Lecture notes on the GW method (draft)

Liam Damewood

May 11, 2012

Abstract

Here are some brief lecture notes on the GW method based on various sources but mainly Solid State Phys. 54, 1 (2000). Most figures are taken from there. Major errors should be reported to *damewood at physics dot ucdavis dot edu*—I'm sure there are quite a few.

1 Introduction

Density functional theory (DFT) provides a method for determining ground state and, in principle, excited state properties of matter using the simple ground state density $n_0(\vec{r})$ as a variational quantity. Unfortunately, the Hoenberg-Kohn (HK) theorem (ref), the basis for DFT, does not provide a mathematical framework to determine the exact wavefunction as a functional of the density. The Kohn-Sham (KS) equations provide a method to obtain independent single particle states(ref), however all but one of these orbitals (the highest occupied state) has physical meaning (ref).

The KS orbitals bear little significance when trying to determine excited state properties. The fundamental gap, for example, is underestimated using the local density approximation (LDA) by roughly 10%. Silicon has a direct bandgap at the Γ-point of 3.4 eV (ref); the LDA calculates a value of around 2.5 eV (ref). Similarly, LDA density of states is not correct. The bandwidth of Al is calculated as 11.06 eV (ref) using the LDA while the experimental value is 10.6 eV (ref).

One method of correcting the excited state properties is to consider a many-body perturbative (MBP) approach. To motivate the theory, I will consider the angle resolved photoemission experiment with some semiconducting sample (figure?) containing N electrons. In a photoemission experiment, an incoming photon with energy $h\nu$ and momentum $\hbar\vec{k}$ ejects an electron from the valence band with initial momentum $\hbar\vec{k}_i$ leaving it with kinetic energy $\hbar k_f^2/2m$ and momentum $\hbar\vec{k}_f$. This information can tell us about the dispersion relation for the valence band, but also leaves the system with one fewer electron. In the inverse photoemission experiment, we can strike the sample with an electron with certain momentum and kinetic energy. The electron can then occupy one of the conduction band states and eject a photon with measurable momentum and kinetic energy. Conservation of energy and momentum can give us information about the conduction band momentum and energy. This will leave the system with one additional electron.

2 Theory

The photoemission experiments leave us with information about the top of the valence band and the excited states including the fundamental gap energy E_g. The photoemission experiment involves adding or removing an electron. When an electron is added to the system the N other electrons feel the bare coulomb potential ($v \sim 1/r$) and rearrange themselves to screen the charge. The rearrangement of the charges polarize the medium and create an oppositely charged "coulomb hole" (figure?). This collection of particles (collectively called a quasiparticle) interacts with other quasiparticles through the screened coulomb potential instead of the bare Coulomb potential. The quasiparticles are not quite eigenstates of the many body Hamiltonian, but have complex eigenvalues with the imaginary part inversely proportional to the quasiparticle lifetime.

Using the photoemission experiment as a guideline, we can determine the binding energy of a quasiparticle state by considering the difference between a N-particle ground state Ψ_0^N and a state within a system of

$N-1$ particles, Ψ_i^{N-1} such that $\epsilon_i = E_0^N - E_i^{N-1}$ when $\epsilon_i < \mu$, the chemical potential. Similarly, the inverse photoemission process adds an electron to the N-particle ground state, so $\epsilon_i = E_i^{N+1} - E_0^N$ when $\epsilon_i > \mu$. The nearly-independent quasiparticle states are then

$$\Psi_i(\vec{r}) = \left\langle N,0 \middle| \hat{\Psi}(\vec{r}) \middle| N+1,i \right\rangle, \epsilon_i = E_{N+1,i} - E_{N,0}, \text{ for } \epsilon_i \geq \mu$$
$$\Psi_i(\vec{r}) = \left\langle N-1,i \middle| \hat{\Psi}(\vec{r}) \middle| N,0 \right\rangle, \epsilon_i = E_{N,0} - E_{N-1,i}, \text{ for } \epsilon_i < \mu$$

For the remainder of the document, I will use the standard shorthand notation for the coordinates

$$f(12) = f(\vec{r}_1, t_1, \vec{r}_2, t_2)$$
$$\int \mathrm{d}1 = \int_{\mathbb{R}^3} \mathrm{d}\vec{r}_1 \int_{-\infty}^{\infty} \mathrm{d}t_1$$
$$f(1^+) = f(\vec{r}, t+\eta)$$

where η is some small positive number.

The quantity of central importance to many body theory is the Green's function

$$\mathrm{i}\hbar G(12) = \left\langle N,0 \middle| T[\hat{\Psi}(1)\hat{\Psi}^\dagger(2)] \middle| N,0 \right\rangle$$

which is the probability of a single particle at $\vec{r}_2$ at time t_2 to propagate to $\vec{r}_1$ at time t_1. The $T[]$ represents the time ordering operator which preserves causality. In many-body theory, we can use the Green's function to develop the quasiparticle Hamiltonian

$$\hat{H}\Psi_i(1) = -\frac{1}{2}\nabla_1^2\Psi_i(1) + v_{\text{ext}}(1)\Psi_i(1) + v_{\text{H}}(1)\Psi_i(1) + \int \Sigma(12)\Psi_i(2) = \epsilon_i\Psi_i(1)$$

where v_{H} is the Hartree potential, and Σ is called the self-energy. It represents the energy (beyond Hartree) required to arrange the system into the quasiparticle. This quantity is directly related to how the system responds to the addition or subtraction of a particle with a bare coulomb potential. In this sense, it is related to the response function and the screened coulomb potential. In order to solve for the quasiparticle wavefunctions, the goal is to find an appropriate approximation to Σ.

3 Screening within linear response

This is a very short overview of the definitions within linearresponse important for GW. In linear response, we consider the effect of a small external perturbation potential added to the Hamiltonian, $h(1) \to h(1) + u_{\text{ext}}(1)$ and then let that potential go to zero. This small potential will perturb the density out of the ground state and the linear response is

$$\chi(12) = \left.\frac{\delta n(1)}{\delta u_{\text{ext}}(2)}\right|_{u_{\text{ext}=0}}$$
$$\delta n(1) \equiv n(1) - n_0(1) = \int \mathrm{d}2\chi(12)u_{\text{ext}}(2)$$

where I have used the functional derivative with respect to the small external potential. The change in the density will contribute to the coulomb energy and I can add this to the external potential and find the effective potential

$$u_{\text{eff}}(1) = u_{\text{ext}}(1) + \int \mathrm{d}3v(13) \underbrace{\int \mathrm{d}2\chi(32)u_{\text{ext}}(2)}_{\delta n(3)}$$

This equation is analogous to the E&M equation $\vec{D} = \vec{E} + 4\pi\vec{P}$. We also can get a definition for the dielectric function ϵ which can be written in terms of the polarization (all functional derivatives are evaluated at $u_{\mathrm{ext}} = 0$)

$$
\begin{aligned}
\epsilon^{-1}(12) &\equiv \frac{\delta u_{\mathrm{eff}}(1)}{\delta u_{\mathrm{ext}}(1)} \\
&= \delta(12) + \int \mathrm{d}3 v(13)\chi(32) \\
P(12) &\equiv \frac{\delta n(1)}{\delta u_{\mathrm{eff}}(1)} \\
\chi(12) &= \frac{\delta n(1)}{\delta u_{\mathrm{ext}}(2)} \\
&= \int \mathrm{d}3 \frac{\delta n(1)}{\delta u_{\mathrm{eff}}(3)} \frac{\delta u_{\mathrm{eff}}(3)}{\delta u_{\mathrm{ext}}(2)} \qquad \text{(chain rule for functional derivatives)} \\
&= \int \mathrm{d}3 P(13)\epsilon^{-1}(32) \\
\epsilon^{-1}(12) &= \delta(12) + \int \mathrm{d}3 v(13)P(32) + \iiint \mathrm{d}3\,\mathrm{d}4\,\mathrm{d}5 v(13)P(34)v(45)P(52) + \cdots \\
\epsilon(12) &= \delta(12) - \int \mathrm{d}3 v(13)P(32)
\end{aligned}
$$

With all of these definitions, we can also write down the screened coulomb potential $W(12)$ which is the bare coulomb potential weighted by the dielectric function

$$
\begin{aligned}
W(12) &= \int \mathrm{d}3 \epsilon^{-1}(13)v(32) \\
&= v(12) + \int \mathrm{d}3\,\mathrm{d}4 v(13)P(34)W(42)
\end{aligned}
$$

All of these definitions should be slightly familiar from E&M although in a different (functional) form. For a similar overview, see the section on general screening in chapter 17 of Ashcroft and Mermin's Solid State Physics textbook. Figure 1 shows the resultant screened potential for Si at different locations.

4 Hedin's equations

Hedin developed the GW method by linking 5 many-body equations together into a self-consistent scheme.

Dyson's equation	$G(12) = G_0(12) + G_0(13)\Sigma(34)G(42)$	(1)
Screening equation	$W(12) = v(12) + v(13)P(34)W(42)$	(2)
Self-energy	$\Sigma(12) = \mathrm{i}\Gamma(4;1,3)G(3,2)W(4,2)$	(3)
Polarizability	$P(1,2) = -2\mathrm{i}\Gamma(1;4,5)G(2,4)G(5,2)$	(4)
Vertex Correction	$\Gamma(1;2,3) = \delta(12)\delta(13) + \Gamma(1;4,7)G(6,4)G(78)\frac{\delta\Sigma(23)}{\delta G(68)}$	(5)

The vertex correction arises from the two-particle Green's function associated with the Coulomb energy. It provides many-body corrections to the exchange and correlation energy. Near the Fermi surface, these

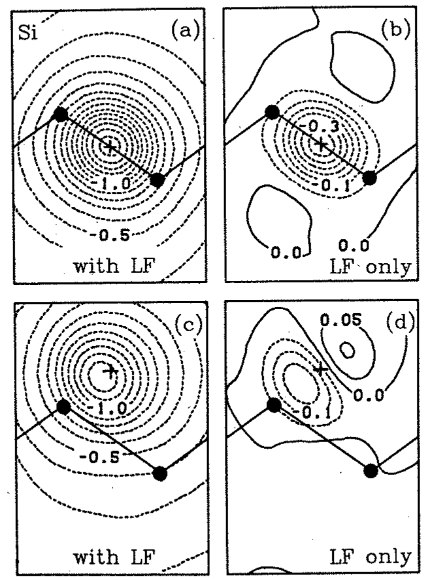
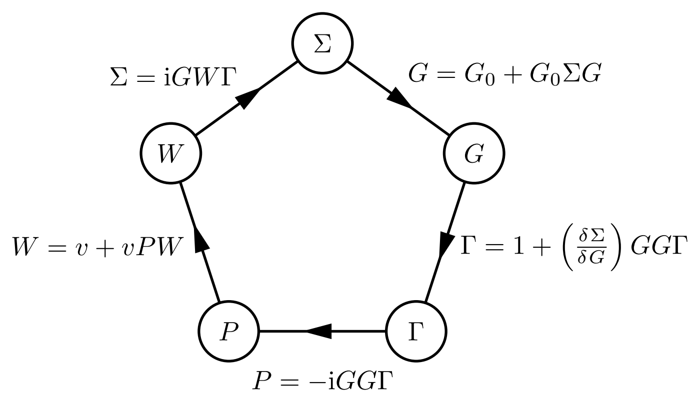
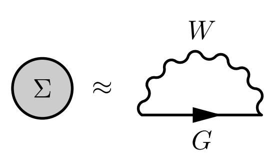


Figure 1: Screened potential contour in response to a $-1/r$ potential added at the x at different locations in Si (ref). The screened potential is a non-local quantity because it depends on two $\vec{r}$ variables. Local fields (LF) show the anisotropy of the screened coulomb potential.

corrections tend to cancel out (ref?). The Hedin equations are often summarized in the following diagram:



Since the vertex correction Γ is difficult to deal with numerically, and since the vertex corrections tend to cancel out, the GW approximation (GWA) is employed such that $\frac{\delta\Sigma}{\delta v_{\mathrm{ext}}} = 0$. Figure 2 shows the path of the GW approximation (GWA) where the complicated vertex terms are ignored. Figure 3 shows the "one-shot" G_0W_0 approximation which is the simplest form of the GWA. The GW approximation for the self energy is often written as the Feynman diagram



By construction, the GW method approximates the self-energy of the Hartree approximation, but it is also possible to start with any approximation (LDA, GGA, exact exchange, LDA+U, ...) by solving the

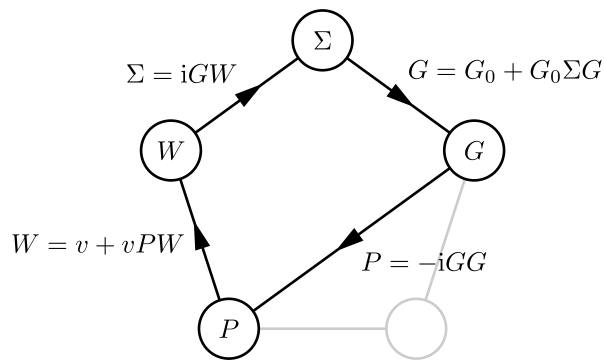


Figure 2: The GW approximation. We can ignore the complicated vertex corrections

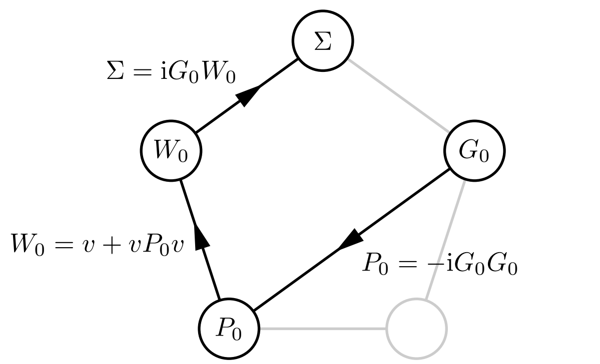


Figure 3: The G_0W_0 approximation. By starting with a "good" initial Green's function, one iteration can be "good enough".

Hamiltonian

$$\epsilon_i \Psi_i(1) = h(1)\Psi(1)_i + v_{\mathrm{Hartree}}(1)\Psi(1)_i + v_{XC}(1)\Psi(1)_i + \int \mathrm{d}2(\underbrace{\Sigma(2) - \delta(12)v_{XC}(2)}_{\Delta\Sigma(12)})\Psi(2)_i$$

Generally, the better the starting approximate Hamiltonian is, the better the G_0W_0 method is.

5 Results

See Figures 4, 5, 6 and 7

6 Other Topics

- COHSEX approximation: separate the coulomb hole (COH) from the screened exchange (SEX) in the calculation for Σ. COH is related to $W_{scr} = W - v$.
- Plasmon pole approximation (and other model dielectric functions): Use an analytic form of the dielectric function to perform the GW convolution.
- RPA susceptibility: No electron-hole interactions. Susceptibility only depends on one wavevector $\vec{q}$.

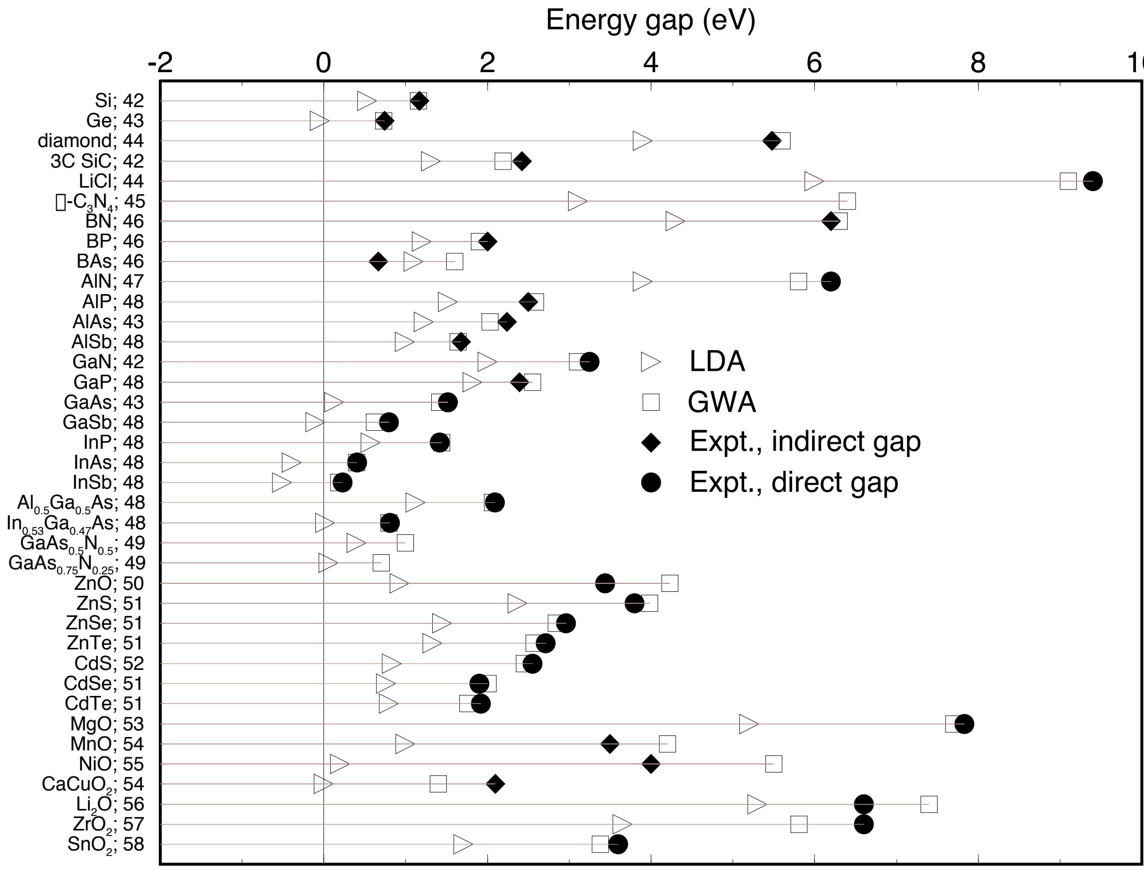


Figure 4: Comparison of characteristic direct and indirect LDA, GWA and experimental energy gaps (ref).

- Local field effects: short range polarization effects (microscopic fields) that occur in screening. Important for localized wavefunctions (bonding regions in insulators). Local fields can dramatically increase the band gap.
- Dynamic effects
- Magnetic GWA & Finite temperature GWA
- Half metallic screening (my research). Local fields play an different role in half metals.
- Core polarization effects: In general, core state cannot be ignored by a frozen core or pseudopotential approach.

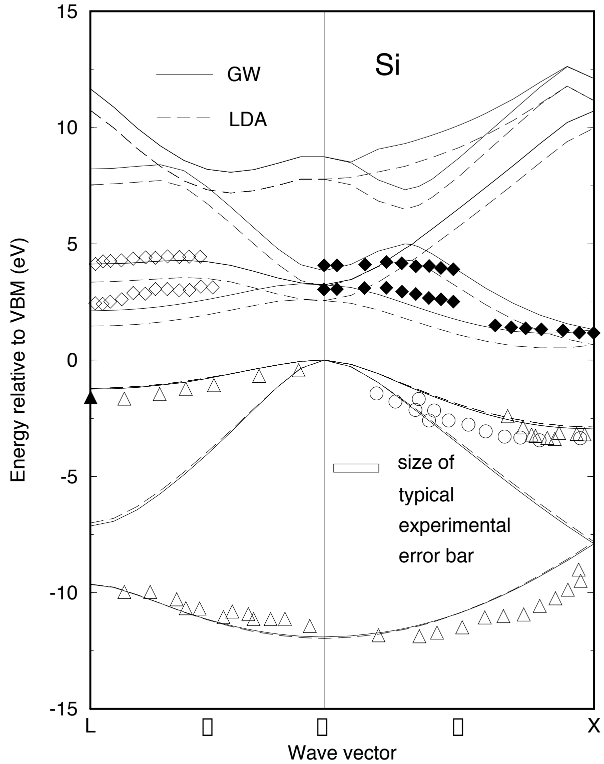


Figure 5: Comparison of LDA and GWA band structures along $L - \Gamma - X$ with photoemission and inverse photoemission experiments for Si (ref)

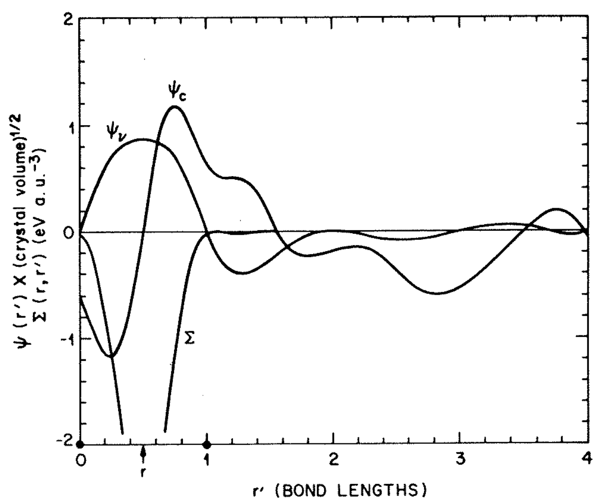


Figure 6: Plot of the self energy of Si where r is chosen to lie at the bond center and r' is varied along the [111] direction. Ψ_{vc} denote the real part of the wave functions (top of the valence, bottom of conduction) close to the Γ point calculated from the LDA. The range of nonlocality is about 1 bond length. It follows that $\langle \Psi_v \,|\, \Sigma \,|\, \Psi_v \rangle$ is large and negative while $\langle \Psi_c \,|\, \Sigma \,|\, \Psi_c \rangle$ has large positive and negative portions. The valence band gets pulled deeper in energy causing the LDA gap to open up.

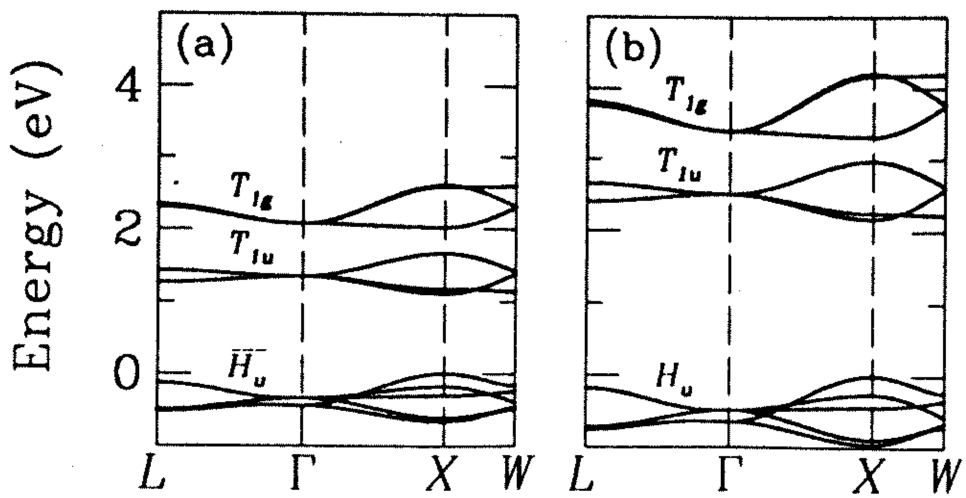


Figure 7: Band structure of the FCC Fm3 structure of solid C_{60} as obtained in LDA (a) and GWA (b) (ref)