies ω_{λ} , and where W_{λ}^p are the amplitudes of $W^p \equiv v_c \chi v_c$, the polarization part of $W = v_c + v_c \chi v_c$ (Eqs (39) and (38)). According to the system, the excitations can be discrete and/or form a continuum.¹ The corresponding energy difference is $\omega_{\lambda} = E_{N, \lambda} - E_N$, the energy of a neutral excitation.

In practice, the GW self-energy is often evaluated by replacing G with an independent-particle Green's function. Then Eqs (37) and (42) lead to^m

$$\Sigma_{xc}(x, x'; \omega) = \Sigma_x(x, x') + \sum_{i, \lambda \neq 0} \frac{\varphi_i(x) \varphi_i^*(x') W_{\lambda}^p(\mathbf{r}, \mathbf{r}')}{\omega + \omega_{\lambda} \operatorname{sgn}(\mu - \varepsilon_i^0) - \varepsilon_i^0} \quad (43)$$

By neglecting the difference $(\omega - \varepsilon_i^0)$, the ‘*COulomb Hole plus Screened EXchange (COHSEX)*,’ COHSEX approximation is obtained.²⁹ The COHSEX self-energy reads

$$\Sigma_{xc}(x, x') = -\rho(x, x') W(\mathbf{r}', \mathbf{r}; \omega = 0) + \frac{1}{2} \delta(x - x') W^p(\mathbf{r}', \mathbf{r}; \omega = 0) \quad (44)$$

Like Eq. (40), it contains a statically screened exchange contribution. Moreover, there is a static ‘Coulomb hole,’ which is the local potential that is created by the adiabatically induced charge density around an external point charge.^{24,25}

The COHSEX self-energy is a static approximation and gives therefore rise to a spectral function of independent-particle form. The frequency dependence that is displayed by the full GW self-energy (Eq (43)) is responsible for the *dynamical effects*. It distinguishes the GW approximation from any static mean-field theory: in a system with a discrete spectrum, the GW spectral function shows a number of excitations that is larger than in the independent-particle case. In the case of a continuous spectrum, it yields broadening and satellites, such as in the picture of Figure 2.

Alternatively to the Galitskii–Migdal expression¹⁰ Eq. (14), one can express the total energy as a variational functional of the interacting Green's function.^{38–41} In the GWA, this functional has interesting properties; in particular, through the appearance of W it introduces long-range polarization effects that give rise to the van der Waals dispersion interaction, a phenomenon that is difficult to capture with simple density functionals. When the random-phase approximation (RPA) (see the following section) is used to calculate W , and when the total energy functional is evaluated with an independent-particle KS–Green's function instead of the self-consistent GW one, the GW total energy becomes^{39,42}

$$E^{\text{GW}}[G_{\text{KS}}] = E_{\text{HF}}[\varphi^{\text{KS}}] + \frac{1}{2} \text{Tr}[\nu_{\text{c}} G_{\text{KS}} G_{\text{KS}} + \ln(1 - \nu_{\text{c}} G_{\text{KS}} G_{\text{KS}})] \quad (45)$$

where Tr stands for a trace over space, spin, and frequencies. The first term is the HF energy calculated with KS orbitals; the remainder is the correlation correction. This expression for the total energy is called RPA in the framework of DFT.^{43–45} It reflects good features of the full GWA functional, in particular it also describes the van der Waals dispersion. More discussion about GWA and RPA total energies can be found, e.g., in Refs 46–50.

Screening in Finite and Infinite Systems

Before moving on to more practical details, it is important to clarify one point that might give rise to confusion. In finite systems, often the so-called Δ self-consistent field (Δ -SCF) approach is used to calculate addition or removal energies. In this approach, the total energy of a system with an additional electron or hole is calculated in some mean-field approximation by populating or depopulating a localized orbital, and the system is allowed to relax adiabatically. This mean-field approximation could be, for example, HF.⁴⁶ The linear response contribution of the relaxation of the Hartree potential simulates the physics of the GWA and indeed, results in finite systems are often similar.^{51–53} Does not this contradict our initial statement where it was said that HF does not contain screening?

Indeed, we have to define carefully what is meant here. First, for example, in the comparison of band structures in Figure 1, HF eigenvalues were used. They correspond to electron addition or removal total energy differences via Koopmans' theorem only if the system is not allowed to relax. In a

finite system this makes a difference, because an additional charge is a strong perturbation that induces significant relaxation in a self-consistent total energy calculation, such as in the Δ -SCF approach. However, we always refer to eigenvalues in our discussion, and this is the sense in which 'HF contains no screening.'

Second, in an infinite periodic system the situation changes. Indeed, if we insert one additional charge in a Bloch state, it is completely smeared out and induces only an infinitesimal relaxation of the system. The relaxation energy is quadratic in the extra charge density, and therefore vanishingly small. Hence Koopmans' theorem is valid, and we cannot get away with Δ -SCF.

The merit of the GWA, or already its static COHSEX approximation, is to overcome this problem. The screening that is introduced by the GWA corresponds implicitly to a different interpretation of what it means to add a charge: it is not to create a delocalized extra charge density whose distribution is given by an eigenstate of the periodic system, but to put an extra localized charge to some place with some probability. The eigenfunction tells us the probability amplitude for the charge to go to a given place; in particular, in a periodic system the Bloch function carries the information that the probability is the same in all periodically repeated units. The addition of a localized charge leads to a significant relaxation energy also in an infinite system, reflected in the screened eigenvalues of the GWA. Therefore in the presence of extended states, where Δ -SCF or similar approaches using simple functionals would have problems, the GWA makes a qualitative difference.

THE GW APPROXIMATION IN PRACTICE

What does it mean to do a GW calculation? The aim of this section is to become more concrete on this question.

The Random-Phase Approximation

To start with, none of the above expressions can be calculated in practice, as we do not know the exact dielectric function: for this, we would have to evaluate Eq. (29), which means that we would have to know how the exact charge density depends on a general time-dependent external potential. To overcome the problem, let us start from χ from Eqs (25) and (27). By expanding the derivative using the

chain rule with the total potential defined in Eq. (34), χ can be written as

$$\chi(1,2) = -iG(1,2)G(2,1^+) - iG(1,\bar{3})G(\bar{4},1^+) \\ \left(\delta(\bar{3},\bar{4})\nu_c(\bar{3},\bar{5}) + \frac{\delta\Sigma_{xc}(\bar{3},\bar{4})}{\delta n(\bar{5})} \right) \chi(\bar{5},2) \quad (46)$$

This is a Dyson equation for χ . However, the exchange–correlation contribution $\delta\Sigma_{xc}/\delta n$ is unknown. The simplest approximation is the RPA, where it is completely neglected.^o Neglecting the variation of the exchange–correlation contribution is perfectly within the spirit of the GW approximation: we have already done this to move from Eq. (35) to Eq. (37). More pragmatically, this approximation makes the calculation of W feasible for realistic systems. For both reasons, the GW approximation usually also implies the RPA for the calculation of W .

The RPA has been introduced by Bohm and Pines⁵⁴ for the homogeneous electron gas. It is equivalent to the time-dependent Hartree approximation,⁵⁵ when Hartree Green's functions instead of G are used to evaluate Eq. (46), with $\delta\Sigma_{xc}/\delta n \rightarrow 0$. In real calculations one usually finds the use of KS Green's functions, or sometimes a self-consistently calculated G , for example, from GW. In any case:

In the RPA the density–density response function χ consists of noninteracting electron–hole pairs (given by a product of two Green's functions) that are connected by the bare Coulomb interaction. This coupling is due to the self-consistent variation of the Hartree potential in the response. The variation of exchange–correlation contributions is neglected.

The resulting Dyson equation (Eq. (46) with $\delta\Sigma_{xc}/\delta n \rightarrow 0$) is depicted in Figure 5.

Things to be Calculated

To make a GW calculation, one needs an input independent-particle electronic structure with which to construct a starting G and W . Preferably, it should already be close to the final result. In extended systems one, uses typically KS eigenfunctions and eigenvalues; hybrid functionals are often a good choice in semiconductors and insulators.



FIGURE 5 | Dyson equation for the polarizability χ in the random-phase approximation (RPA).

At the heart of a GW calculation lies the determination of screening. It starts with the calculation of the independent-particle response function $\chi_0(1,2) \equiv -iG(1,2)G(2,1^+)$, using the input electronic structure to build G . Subsequently one solves the screening Dyson equation depicted in Figure 5, which allows one to obtain the inverse dielectric function Eq. (39) and finally, the screened Coulomb interaction W from Eq. (38). Now the self-energy can be set up along the lines of Eq. (43). In extended systems, the sum over excitation energies is an integral over frequencies. Several approaches exist to deal with this numerical problem, e.g. Refs 56–64.

Once the self-energy is obtained, quasi-particle energies and/or a new Green's function can be calculated on various levels of approximation. Often calculations are limited to an approximate calculation of quasi-particle energies ϵ_i from first-order perturbation theory with respect to KS,

$$\epsilon_i = \epsilon_i^{\text{KS}} + [\Sigma_{xc}(\epsilon_i) - \nu_{xc}]_{ii} \approx \epsilon_i^{\text{KS}} + Z_{ii} [\Sigma_{xc}(\epsilon_i^{\text{KS}}) - \nu_{xc}]_{ii} \quad (47)$$

where

$$Z_{ii} \equiv \left(1 - \frac{\partial [\Sigma_{xc}(\omega)]_{ii}}{\partial \omega} \Big|_{\omega = \epsilon_i^{\text{KS}}} \right)^{-1} : \quad (48)$$

here $\Sigma_{xc} - \nu_{xc}$ is treated as a small perturbation with respect to KS, so quasi-particle and KS orbitals φ_i are equal. Moreover, $\Sigma(\omega)$ is considered to be approximately linear in frequency around the quasi-particle energy.^{60p} A calculation where G and W are built with a starting independent-particle electronic structure is called 'G₀W₀.' This and the further approximations in Eq. (47) can be dropped, and one can go up to a fully self-consistent solution of the Dyson equation, which requires to re-calculate G several times. An intermediate, today often adopted, level of self-consistency is given by the quasi-particle self-consistent GW approximation (QSGW), where the fully frequency-dependent GW self-energy is replaced by an optimized static nonlocal operator in the self-consistency cycle.^{65–67} This approach is easier to use, and it avoids the drawbacks of a fully dynamical self-consistent calculation, in particular the violation of the f -sum rule that one finds if W is calculated in the RPA with fully interacting Green's functions.⁶⁸

As an output of a GW calculation one obtains quasi-particle energies, spectral functions and/or total energies, and when one goes beyond the first-order perturbation theory approach, also updated charge densities (see, e.g., Refs 69–73), and density matrices. Of course,

dealing with nonlocal objects has a computational price: the number of atoms N per unit cell that can be treated in typical ground-state DFT calculations today can be of an order of magnitude of 10^4 , whereas GW calculations usually deal at best with some 10^3 atoms. This goes hand-in-hand with the scaling properties: the tendency in DFT codes is to have a dominant scaling of N^2 , whereas most GW codes scale as N^3 or N^4 . Still, with these orders of magnitude, many interesting applications are within the reach of GW calculations.

Applications

GW Calculations for Extended and Finite Systems

The first and probably most well-known success of the GWA was to overcome the band gap underestimate of KS eigenvalue differences in simple semiconductors.^{60,74–76} The example of silicon is shown in Figure 1: the GWA gap is about twice as large as the eigenvalue difference calculated in KS-LDA. On the other hand, the GWA gap is significantly smaller than the HF eigenvalue gap, and overall the GWA band structure is in good agreement with experiment. The band structure or single aspects such as effective masses, band offsets, or band widths have been calculated successfully for many other simple semiconductors and insulators.^{22,77,78} Even the band structure of materials containing strongly localized electrons such as d- or f- electrons in transition metal oxides can often⁹ be well described; in these cases, it is necessary to carefully chose the starting point of a G_0W_0 calculation, or to be at least partially self-consistent, in order to have a good description of the charge density^{65–80} (see also *The Starting Point*). Furthermore, the GWA is used to calculate finite lifetimes due to impact ionization.⁸¹ It also yields satellite structure, but in many cases the agreement with experiment is poor. It can, however, be improved by using the GW self-energy in a cumulant expansion (see *What Is Wrong With the GW Approximation?*).

The fact that the GWA contains screening is also important in the calculation of total energies. For example, the GWA can treat van der Waals dispersion, a major challenge for DFT functionals. Screening, moreover, is important to describe image states outside a surface or a cluster; the GWA can therefore do this successfully.⁸² For an overview and more references concerning applications, see. Ref 22.

When the dimension and size of a system is reduced, states evolve from delocalized Bloch states in a crystal to more localized states, and there is less

screening. Therefore, the screened exchange contribution increases, toward the Fock exchange. However, even for quite small systems, down to an atom or a few atoms, one cannot in general neglect screening completely, and the GWA is clearly superior to HF for the calculation of ionization energies and electron affinities. At the same time, it also corrects KS results, where the smallest ionization energy is *in principle* correct, but results are spoiled by the use of approximate functionals. Therefore, the GWA meets an increasing interest in the chemistry community (see, e.g., calculations of energy levels in molecules described in Ref 83, and references therein).

Of course, a method that is born in the homogeneous electron gas has to be thoroughly re-examined, and eventually re-optimized, when it is to be used for small finite systems. Therefore a major part of the ongoing effort is devoted to benchmarking the GWA in systems relevant for chemistry, to find the most efficient implementations, and the most reliable level of approximation. Many issues have still to be settled. The rest of this section is not meant to contain an exhaustive literature review, but to highlight and illustrate questions and directions of research that are specific for quantum chemistry applications, with the help of a few examples.

Quantum Chemistry Implementations

Technical choices are not necessarily the same for extended and finite systems. One can easily understand that in periodic systems basis sets such as plane waves are often a good choice, whereas in small systems a localized basis may be more suitable. Various implementations have been proposed and described in detail (see, e.g., Refs 64 and 84–88), which are often adapted to standard quantum chemistry packages. These implementations have been used for extensive benchmarking. For example, in Ref 64, a test set of 27 molecules (called GW27) ranging from H_2 to naphthacene has been considered. The work of Ref 84 focused on a test set of 29 molecules ranging from H_2 to tetrathiafulvalene. The systems examined in Ref 85 include the benzene and other molecules, water clusters containing up to 480 atoms, small CdSe nanoparticles, and linear acenes, whereas the work of Ref 86 concentrates on relativistic effects, which are studied in molecules, solids, and nanocrystals. A large set of benchmark molecules proposed to benchmark GWA calculations is the GW100 set of Ref 89.

The Starting Point

To build the GW self-energy Eq. (37), one would in principle need G and W , but the Green's function is a *result* of the calculation. Many calculations use a 'best

guess' for G and W to determine the self-energy, for example, a KS Green's function and an RPA W calculated from KS energies and orbitals. As explained in *Things to be calculated*, this is called a G_0W_0 calculation. Of course, results depend on this starting point, and the choice of an optimal starting point is crucial. In Ref 90, GW calculations were performed for acenes for various lengths of the chains, ranging from benzene to hexacene. Different hybrid functionals were used as starting points, and eigenvalues (not eigenfunctions) were calculated self-consistently. Ionization potentials were well described, but electron affinities were not satisfactory for the larger acenes, pointing to the need for a more complete self-consistency scheme, for a still improved starting point, or for corrections terms beyond the GWA. In Ref 91, a wide range of starting points for G_0W_0 was examined, including HF, LDA, PBE, PBE0, B3LYP, HSE06, BHPHLYP, CAM-B3LYP, and tuned CAM-B3LYP. For the ionization energy, the hybrid functionals, and in particular those with a high proportion of exact-exchange, yielded the best results. Other important classes of systems are molecules containing transition metal atoms, and molecules for photovoltaic applications. For these, the effect of the starting point and partial self-consistency in the GWA calculations has been explored.⁸³ It was suggested that G_0W_0 on top of PBE0 emerges as a method that is reliable and efficient enough to treat systems of technological interest within this class of materials.

An illustration for a comparison of starting points for a pyridazine molecule is given in Figure 6, taken from Ref 92.

Concerning the question of the starting point, there can be no miracle solution. Because G_0W_0 is a severe approximation, the choice of the starting point is necessarily system dependent. As a guideline, one should make sure that the charge density is well described by the chosen starting functional, in order to capture the important electrostatic effects.

Self-Consistency

In order to improve the results and make calculations more starting point independent, the calculations can be partially or fully iterated. An important step is the update of eigenvalues. For example, eigenvalue self-consistency has been shown to be beneficial for the case of water in Ref 93. Also the quasi-particle wavefunctions may change significantly when the calculations are iterated toward self-consistency. Imagine, for example, a molecule with its first unoccupied level resulting below the vacuum level in a KS calculation. If the GWA increases its energy, it may move above the vacuum level and delocalize. This, in turn, can yield a very different one-body

Green's function and, in the next step of a self-consistent GW calculation, a big change in the self-energy. *Full* self-consistency is more than just an update of quasi-particle eigenvalues and wavefunctions: it requires the solution of the Dyson equation. As the GW self-energy is frequency dependent, after the first self-consistency step the Green's function has no longer an independent-particle form Eq. (16).

There are different arguments, sometimes contrasting, for or against full or partial self-consistency. This is natural, because the GWA is an approximation, and according to the system and the quantity that is studied, contributions from self-consistency and terms that are neglected in the GWA may have a tendency to cancel. Therefore, it is very important to gain insight and experience using different test sets, in order to propose the most reliable versions of GW calculations for certain classes of applications.

Various levels of self-consistency, up to fully self-consistent GW, were examined in Ref 92. Their performance was assessed for electron removal quasi-particle spectra of benzene, pyridine, and the diazines. It was found that fully self-consistent GW does not yield the best results, whereas nonself-consistent or partially self-consistent calculations may be better, if the starting point is carefully chosen. In particular, when the ground-state charge density has a dramatic influence on spectra, it may be wise to use GW with a level of self-consistency which updates wavefunctions. This is the case for the HOMO–LUMO alignment in donor–acceptor systems studied in Ref 94. It has been shown that for these systems it is not enough to choose one of the widely used starting points, such as PBE0, but that in these donor–acceptor compounds, and more generally in systems where the description of ground-state properties depends on the relative alignment of the frontier orbitals of different components (such as, e.g., molecules absorbed on surfaces), partially self-consistent GW significantly changes the density and brings spectral results in better agreement with photo emission experiments and accurate quantum chemistry calculations. The importance of partial (quasi-particle - or eigenvalue-) self-consistency has also been stressed in Ref 84, where GWA results for the dipole moments of five molecules relevant for organic photovoltaics have been benchmarked against experimental results. The effect of an update of the quasi-particle wavefunctions is illustrated in Figure 7. One might argue that a suitably chosen starting point could circumvent the need for self-consistency. This is true; for example, in the charge transfer example of Ref 94 discussed above, hybrids with large fraction of exact exchange lead to

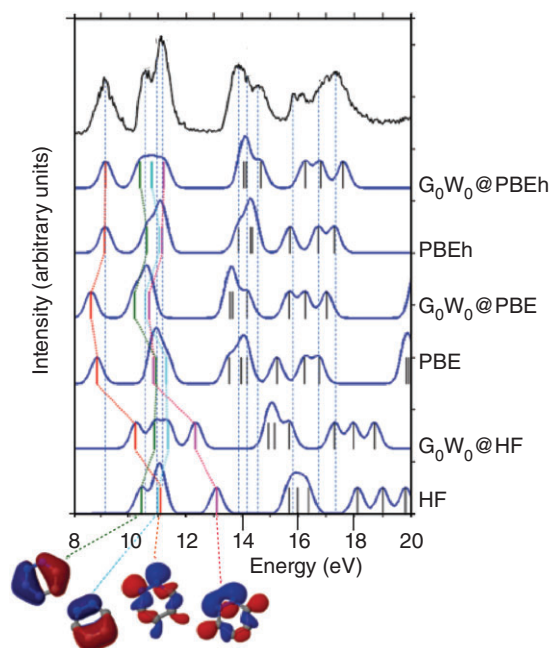


FIGURE 6 | Electron removal spectrum of pyridazine, comparison of theory and experimental photo emission done in the gas phase, from Ref 92. Calculated results include Hartree–Fock and two density functional theory functionals, as well as G_0W_0 using the DFT calculations as starting points. A broadening of 0.3 eV was used. At the bottom the density of the frontier orbitals is shown (Adapted with permission from Ref. 92. Copyrighted by the American Physical Society).

satisfactory results. On the other hand, the strategy of partial self-consistency is certainly more transferable than a case-to-case search for an optimal starting point.

Overall, it emerges that in general QSGW, or at least an update of eigenvalues combined with a good starting density, should be preferred with respect to full self-consistency.

Total Energies

Also GWA total energy calculations for molecules are a topic of current exploration. The textbook example of the H_2 molecule is examined in Refs 95 and 96. In the dissociation limit, GWA results are worse than DFT–RPA ones, although they correspond to similar diagrams (see *Physics of the GW Approximation*). More work is needed to make GWA total energy calculations operative and competitive with respect to DFT ones, especially as GWA calculations are significantly more computationally demanding. It should be noted that GWA total energies can in principle also be computed using variational functionals. In that case, even a nonself-consistent Green’s function should yield results close to fully self-consistent one, which avoids the need of a self-consistency cycle and

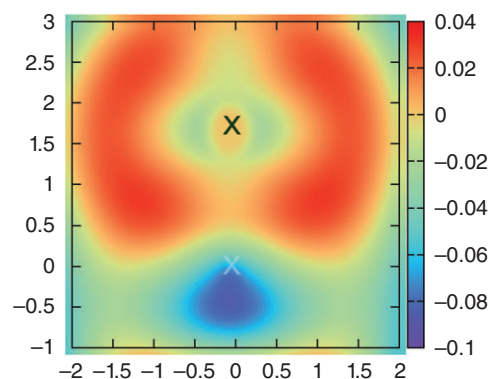


FIGURE 7 | Effect of the update of quasi-particle wavefunctions in a GWA calculation for hydrogen fluoride: shown is the difference between the densities calculated in density functional theory (DFT)-PBE and quasi-particle self-consistent GW approximation, respectively. The result is normalized with respect to the DFT density. The fluoride atom is indicated by a green cross and the hydrogen atom by a gray cross (Reproduced with permission Ref 84. Copyright 2016 American Chemical Society).

leads to a significant reduction of computer time. Work is ongoing in this direction; see, e.g. Refs 38–41.

Overall, and in view of the fact that with today’s implementations of the GWA systems of hundreds or a few thousands of electrons can be studied (see, e.g., Ref 86), it seems worthwhile to further investigate the potential of GWA calculations in a context of quantum chemistry.

WHAT IS WRONG WITH THE GW APPROXIMATION?

The GWA is today part of the standard toolbox in condensed matter physics, and it is also increasingly used in chemistry, with applications to molecules and clusters. As we have discussed above, it contains several pieces of important physics. However, it is an approximation, and as such bound to fail sooner or later. The present review would be misleading without mentioning some of the main problems.

Examples of Failures

To start with a toy system, let us look at a Hubbard dimer with one electron in its ground state. This system consists of two equivalent sites, each with one orbital. The Hamiltonian reads

$$H = -t \sum_{\sigma} [c_{1\sigma}^{\dagger} c_{2\sigma} + c_{2\sigma}^{\dagger} c_{1\sigma}] + U \sum_{i=1,2} n_{i\downarrow} n_{i\uparrow} \quad (49)$$

it contains a kinetic term proportional to the hopping parameter t , and an interaction contribution that is proportional to the onsite interaction U . This is probably the simplest model for the H_2^+ molecule. For electron removal, the exact solution is trivially noninteracting, but for electron addition this is an interacting problem that can be solved exactly (see, e.g., Refs 97 and 98). The ratio U/t is a measure of the effective interaction strength. The larger U/t , the more perturbation theory using the interaction as expansion parameter has problems. Indeed, in the atomic limit ($t \rightarrow 0$) this is a textbook example of strong correlation.

Two typical spectral functions are shown in Figure 8. They have been calculated for the case where the ground state consists of one spin-up electron. In the left panel, the spin-up spectral function is shown for moderate interaction strength ($U = 5$, $t = 1$). The exact spectral function is noninteracting, also for electron addition, because the second spin-up electron must go on the empty site, and the interaction is short ranged. Therefore, the spectral function shows two peaks separated by a gap of $2t$, the bonding–antibonding separation. The GWA slightly overestimates this gap. Moreover, it creates spurious satellites. In particular, the GWA transforms a noninteracting problem (the removal of a single electron) into an erroneously interacting one; this is due to the fact that the electron screens itself, since it enters the calculation of W . It is a clear illustration of the self-screening (also called self-correlation) problem of the GWA. In the right panel, the spin-down spectral function in the atomic limit is shown. The exact result has two peaks, as an additional electron can meet an empty or an occupied site. The GWA places only one peak, in an average position, because it ‘sees’ half an electron on each site; this corresponds to an interpretation of the system density as a classical charge distribution.^r

Also in the case of two electrons in the ground state of the Hubbard dimer, the GWA fails to reproduce the exact spectral functions of electron addition or removal when the interaction is much larger than the hopping parameter. The problems are reflected in the total energy of a H_2 molecule⁹⁵: in particular, GW is not size extensive, which means that one does not obtain the sum of the energies of two isolated atoms in the dissociation limit.

Even in extended systems the GWA can sometimes have severe problems. Similarly to the atomic limit of the Hubbard dimer, unless there is symmetry breaking the GWA cannot describe the gap of a Mott insulator, which is by definition purely due to correlation. For example, paramagnetic NiO is an insulator,

but the GWA yields a metal. In the low-temperature antiferromagnetic phase where the spins are fixed and therefore the translational symmetry is broken, GW band structure results are instead very good.⁸⁰

It is also difficult for the GWA to describe satellites. This may be astonishing, because, as one can see, for example, from Eq. (43), the appearance of a frequency-dependent W leads to the coupling of the propagating fermion to neutral excitations of the system.^s However, there are several problems:

1. The RPA for W . This may exclude important exchange–correlation effects, including some multiplets, which should be seen in the satellites.
2. The appearance of W itself, even if one could calculate it exactly, is an approximation: the derivation in *From Wavefunctions to Green's Functions* shows that the correct effective interaction to be used is W^{tot} from Eq. (36), not the measurable test-charge–test-charge W . Using W instead of W^{tot} has several important consequences: the first is the self-screening problem, which has been illustrated for the Hubbard dimer above. More generally, the full W^{tot} in the exact self-energy Eq. (35) contains the coupling to excitations that are not contained in W and therefore in the GWA, such as spin-flip, or hole–hole excitations.

The second issue is responsible for many of the observed GWA failures, for example, the absence of a satellite at 6 eV binding energy in the photoemission spectrum of bulk nickel, which is believed to be due to hole–hole excitations.⁹⁹ The GWA also finds it difficult to describe multiple satellites, at least in the (non-self-consistent) G_0W_0 approximation. For example, the photoemission spectrum of bulk silicon displays a series of satellites that follows the quasi-particle valence band region between 0 and –12 eV. This satellite series is due to the presence of one, two, and more plasmon excitations. G_0W_0 , instead, produces only one satellite, further away from the quasi-particle than the first satellite in the experimental spectrum. This is illustrated in Figure 9 taken from Ref 100.

Why is this so? In principle it would be possible to express W^{tot} , and hence the self-energy, exactly in terms of W . The GW approximation limits the self-energy to the first-order term, where $W^{\text{tot}} \approx W$, and higher order terms are neglected. The correction to this problem can be expressed in two ways, which we will briefly discuss in the following section.

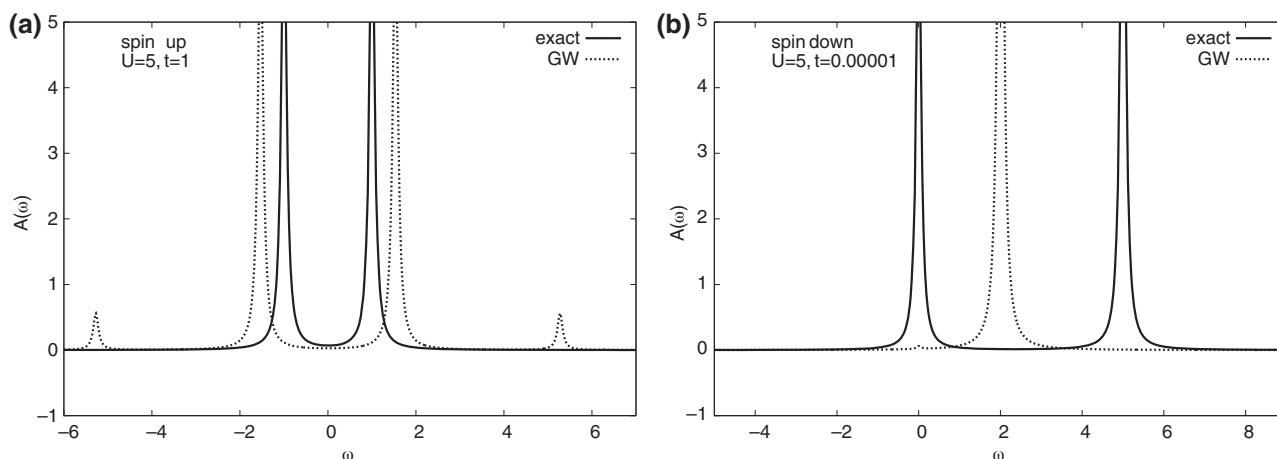


FIGURE 8 | Electron addition and removal spectral functions of the Hubbard dimer. There is one spin-up electron in the ground state. The exact result is given by the continuous line, the GWA result by the dotted line. Left: The spin-up spectral function for $U/t = 5$. The GWA suffers from self-screening in the electron removal and creates spurious satellites. Right: The spin-down spectral function for the atomic limit $t \rightarrow 0$. The GWA cannot reproduce the gap in this limit.

BEYOND THE GWA

Often the GWA is introduced as an approximation to a set of equations formalized by Hedin.²⁵ This in principle exact set of equations consists of the Dyson equation and a rewriting of Eqs (36) and (35), namely

$$\Sigma_{xc}(1,2) = iG(1,\bar{4})W(1^+, \bar{3})\tilde{\Gamma}(\bar{4}, 2; \bar{3}) \quad (50)$$

$$W(1,2) = v_c(1,2) + v_c(1, \bar{3})P(\bar{3}, \bar{4})W(\bar{4}, 2) \quad (51)$$

$$P(1,2) = -iG(1, \bar{3})G(\bar{4}, 1)\tilde{\Gamma}(\bar{3}, \bar{4}; 2) \quad (52)$$

$$\begin{aligned} \tilde{\Gamma}(1, 2; 3) = & \delta(1, 2)\delta(1, 3) + \frac{\delta\Sigma_{xc}(1, 2)}{\delta G(\bar{4}, \bar{5})} \\ & \times G(\bar{4}, \bar{6})G(\bar{7}, \bar{5})\tilde{\Gamma}(\bar{6}, \bar{7}; 3) \end{aligned} \quad (53)$$

$$G(12) = G_0(12) + G_0(1, \bar{3})\Sigma(\bar{3}, \bar{4})G(\bar{4}, 2) \quad (54)$$

This way of formulating the equations highlights some important ingredients:

1. the *irreducible polarizability* P in Eq. (52) is the response of the system to the total classical perturbation. It consists of pairs of particles (electrons or holes) that are interacting through the irreducible *vertex function* $\tilde{\Gamma}$.
2. The *effective classical interaction* W in Eq. (51) is the test-charge–test-charge screened interaction in Eq. (38).
3. The difference between W^{tot} and W is expressed by $\tilde{\Gamma}$ in Eq. (50). It stems from

variations of Σ_{xc} (see Eq. (53)). In other words, Hedin's equations highlight the GWA idea, and the vertex function $\tilde{\Gamma}$ that contains possible corrections to the GWA.

Many attempts have been made to find approximations for $\tilde{\Gamma}$, but deriving a reliable and still feasible expression has turned out to be exceedingly difficult. Note that Hedin's equations express a point of view: they focus on W (and therefore indirectly on the density-density response function χ), and the rest is presented as a correction. There are other ways to formulate the self-consistent set of relations, highlighting correlation functions other than χ , for example, hole–hole correlation. This leads in lowest order to the T-matrix approximation, rather than the GWA.⁴¹

In spite of all the good arguments for the use of Dyson equations, one may also have a doubt about the limits of usefulness of the self-energy concept itself. Let us first take the simple case where a matrix element of the Green's function has only one pole ϵ corresponding to a quasi-particle, and let us suppose that this pole is shifted with respect to the independent-particle result ϵ^0 , similar to Eqs (21) and (22). In this case, the self-energy is

$$\Sigma(\omega) = G_0^{-1}(\omega) - G^{-1}(\omega) = \epsilon - \epsilon^0 \quad (55)$$

a simple constant.

Let us now suppose that we have also a satellite due to the coupling to an excitation with energy ω_p . In order to conserve the total weight of the spectrum, in this case the Green's function reads

$$G(\omega) = \frac{a}{\omega - \varepsilon} + \frac{1-a}{\omega - \varepsilon + \omega_p} \quad (56)$$

and the self-energy becomes

$$\Sigma(\omega) = G_0^{-1}(\omega) - G^{-1}(\omega) = \varepsilon - \varepsilon^0 - (1-a)\omega_p + \frac{a(1-a)\omega_p^2}{\omega - \varepsilon + a\omega_p} \quad (57)$$

Now the ω -independent contribution to the self-energy is renormalized by the coupling to the excitation. Moreover, the self-energy is frequency-dependent, with a pole at $\omega = \varepsilon - a\omega_p$. The GW self-energy Eq. (43) has a similar structure, but there is a major difference: in the GWA, the pole of Σ_{xc} is at $\omega = \varepsilon - \omega_p$ instead of $\varepsilon - a\omega_p$. For this reason, and because $0 < a < 1$, satellites tend to be too far from the quasi-particle peak, even when ω_p is calculated beyond the RPA. It is difficult to find the necessary correction to the self-energy from physical intuition alone, and the task becomes increasingly difficult for multiple satellites.

One way to overcome this problem is to abandon the use of a self-energy, and move to another framework, such as that of cumulant expansions (see, e.g., Ref 29). One may roughly understand the cumulant idea by rewriting Eq. (32) as

$$G_u(1,1') = G_{Hu}(1,1') + iG_{Hu}(1,\bar{2}) \times W_u(\bar{2},\bar{3}) \frac{\delta G_u(\bar{2},1')}{\delta u_{cl}(\bar{3}^+)} \quad (58)$$

where W_u is the screened Coulomb interaction W as a functional of u (or equivalently, u_{cl}). Here $u_{cl} = u_{ext} + u + v_{Hu}$ is the total classical potential including the original and the additional external potentials, u_{ext} and u , and $G_{Hu} = G_0 + G_0 u_{cl} G_{Hu}$ is the Hartree Green's function in presence of u . The Eq. (58) is exact. To obtain the GWA, one would approximate $(\delta G_u / \delta u_{cl}) \approx GG$ and then set $u \rightarrow 0$ (this is an alternative way to derive the GWA). A different family of approximations is obtained by neglecting in Eq. (58), the dependence of W_u on the external perturbing potential, which means, by making the approximation of linear response. Then the differential equation can be solved approximately (see, e.g., Ref 101). The result is a Green's function of the form $G_{ii}(\tau) = G_{H,ii}(\tau) e^{C_i(\tau)}$, where $\tau = t_1 - t_2$. The function C in the exponent is called cumulant. In this approximation it depends linearly on the GW self-energy (and hence on W), and the expansion of the exponential yields single, double, and multiple excitations from a single

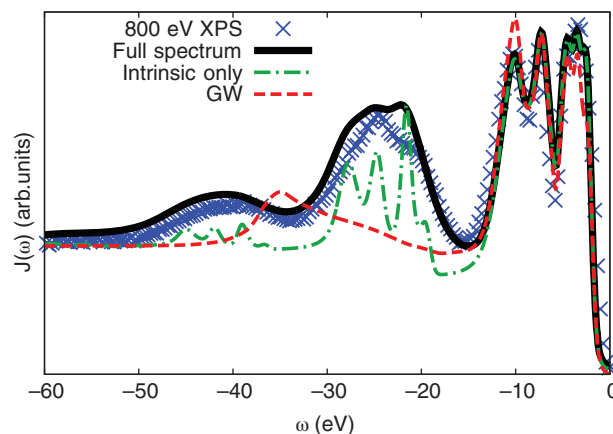


FIGURE 9 | Photo emission spectrum of bulk silicon, from Ref 100. Crosses are experimental results, the dashed line is from a G_0W_0 calculation. The thick continuous line is the intrinsic spectral function from a cumulant expansion calculation. The dot-dashed line includes moreover extrinsic losses of the outgoing photoelectron and interference effects. Cross sections and a secondary electrons background have been added to all calculated spectral functions (Reproduced with permission from Ref 100. Copyright 2011 American Physical Society).

excitation in W (see, e.g., Ref 102). In this respect, the approach bares a strong similarity to the coupled cluster method.¹⁰³ When the excitations that are contained in W , for example plasmons, are responsible for the observed satellites, the cumulant approximation leads to strong improvement of satellite spectra with respect to the GWA. This can be seen at the example of bulk silicon in Figure 1. Instead, in this simple form it still cannot, as a matter of principle, solve the problem of satellites due to other excitations, such as magnons, which are not contained in W .

CONCLUSIONS

In first principles calculations, electron addition and removal energies for finite and extended systems are often determined in the framework of density functional theory, by calculating KS eigenvalues. However, such an approach has two problems: first, these eigenvalues are not meant to be electron addition or removal energies. Second, one has to approximate the in general unknown exchange–correlation potential of the KS Hamiltonian. Altogether, this often introduces significant errors in ionization energies and electron affinities, and it leads to an underestimate of band gaps in solids which is often called the ‘band gap problem.’

To overcome these issues, today the state-of-the-art for calculations of the band structure of

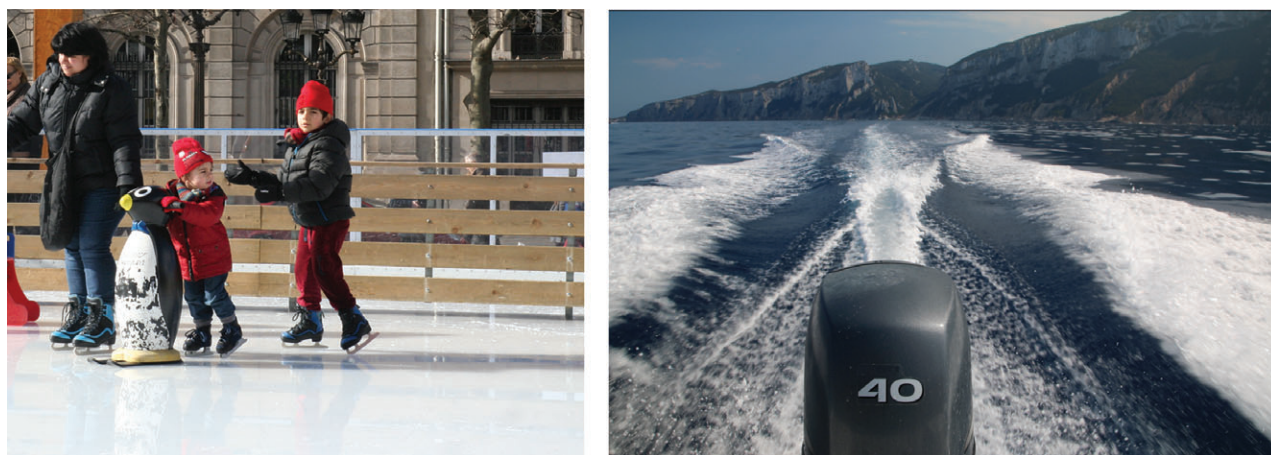


FIGURE 10 | Pictorial representation of the Hartree–Fock (left panel) and GW (right panel) approximations. Hartree–Fock eigenvalues reflects electron addition or removal with respect to a rigid system, in the GW approximation the system responds dynamically. (Credits to Andrea Cucca, 2017).

extended systems is the use of Green's functions and many-body perturbation theory. With respect to KS, the exchange–correlation potential is replaced by a nonlocal and frequency-dependent self-energy. The Hartree plus exchange potentials of HF constitute the first-order, frequency-independent, approximation to the self-energy. To go beyond, the most widely used approximation in extended systems is the GW approximation; with respect to HF, the bare Coulomb interaction in the exchange term is replaced by a dynamically screened interaction. This reduces the way too large HF eigenvalue gap in solids and brings it in general close to experiment.

The GWA is used to calculate band structures, lifetime broadening, spectral functions, or total energies, for a wide range of extended systems. More recently, it has also been used in finite systems, and one can find an increasing number of implementations and benchmarks. Results are quite promising, although more experience is needed to elucidate questions such as the optimal level of self-consistency that is needed to obtain the most reliable results.

Successes and limitations of the GWA can be understood from a picture where an electron that propagates in a system is represented by an object traveling on water (see Figure 10): in HF, the water is frozen and the electron is an ice-skater: there is no relaxation, which explains the too high energies. In the GWA, the electron acts like a boat on water, creating waves (as it couples to other excitations in the system) that act back on its propagation. It should be noted, however, that the response of the ‘water’ is a mean-field response: in the GWA, the additional particle does not see the individual water molecules, in other words, it does not correlate with the individual

system particles. In situations of low density such as in the dissociation limit of certain molecules, including H_2 and H_2^+ , this becomes important and is a limitation of the GWA.

Overall, the GWA constitutes a good compromise between precision, range of validity and computational efficiency. Moreover, it yields a convenient starting point for advanced calculations of response functions, for example, for the determination of excitonic effects in optical spectra. It is therefore increasingly used for applications that may be interesting in physics, chemistry, or even biology, such as the calculation of energy levels of molecules in water, or spectra of nanostructured materials for photovoltaics. In cases where processes are important that are not described by the GWA, it can still be used as a valuable ingredient for methods that go beyond, thanks to its clearly defined content in terms of exchange and dynamical screening.

NOTATION AND FOURIER TRANSFORMS

In order to ease the reading of this article, in the following we summarize the most important notations and the convention for Fourier transforms:

$$t^+ \equiv \lim_{\eta \rightarrow 0^+} t + \eta \quad (59)$$

$$t^{++} \equiv \lim_{\eta \rightarrow 0^+} t + 2\eta \quad (60)$$

$$\cdot(1) \equiv (x_1, t_1) \equiv (\mathbf{r}_1, \sigma_1, t_1) \quad (61)$$

$$f(\overline{1})g(\overline{1}) \equiv \int d1 f(1)g(1) \quad (62)$$

$$G(\omega) = \int_{-\infty}^{+\infty} d\tau e^{i\omega\tau} G(\tau) \quad (63)$$

$$G(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} d\omega e^{-i\omega\tau} G(\omega) \quad (64)$$

The time ordering of operators is defined as

$$T[\hat{A}(t_1)\hat{B}(t_2)] \equiv \begin{cases} \hat{A}(t_1)\hat{B}(t_2), & t_1 > t_2 \\ \hat{B}(t_2)\hat{A}(t_1), & t_1 < t_2 \end{cases} \quad (65)$$

NOTES

^a Note that throughout this article we will use *atomic units*, where $\hbar = m_e = e = 4\pi/\epsilon_0 = 1$.

^b In the following, we will sometimes refer to spin-unpolarized systems with spin-independent interactions. In that case spin appears only as a prefactor, for example the density is $n(\mathbf{r}) = 2\langle\Psi|\hat{\psi}^\dagger(\mathbf{r},t)\hat{\psi}(\mathbf{r},t)|\Psi\rangle$.

^c Note that the eigenvalue of the highest occupied molecular orbital is in principle a removal energy, but not the other KS eigenvalues.

^d In a finite systems, this can also happen when there is coupling to some continuum or very dense excitation spectrum, which can stem, for example, from the environment.

^e Usually plasmons are measured using an external probe, such as X-rays in inelastic X-ray scattering, or a beam of fast electrons in electron energy loss spectroscopy. In the case discussed here, the propagation of a system-internal electron or hole excites the plasmons.

^f For the sake of clarity of the comparison, here we use the spin-resolved version.

^g This statement refers to the fact that the wavefunction is a single Slater determinant, and to the HF equations. This does not preclude the fact that correlation can also be simulated within HF, by letting a system relax in a Δ -SCF approach, as it is explained in *Screening in Finite and Infinite Systems*.

^h Note that a response function is a retarded quantity, whereas here we work with time-ordered objects. However, retarded and time-ordered quantities can be linked in a straightforward way, and we do not make an explicit distinction here.

ⁱ Note that in order to fully represent L , the applied external potential must be time-dependent. There are subtleties for the use of Green's function s in presence of a time-dependent external potential, which means that the system

is driven out of equilibrium. A formulation such as the one exposed here remains rigorous also out of equilibrium if the concept of a time-contour is introduced. We will not deepen this question here and refer the reader to the modern books (Refs 1 and 22).

^j Note that the X-rays or fast electrons can be considered to be classical probes, which explains why these experiments can be described by the screening that is due to the Hartree potential only.

^k There is, however, self-screening in the GWA, because *variations* of the Hartree potential are not combined with the corresponding variations of the exchange self-energy, as can be seen in Eq. (39).

^l There is also an infinitesimal imaginary part included in ω_λ , similar to Eq. (16), but we do not display this here and in the following in order to keep the equations simpler.

^m A similar expression holds when G is the interacting Green's function. In that case the independent-particle orbitals are replaced by Dyson orbitals, and the independent-particle eigenvalues by total energy differences, $\epsilon_i = E_{N+1,i} - E_N$ for electron addition and $\epsilon_i = E_N - E_{N-1,i}$ for electron removal.

ⁿ Today, in practical applications better mean-field approximations are used, but here we use HF as a reference for the discussion.

^o Note that the term RPA is sometimes extended to include the variation of the Fock exchange, which is also called Random-Phase Approximation with Exchange, but our definition does not contain this contribution.

^p The self-energy in Figure 2 is an illustration for a case where the linear approximation is reasonable.

^q Though not always: the GWA cannot create a Mott gap, which is purely due to correlation.

^r Note that this continues the discussion at the end of *Screening in Finite and Infinite Systems*: as it is stated there, the GWA 'understands' that the squared wavefunction of the additional electron is not a classical charge density. However, it still treats the system to which the charge is added as a (polarizable) mean field. There is no explicit correlation between the extra charge and the system particles.

^s This is consistent with the fact that the many-body states are superpositions of many excited Slater determinants, which is explained in *Spectral Functions*.

ACKNOWLEDGMENTS

Stimulating discussions with many members of the Palaiseau Theoretical Spectroscopy Group and with Andreas Savin and Evert Jan Baerends are gratefully acknowledged. Financial support has been provided by the European Research Council under the EU FP7 framework program (ERC grant No. 320971).

REFERENCES

1. Stefanucci G, van Leeuwen R. *Many-Body Theory of Quantum Systems*. Cambridge: Cambridge University Press; 2013.
2. Born M, Oppenheimer JR. Zur Quantentheorie der Molekeln. *Ann Phys* 1927, 84:457.
3. Kohn W. Nobel lecture: electronic structure of matter—wave functions and density functionals. *Rev Mod Phys* 1999, 71:1253–1266.
4. Fetter AL, Walecka JD. *Quantum Theory of Many-Particle Systems*. New York, NY: McGraw-Hill; 1971.
5. Kohn W, Sham LJ. Self-consistent equations including exchange and correlation effects. *Phys Rev* 1965, 140:A1133–A1138.
6. Holleboom LJ, Snijders JG, Baerends EJ. Natural energy orbitals and the one-particle Green's function. *Int J Quant Chem* 1988, 34:289–300.
7. Baerends EJ, Gritsenko OV, van Meer R. The Kohn–Sham gap, the fundamental gap and the optical gap: the physical meaning of occupied and virtual Kohn–Sham orbital energies. *Phys Chem Chem Phys* 2013, 15:16408–16425.
8. Ohkoshi I. An ab initio Hartree-Fock pseudopotential approach to the ground-state properties of Si. *J Phys C-Solid State Phys Ther* 1985, 18:5415.
9. Godby RW, Schlüter M, Sham LJ. Self-energy operators and exchange-correlation potentials in semiconductors. *Phys Rev B* 1988, 37:10159–10175.
10. Galitskii VM, Migdal AB. Application of quantum field theory methods to the many body problem. *Soviet Phys JETP* 1958, 7:96.
11. Linderberg J, Öhrn Y. *Propagators in Quantum Chemistry*. Hoboken, NJ: John Wiley and Sons; 2004.
12. Ortiz JV. *Computational Chemistry: Reviews of Current Trends*, vol. 2. Singapore: World Scientific; 1997, 1–61.
13. Cederbaum LS. Green's Functions and Propagators for Chemistry. In: *Encyclopedia of Computational Chemistry*. Chichester: John Wiley and Sons Ltd; 2002.
14. Cederbaum LS. On green's functions and their applications. *Int J Quant Chem* 1990, 38:393–404.
15. Cederbaum LS. Green's functions for molecules, in few-body problems in physics. In: *Proceedings of the XIIIth European Conference on Few-Body Physics*, Marciana Marina, Isola d'Elba, Italy, September 9–14, 1991. Vienna: Springer; 1992, 595–604.
16. Schirmer J, Cederbaum LS, Walter O. New approach to the one-particle Green's function for finite Fermi systems. *Phys Rev A* 1983, 28:1237–1259.
17. Holleboom L, Snijders J, Baerends E, Buijse M. A correlation potential for molecular systems from the single particle Green's function. *J Chem Phys* 1988, 89:3638–3653.
18. Ortiz JV. Electron propagator theory: an approach to prediction and interpretation in quantum chemistry. *WIREs Comput Mol Sci* 2013, 3:123.
19. Perdew JP, Zunger A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys Rev B* 1981, 23:5048–5079.
20. Mattuck RD. *A Guide to Feynman Diagrams in the Many-Body Problem*. Dover Books on Physics. Mineola, NY: Dover Publications; 1992.
21. Strinati G. Application of the Green's functions method to the study of the optical properties of semiconductors. *Riv Nuovo Cimento* 1988, 11:1.
22. Martin RM, Reining L, Ceperley DM. *Interacting Electrons: Theory and Computational Approaches*. Cambridge: Cambridge University Press; 2016.
23. Berger JA, Romaniello P, Tandetzky F, Mendoza BS, Brouder C, Reining L. Solution to the many-body problem in one point. *New J Phys* 2014, 16:113025.
24. Casida ME, Chong DP. Physical interpretation and assessment of the Coulomb-hole and screened-exchange approximation for molecules. *Phys Rev A* 1989, 40:4837–4848.
25. Hedin L. New method for calculating the one-particle Green's function with application to the electron-gas problem. *Phys Rev* 1965, 139:A796–A823.
26. Seidl A, Görling A, Vogl P, Majewski JA, Levy M. Generalized Kohn-Sham schemes and the band-gap problem. *Phys Rev B* 1996, 53:3764–3774.
27. Bylander DM, Kleinman L. Good semiconductor band gaps with a modified local-density approximation. *Phys Rev B* 1990, 41:7868–7871.
28. Wigner EP. Effects of the electron interaction on the energy levels of electrons in metals. *Trans Faraday Soc* 1938, 34:678.
29. Hedin L. On correlation effects in electron spectroscopies and the GW approximation. *J Phys C* 1999, 11:R489.
30. Hedin L, Lundqvist S. Effects of electron-electron and electron-phonon interactions on the one-electron states of solids. *Solid State Phys* 1969, 23:1.
31. Vidal J, Botti S, Olsson P, Guillemales J-F, Reining L. Strong interplay between structure and electronic properties in CuIn(S,Se)2: a first-principles study. *Phys Rev Lett* 2010, 104:056401.
32. Stoll H, Savin A. *Density Functional Methods in Physics*. New York, NY: Plenum; 1985.
33. Heyd J, Scuseria GE, Ernzerhof M. Hybrid functionals based on a screened Coulomb potential. *J Chem Phys* 2003, 118:8207–8215.
34. Krukau AV, Vydrov OA, Izmaylov AF, Scuseria GE. Influence of the exchange screening parameter on the

- performance of screened hybrid functionals. *J Chem Phys* 2006, 125:224106.
35. Yanai T, Tew DP, Handy NC. A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem Phys Lett* 2004, 393:51–57.
36. Vydrov OA, Scuseria GE. Assessment of a long-range corrected hybrid functional. *J Chem Phys* 2006, 125:234109.
37. Vydrov OA, Heyd J, Krukau AV, Scuseria GE. Importance of short-range versus long-range Hartree-Fock exchange for the performance of hybrid density functionals. *J Chem Phys* 2006, 125:074106.
38. Almladh CO, Von Barth U, Van Leeuwen R. 9th international conference on recent progress in many-body theories. *Int J Mod Phys B* 1999, 13:535.
39. Dahlen NE, van Leeuwen R, von Barth U. Variational energy functionals of the green function and of the density tested on molecules. *Phys Rev A* 2006, 73:012511.
40. Agnihotri MP, Apel W, Weller W. The Luttinger-Ward method applied to the 2D Coulomb gas. *Phys Stat Sol B* 2008, 245:421–427.
41. van Leeuwen R, Dahlen NE, Stan A. Total energies from variational functionals of the green function and the renormalized four-point vertex. *Phys Rev B* 2006, 74:195105.
42. Dahlen N, Barth U v. Variational energy functionals tested on atoms. *Phys Rev B* 2004, 69:195102.
43. Langreth DC, Perdew JP. Exchange-correlation energy of a metallic surface: wave-vector analysis. *Phys Rev B* 1977, 15:2884–2901.
44. Furche F. Molecular tests of the random phase approximation to the exchange-correlation energy functional. *Phys Rev B* 2001, 64:195120.
45. Fuchs M, Gonze X. Accurate density functionals: approaches using the adiabatic-connection fluctuation-dissipation theorem. *Phys Rev B* 2002, 65:235109.
46. Paier J, Ren X, Rinke P, Scuseria GE, Grueneis A, Kresse G, Scheffler M. Assessment of correlation energies based on the random-phase approximation. *New J Phys* 2012, 14:043002.
47. Aryasetiawan F, Miyake T, Terakura K. Total energy method from many-body formulation. *Phys Rev Lett* 2002, 88:166401.
48. Miyake T, Aryasetiawan F, Kotani T, van Schilfgaarde M, Usuda M, Terakura K. Total energy of solids: an exchange and random-phase approximation correlation study. *Phys Rev B* 2002, 66:245103.
49. Fuchs M, Burke K, Niquet Y-M, Gonze X. Comment on “Total energy method from many-body formulation”. *Phys Rev Lett* 2003, 90:189701.
50. Aryasetiawan F, Miyake T, Terakura K. Aryasetiawan, Miyake, and Terakura Reply. *Phys Rev Lett* 2003, 90:189702.
51. Bruneval F. GW approximation of the many-body problem and changes in the particle number. *Phys Rev Lett* 2009, 103:176403.
52. Frey K, Idrobo JC, Tiago ML, Reboredo F, Ögüt S. Quasiparticle gaps and exciton Coulomb energies in Si nanoshells: first-principles calculations. *Phys Rev B* 2009, 80:153411.
53. Massobrio C, Pasquarello A, Car R. First principles study of photoelectron spectra of Cu_n^- clusters. *Phys Rev Lett* 1995, 75:2104–2107.
54. Bohm D, Pines D. A collective description of electron interactions: III. Coulomb interactions in a degenerate electron gas. *Phys Rev* 1953, 92:609–625.
55. Lindhard J. Electron capture and loss by heavy ions penetrating through matter. *Kgl Danske Videnskab Selskab, Mat-Fys Medd* 1954, 28:1.
56. Rieger MM, Steinbeck L, White ID, Rojas HN, Godby RW. The GW space-time method for the self-energy of large systems. *Comput Phys Commun* 1999, 117:211–228.
57. Rojas HN, Godby RW, Needs RJ. Space-time method for ab initio calculations of self-energies and dielectric response functions of solids. *Phys Rev Lett* 1995, 74:1827–1830.
58. Aryasetiawan F. Self-energy of ferromagnetic nickel in the GW approximation. *Phys Rev B* 1992, 46:13051–13064.
59. Lebègue S, Arnaud B, Alouani M, Bloechl PE. Implementation of an all-electron GW approximation based on the projector augmented wave method without plasmon pole approximation: application to Si, SiC, AlAs, InAs, NaH, and KH. *Phys Rev B* 2003, 67:155208.
60. Hybertsen MS, Louie SG. Electron correlation in semiconductors and insulators: band gaps and quasiparticle energies. *Phys Rev B* 1986, 34:5390–5413.
61. Godby RW, Needs RJ. Metal-insulator transition in Kohn-Sham theory and quasiparticle theory. *Phys Rev Lett* 1989, 62:1169–1172.
62. Soininen JA, Rehr JJ, Shirley EL. Electron self-energy calculation using a general multi-pole approximation. *J Phys Condense Matter* 2003, 15:2573–2586.
63. Soininen JA, Rehr JJ, Shirley EL. Final-state rule vs the Bethe-Salpeter equation for deep-core x-ray absorption spectra. *Phys Scr* 2005, T115:243.
64. van Setten MJ, Weigend F, Evers F. The GW-method for quantum chemistry applications: theory and implementation. *J. Chem Theory Comput* 2013, 9:232–246.
65. van Schilfgaarde M, Kotani T, Faleev S. Quasiparticle self-consistent GW theory. *Phys Rev Lett* 2006, 96:226402.

66. Faleev SV, van Schilfgaarde M, Kotani T. All-electron self-consistent GW approximation: application to Si, MnO, and NiO. *Phys Rev Lett* 2004, 93:126406.
67. Kotani T, van Schilfgaarde M, Faleev SV. Quasiparticle self-consistent GW method: a basis for the independent-particle approximation. *Phys Rev B* 2007, 76:165106.
68. Holm B, von Barth U. Fully self-consistent GW self-energy of the electron gas. *Phys Rev B* 1998, 57:2108–2117.
69. Schindlmayr A, García-González P, Godby RW. Diagrammatic self-energy approximations and the total particle number. *Phys Rev B* 2001, 64:235106.
70. García-González P, Godby RW. Self-consistent calculation of total energies of the electron gas using many-body perturbation theory. *Phys Rev B* 2001, 63:075112.
71. Rieger MM, Godby RW. Charge density of semiconductors in the GW approximation. *Phys Rev B* 1998, 58:1343–1348.
72. Stan A, Dahlen NE, van Leeuwen R. Levels of self-consistency in the GW approximation. *J Chem Phys* 2009, 130:114105.
73. Schindlmayr A. Violation of particle number conservation in the GW approximation. *Phys Rev B* 1997, 56:3528–3531.
74. Strinati G, Mattausch HJ, Hanke W. Dynamical correlation effects on the quasiparticle Bloch states of a covalent crystal. *Phys Rev Lett* 1980, 45:290–294.
75. Strinati G, Mattausch HJ, Hanke W. Dynamical aspects of correlation corrections in a covalent crystal. *Phys Rev B* 1982, 25:2867–2888.
76. Hybertsen MS, Louie SG. First-principles theory of quasiparticles: calculation of band gaps in semiconductors and insulators. *Phys Rev Lett* 1985, 55:1418–1421.
77. Aulbur WG, Jonsson L, Wilkins JW. Quasiparticle calculations in solids. *Solid State Phys* 2000, 54:1.
78. Aryasetiawan F, Gunnarsson O. The GW method. *Rep Prog Phys* 1998, 61:237–312.
79. Gatti M, Bruneval F, Olevano V, Reining L. Understanding correlations in vanadium dioxide from first principles. *Phys Rev Lett* 2007, 99:266402.
80. Rödl C, Fuchs F, Furthmüller J, Bechstedt F. Quasiparticle band structures of the antiferromagnetic transition-metal oxides MnO, FeO, CoO, and NiO. *Phys Rev B* 2009, 79:235114.
81. Chulkov EV, Borisov AG, Gauyacq JP, Sanchez-Portal D, Silkin VM, Zhukov VP, Echenique PM. Electronic excitations in metals and at metal surfaces. *Chem Rev* 2006, 106:4160–4206.
82. Rinke P, Delaney K, Garcia-Gonzalez P, Godby RW. Image states in metal clusters. *Phys Rev A* 2004, 70:063201.
83. Koerbel S, Boulanger P, Duchemin I, Blase X, Marques MAL, Botti S. Benchmark many-body GW and Bethe–Salpeter calculations for small transition metal molecules. *J Chem Theory Comput* 2014, 10:3934–3943.
84. Kaplan F, Harding ME, Seiler C, Weigend F, Evers F, van Setten MJ. Quasi-particle self-consistent GW for molecules. *J Chem Theory Comput* 2016, 12:2528–2541.
85. Wilhelm J, Del Ben M, Hutter J. GW in the Gaussian and plane waves scheme with application to linear acenes. *J Chem Theory Comput* 2016, 12:3623–3635.
86. Scherpelz P, Govoni M, Hamada I, Galli G. Implementation and validation of fully relativistic GW calculations: spin–orbit coupling in molecules, nanocrystals, and solids. *J Chem Theory Comput* 2016, 12:3523–3544.
87. Faber C, Attaccalite C, Olevano V, Runge E, Blase X. First-principles GW calculations for DNA and RNA nucleobases. *Phys Rev B* 2011, 83:115123.
88. Blase X, Attaccalite C, Olevano V. First-principles GW calculations for fullerenes, porphyrins, phthalocyanine, and other molecules of interest for organic photovoltaic applications. *Phys Rev B* 2011, 83:115103.
89. van Setten MJ, Caruso F, Sharifzadeh S, Ren X, Scheffler M, Liu F, Lischner J, Lin L, Deslippe JR, Louie SG, et al. GW100: Benchmarking G0W0 for molecular systems. *J Chem Theory Comput* 2015, 11:5665–5687.
90. Rangel T, Hamed SM, Bruneval F, Neaton JB. Evaluating the GW approximation with CCSD(T) for charged excitations across the oligoacenes. *J Chem Theory Comput* 2016, 12:2834–2842.
91. Bruneval F, Marques MAL. Benchmarking the starting points of the GW approximation for molecules. *J Chem Theory Comput* 2013, 9:324–329.
92. Marom N, Caruso F, Ren X, Hofmann OT, Körzdörfer T, Chelikowsky JR, Rubio A, Scheffler M, Rinke P. Benchmark of GW methods for azabenzenes. *Phys Rev B* 2012, 86:245127.
93. Blase X, Boulanger P, Bruneval F, Fernandez-Serra M, Duchemin I. GW and Bethe–Salpeter study of small water clusters. *J Chem Phys* 2016, 144:034109.
94. Caruso F, Atalla V, Ren X, Rubio A, Scheffler M, Rinke P. First-principles description of charge transfer in donor-acceptor compounds from self-consistent many-body perturbation theory. *Phys Rev B* 2014, 90:085141.
95. Caruso F, Rohr DR, Hellgren M, Ren X, Rinke P, Rubio A, Scheffler M. Bond breaking and bond formation: how electron correlation is captured in many-body perturbation theory and density-functional theory. *Phys Rev Lett* 2013, 110:146403.

96. Hellgren M, Caruso F, Rohr DR, Ren X, Rubio A, Scheffler M, Rinke P. Static correlation and electron localization in molecular dimers from the self-consistent RPA and GW approximation. *Phys Rev B* 2015, 91:165110.
97. Romaniello P, Guyot S, Reining L. The self-energy beyond GW: local and nonlocal vertex corrections. *J Chem Phys* 2009, 131:154111.
98. Romaniello P, Bechstedt F, Reining L. Beyond the GW approximation: combining correlation channels. *Phys Rev B* 2012, 85:155131.
99. Springer M, Aryasetiawan F, Karlsson K. First-Principles T-matrix theory with application to the 6 eV satellite in Ni. *Phys Rev Lett* 1998, 80:2389–2392.
100. Guzzo M, Lani G, Sottile F, Romaniello P, Gatti M, Kas JJ, Rehr JJ, Silly MG, Sirotti F, Reining L. Valence electron photoemission spectrum of semiconductors: ab initio description of multiple satellites. *Phys Rev Lett* 2011, 107:166401.
101. Zhou JS, Kas JJ, Sponza L, Reshetnyak I, Guzzo M, Giorgetti C, Gatti M, Sottile F, Rehr JJ, Reining L. Dynamical effects in electron spectroscopy. *J Chem Phys* 2015, 143:184109.
102. Aryasetiawan F, Hedin L, Karlsson K. Multiple plasmon satellites in Na and Al spectral functions from ab initio cumulant expansion. *Phys Rev Lett* 1996, 77:2268–2271.
103. Bartlett RJ. Many-body perturbation theory and coupled cluster theory for electron correlation in molecules. *Ann Rev Phys Chem* 1981, 32:359–401.