Predictive nature of Self-consistent Quasiparticle Theory

Mark van Schilfgaarde, King's College London

The notion that connects these disparate efforts is that with the increasing sophistication of high performance computation, both hardware and software, the complex problem of building a first-principles understanding of materials will eventually be possible. The key intellectual challenge we wish to discuss is to identify tools that explain a sufficiently broad range of the rich spectrum of behaviors observed in complex materials to provide the impetus for moving the field beyond "explanation" to "prediction", a much harder task. Ultimately, this approach goes directly to the heart of emergent phenomena: to what extent can we, with our advanced computational tools and our experience with known emergent phenomena, predict new materials' properties? Discussing this among a group of the world's leading researchers in the field is the goal of this workshop.

Self-consistent Quasiparticle Based Theory of Defects

Mark van Schilfgaarde, King's College London

Quasiparticle Self-Consistent *GW* (QSGW)

✔ Advantages:

- A optimal means to design one-body hamiltonians
- An optimal way for many-body perturbation theory
- Dual character makes a potentially powerful tool to study electronic structure without ambiguities inherent in LDA+U, or hybrid 80%LDA + 20% HF, or LDA+GW, etc.

✘ Limitations:

- Standard implementation of QS*GW* expensive: N^4 scaling

How to surmount the scaling problem?

The Many-Body Wave Function

- The many-electron eigenfunction:

$$H\,\Psi(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N) = E\,\Psi(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N)$$

  contains $3N \sim 10^{23}$ degrees of freedom in a macroscopic solid.

- Easy if H separates $\Rightarrow$ Ψ would factor into a collection of solutions for independent electrons: i.e.

$$\Psi(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N) \rightarrow \psi_1(\mathbf{r}_1) \times \psi_2(\mathbf{r}_2) \times \psi_N(\mathbf{r}_N)$$

- Without this factorization, cannot isolate a single electron and trace its evolution even in principle. Example: excitons, Cooper pairs .
- Practically all of our intuitive understanding is based on the notion of independent particles. Very difficult to understand anything without the independent-particle concept.

Quasiparticles

- How to cast many-body problem into a collection of independent particles ? ($1/|\mathbf{r}-\mathbf{r}'|$ not factorizable)

$$\Psi(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N) \approx \psi_1(\mathbf{r}_1) \times \psi_2(\mathbf{r}_2) \times \psi_N(\mathbf{r}_N)$$

- Resolution: each e^- contributes some effective external field to the entire system.
- All e^- move in the presence of the effective field.
- Quasiparticles (Landau): a "particle," e.g. electron, really consists of a normal ("bare") electron + cloud of other "stuff."
- Quasiparticles behave as though they are nearly independent of each other. Residual interactions ⇒ quasiparticles decay after finite time. Lifetime cannot be too short if QP picture is to be meaningful.
- Q: How to formulate a theory for the effective field?

Density-Functional Theory

- W. Kohn proved (1964) that there exists an energy functional $E[n]$ of the electron density $n(\mathbf{r})$.
- A "deep" result: that $E=E[n]$ alone: nothing else in the vastly more complicated $\Psi(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N)$ is needed.
- Carry out in practice by solving effective SE with

$$V^{\text{eff}}(\mathbf{r}) = \frac{\delta}{\delta n(\mathbf{r})}\left[E_{\text{el-nuc}} + E_{\text{Hartree}} + E_{\text{xc}} \right]$$

Unknown : make ansatz for it

Key point: all electrons see same $V^{\text{xc}}(\mathbf{r})$ (Analog of Σ in DMFT)

Each electron should see a different $V^{\text{eff}}(\mathbf{r})$).

Hartree-Fock: nonlocal in space $V^{\text{x}}(\mathbf{r},\mathbf{r}')$

DMFT: nonlocal in time: $\Sigma(\omega)$

Locality brings: (1) advantage because theory really simple

(2) Cost because $V^{\text{eff}}(\mathbf{r})$ is fictitious $\Rightarrow$ ψ,ε are fictitious

- The 3 most cited papers, and 6 of the 10 most cited papers in the Physical Review series (Phys. Rev. B, Phys. Rev. Lett., Rev. Mod. Physics) all have to do with ab initio approaches to solving the Schrodinger equation for the electrons.
- Author of "Microsoft Version" of LDA code (Kresse, who wrote VASP) … has several papers with ~3000 citations

Table 1. *Physical Review* Articles with more than 1000 Citations Through June 2003

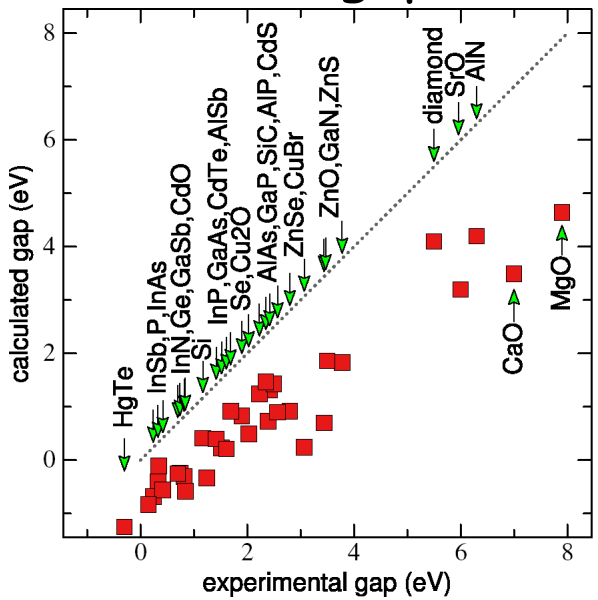
Publication	# cites	Av. age	Title	Author(s)	
PR **140**, A1133 (1965)	3227	26.7	Self-Consistent Equations Including Exchange and Correlation Effects	W. Kohn, L. J. Sham	←
PR **136**, B864 (1964)	2460	28.7	Inhomogeneous Electron Gas	P. Hohenberg, W. Kohn	←
PRB **23**, 5048 (1981)	2079	14.4	Self-Interaction Correction to Density-Functional Approximations for Many-Electron Systems	J. P. Perdew, A. Zunger	←
PRL **45**, 566 (1980)	1781	15.4	Ground State of the Electron Gas by a Stochastic Method	D. M. Ceperley, B. J. Alder	←
PR **108**, 1175 (1957)	1364	20.2	Theory of Superconductivity	J. Bardeen, L. N. Cooper, J. R. Schrieffer	
PRL **19**, 1264 (1967)	1306	15.5	A Model of Leptons	S. Weinberg	
PRB **12**, 3060 (1975)	1259	18.4	Linear Methods in Band Theory	O. K. Anderson	←
PR **124**, 1866 (1961)	1178	28.0	Effects of Configuration Interaction of Intensities and Phase Shifts	U. Fano	
RMP **57**, 287 (1985)	1055	9.2	Disordered Electronic Systems	P. A. Lee, T. V. Ramakrishnan	
RMP **54**, 437 (1982)	1045	10.8	Electronic Properties of Two-Dimensional Systems	T. Ando, A. B. Fowler, F. Stern	
PRB **13**, 5188 (1976)	1023	20.8	Special Points for Brillouin-Zone Integrations	H. J. Monkhorst, J. D. Pack	←

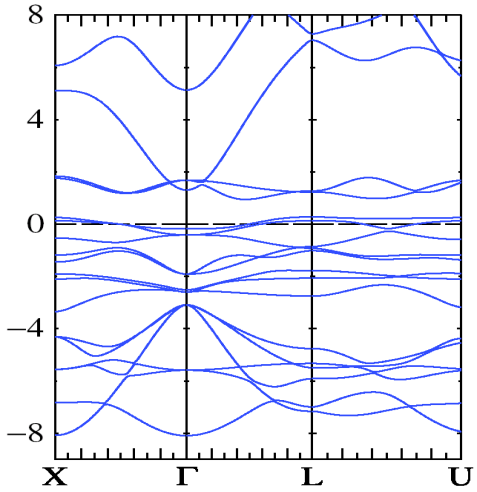
PR, Physical Review; PRB, Physical Review B; PRL, Physical Review Letters; RMP, Reviews of Modern Physics.

- Source: Physics Today 58, 49 (2005)

Failures in the Local Density Approximation

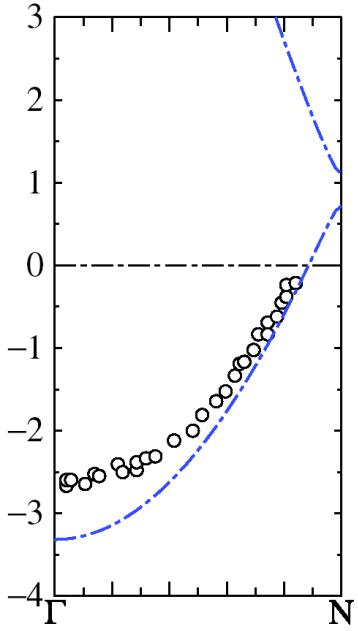
Semi. bandgaps too small

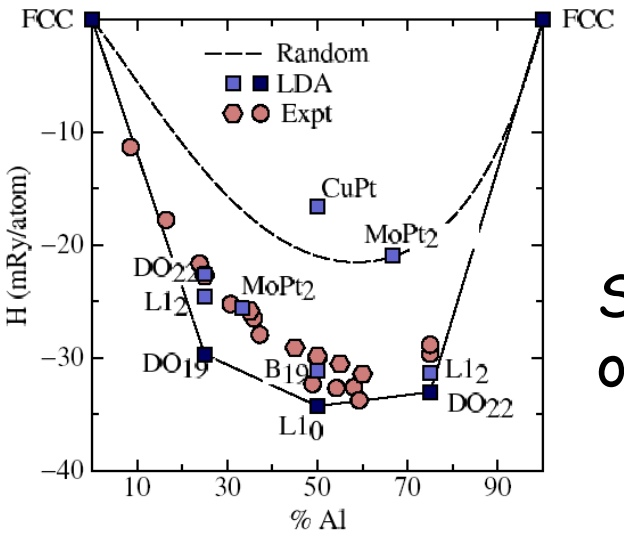




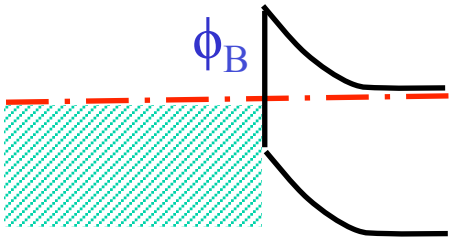
LDA CoO is metalllic but real CoO is AFM insulator (~4 eV gap)

Poor Na bands





Systematic overbinding



Schottky barriers at metal-semi contacts fall too close to Valence Band

Two possible explanations for LDA error

What is the dominant source of difficulty in the L(S)DA?

Explanation I: Ansatz for $E^{xc}[n]$ is the primary cause.

Explanation II: Kohn-Sham ψ_i and eigenvalues ε_i the Lagrange multipliers of the KS hamiltonian

$$\hat{H}^{\sigma}_{KS} = -\frac{\hbar^2}{2m}\nabla^2 + [V^{\sigma}_{KS}(\mathbf{r}) = V_H(\mathbf{r}) + V_{ext}(\mathbf{r}) + V^{\sigma}_{xc}(\mathbf{r})]$$ are fictitious.

Q: How do we assess the source of error?

A: Density-functionalize nonlocal functionals and check.

Not a strict division (there is an interplay between them).
But roughly:

For ground state properties, I is the primary problem

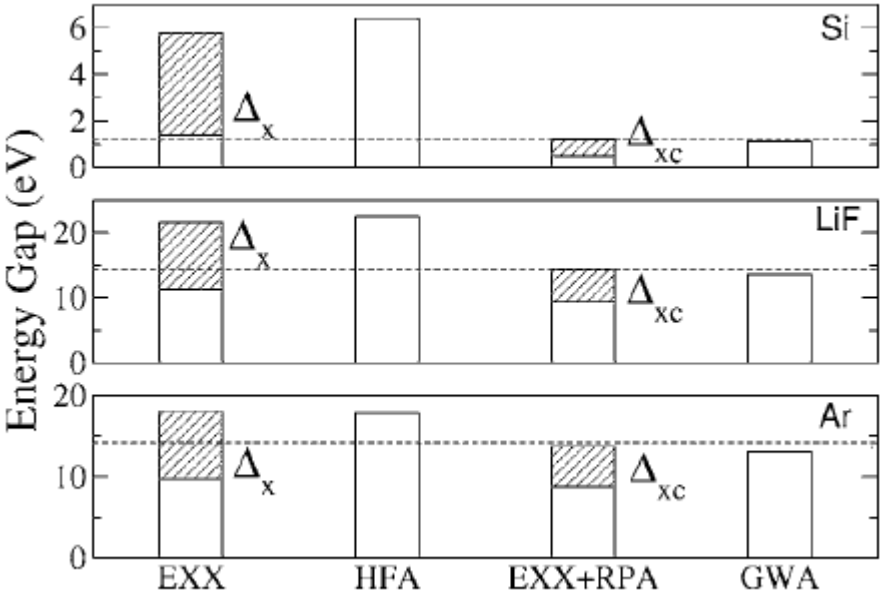
For excited state properties II is the primary problem.

Connection between DFT and QP levels

ψ_i and ε_i fictitious ⇒ discontinuity Δ in V^{xc} betw/ highest occ state and next higher one

Grüning, Marini, Rubio, (J. Chem. Phys. 124, 154108) evaluated Δ_{xc} by making OEP (density functionalized) GW for Si, LiF,Ar

$$\Delta_{\rm xc} = \langle \psi_{N+1} | \Sigma^{GW}_{\rm xc}(\varepsilon_{N+1}) - v_{\rm xc} | \psi_{N+1} \rangle - \langle \psi_N | \Sigma^{GW}_{\rm xc}(\varepsilon_N) - v_{\rm xc} | \psi_N \rangle .$$



Results show:

OEP gap (EXX+RPA) close to usual LDA gap. Thus Explanation II: the fictitious nature of ψ_i and ε_i are the primary problem

Many attempts to extend the LDA

- Good ground-state properties in weakly correlated systems.
- Excited state properties are much worse.

Many attempts to extend, improve on the LDA

- Self-Interaction Correction (Perdew, Zunger, PRB 23, 5048 (1981))
- LDA+U (Anisimov, Zaanen, Andersen, Phys. Rev. B 44, 943 (1991))
- LDA+Screened exchange, (Seidl et al, PRB 53, 3764 (1996))
- LDA+DMFT (Anisimov et al, J. Phys. C9, 7359 (1997))
- Mix Hartree-Fock with LDA (B3LYP, Becke; Scuseria)
- Optimized Effective Potential (Kotani, PRL 74, 2989 (1995)

  Exact exchange, EXX+RPA closest to LDA: local potential $V^{\mathrm{eff}}(\mathbf{r})$.

- All have significant successes to their credit, but improve one or another property in some special cases.
- Removing locality is essential … but without removing the ansatz LDA starts with, hard to *systematically* improve on the basic framework

GW Approximation

- Hedin's GW approximation (1965): a major advance on Hartree-Fock theory. Conceptually, the Fock V_x gets replaced by GW.
  - G = Green's Function, W = screened coulomb interaction
- Hartree Fock: e^- senses an effective potential V_x owing to correlated motion. V_x = functional derivative of E_x and can be written in terms of Green's functions as:

$$V_x(\mathbf{r}) = i\int G(\mathbf{r},\mathbf{r}')\frac{1}{|\mathbf{r}-\mathbf{r}'|}d^3\mathbf{r}' = iGv$$

GW: Hartree-Fock like but the coulomb e^-–e^- repulsion $1/|\mathbf{r}-\mathbf{r}'|$ is dynamically screened:

$$V_{\text{bare}}(\mathbf{r},\mathbf{r}') = \frac{1}{|\mathbf{r}-\mathbf{r}'|} \rightarrow W(\mathbf{r},\mathbf{r}',\omega) = \frac{1}{\varepsilon(\mathbf{r},\mathbf{r}',\omega)}\frac{1}{|\mathbf{r}-\mathbf{r}'|}; \quad \Sigma = iGW$$

- More rigorously: GW = the lowest order term in an exact expansion in W (many-body perturbation theory)

Advantages of the GW Approximation

- The GW approximation can potentially redress the worst failings inherent in both Hartree-Fock and LDA:
- HF : the nonlocality present, but not screened (disaster)
- LDA: nonlocality ⇒ pathologies in local potential. (Many problems, e.g. cannot break reversal symmetry).

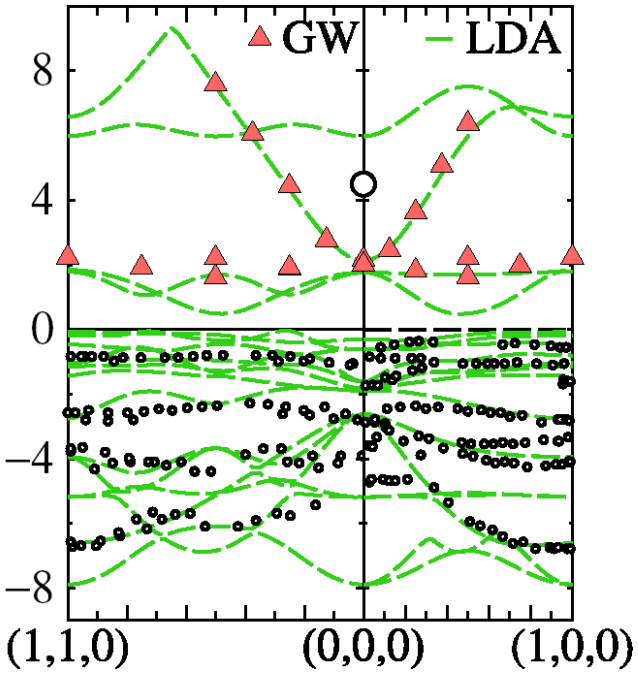
- But … GW is a perturbation theory: first term in an expansion in W. Perturbation theory must be carried out around some starting point H_0. How choose H_0?

- Major development (Hybertsen and Louie, 1987): use LDA as starting point

$$H_0 = H^{\mathrm{LDA}} \Rightarrow G = G^{\mathrm{LDA}},\ W = W^{\mathrm{LDA}};\ \Sigma = iG^{\mathrm{LDA}} W^{\mathrm{LDA}}$$

- Hugely successful in semiconductors

Failings in LDA-based GW

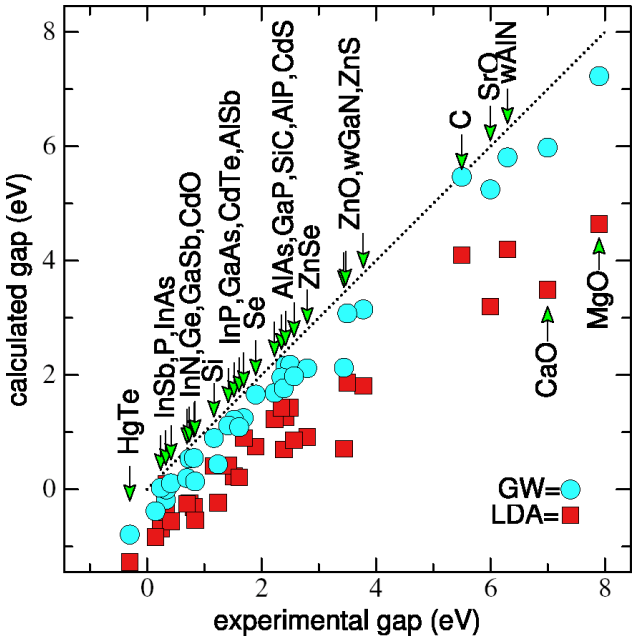


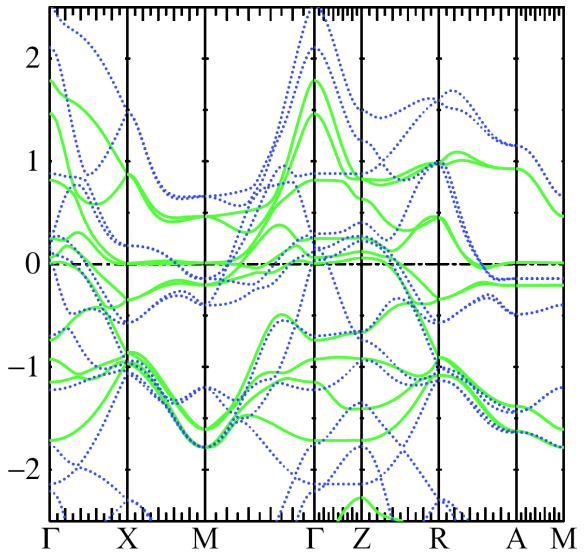
If $H_0 = H^{\mathrm{LDA}}$, then $GW \rightarrow G^{\mathrm{LDA}}W^{\mathrm{LDA}}$

←

NiO only slightly improved over LDA

$G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ bandgaps are underestimated →





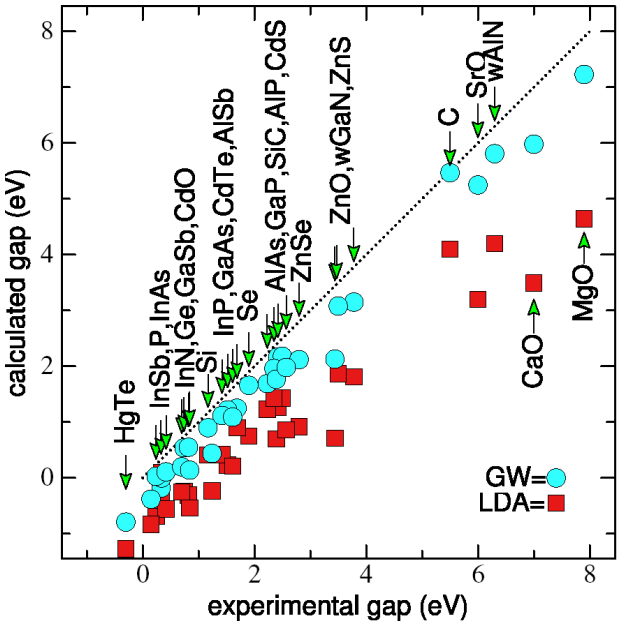
← Fermi surfaces, magnetic structure in metals can be nonsensical, e.g. in FeTe.

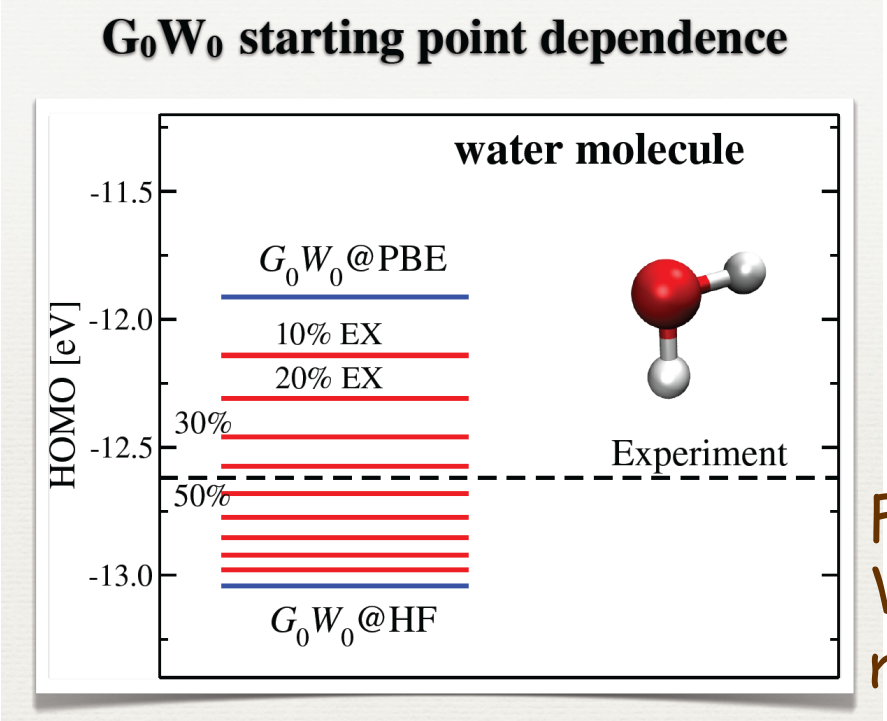
Many other problems;
see PRB B74, 245125 (2006)

Ambiguities in GW from ambiguities in H_0

The GW approximation significantly ameliorates errors in E_G.
Unlike many extensions to the LDA, (hybrid functionals) it is true *ab initio*.

But GW is a perturbation theory usually calculated around a noninteracting H_0.





But $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ works well only in special cases.
Change H_0 ⇒ improve result, but not universal or predictive.
Ambiguities cannot be avoided.

From Patrick Rinke, CECAM Workshop Green's function methods, Toulouse, 5 June 2013

Quasiparticle self-consistent GW Approximation

A new, first-principles approach to solving the Schrodinger equation within Hedin's GW theory.

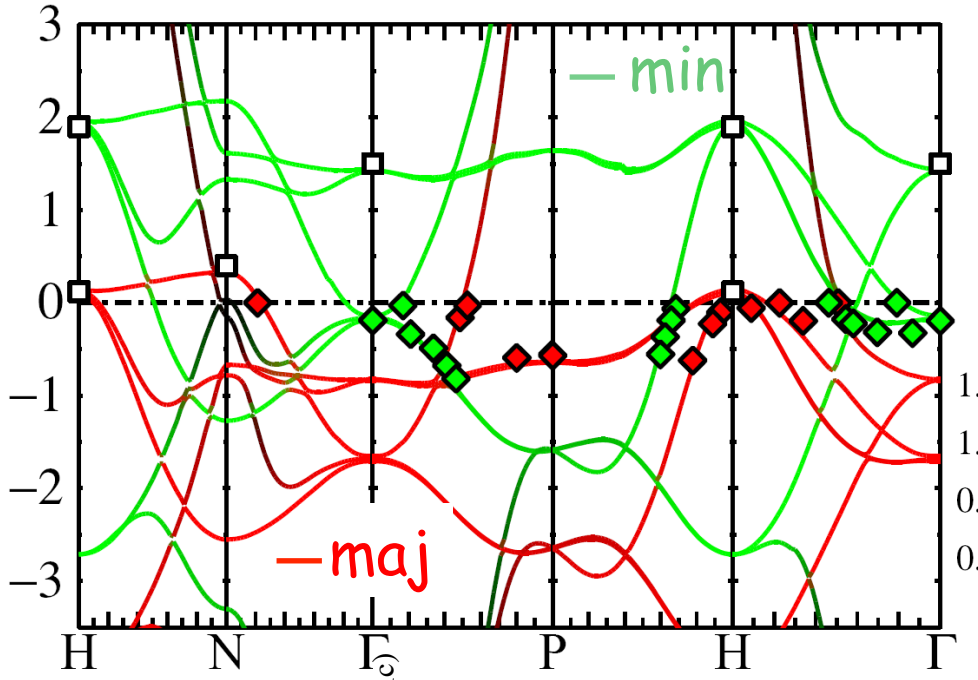
Principle : Can we find a good starting point H_0 in place of H^{LDA} ? How to find the best possible H_0 ?

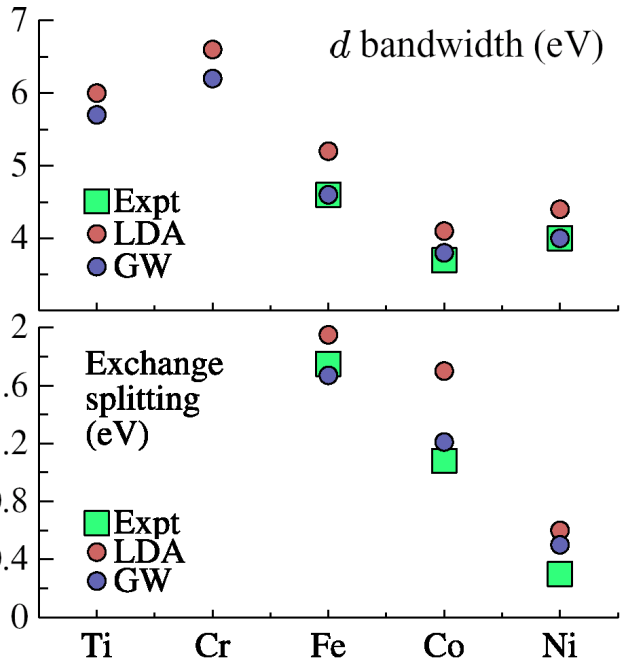
Requires a prescription for minimizing the difference between the full hamiltonian H and H_0.

QSGW : a self-consistent perturbation theory where self-consistency determines the best H_0 (within the GW approximation) PRL 96, 226402 (2006)

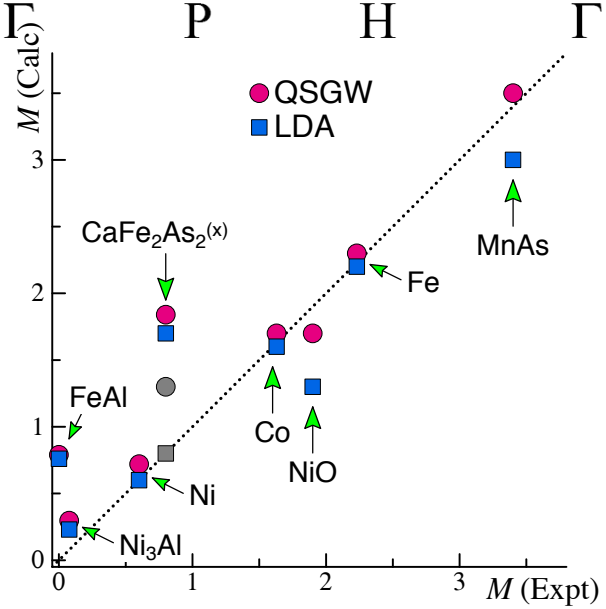
QSGW theory applied to elemental *d* systems

Fe in FL regime: matches ARPES within resol'n of expt



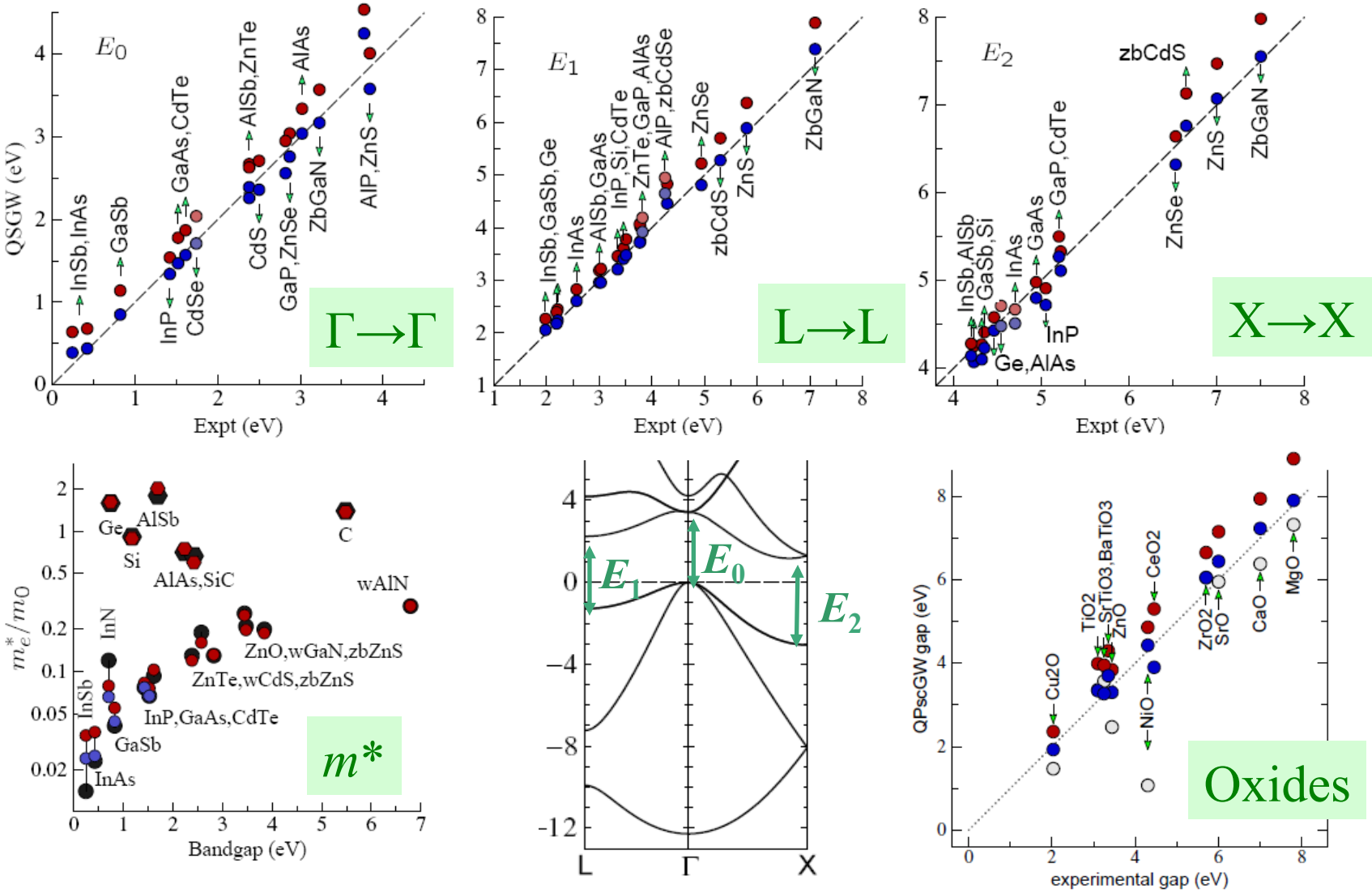


- Magnetic moments reliably predicted when local-moment picture applies.
- Breaks down in itinerant magnets



* *d* band exchange splitting and bandwidths well described (exception: Ni)

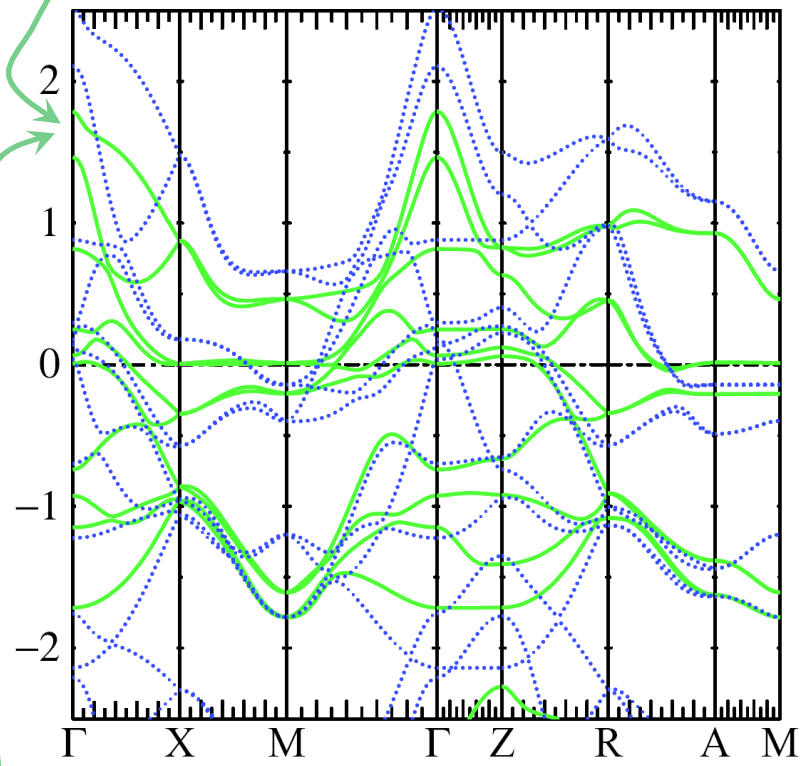
Critical points, m* in *sp* bonded systems



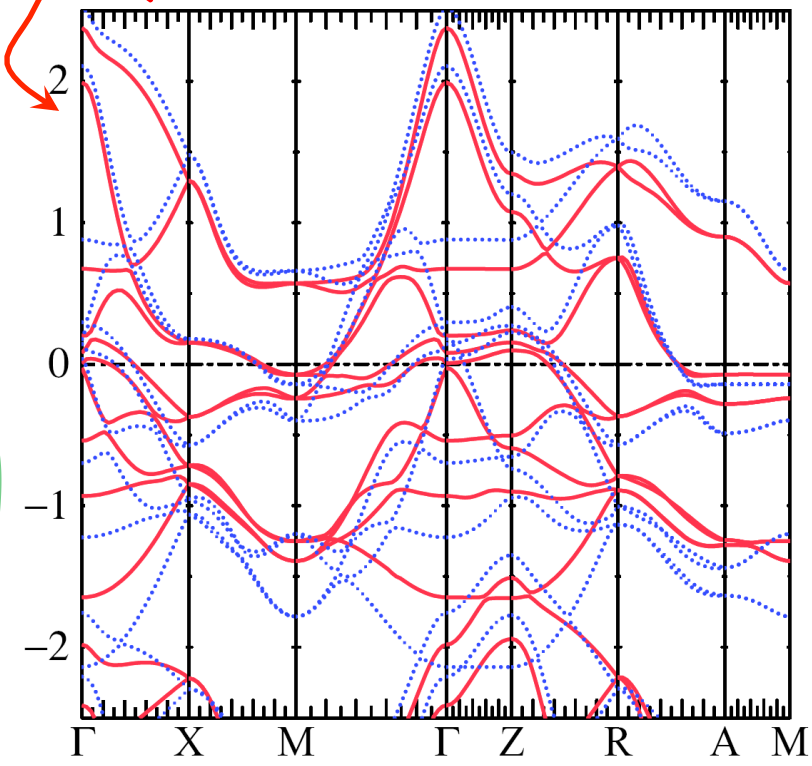
CP's always slightly overestimated (●); m* mostly quite good

QS*GW* can dramatically improve the quality of *GW*

$G^{LDA}W^{LDA}$ vs LDA, FeTe



QSGW vs LDA, FeTe



$G^{LDA}W^{LDA}$: Fe *d* /As *p* alignment shifts by ~1/2 eV ⇒ puts Fe *d* states too high.

Fermi surface nonsensical.

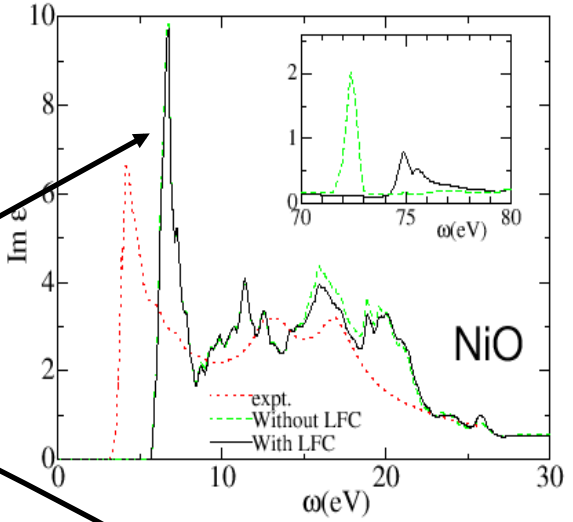
QSGW: marked narrowing of Fe *d* bandwidth.

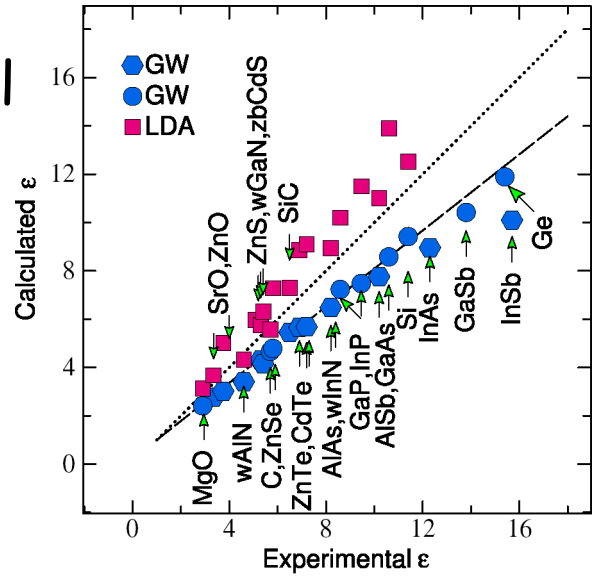
But still too large compared to experiment …

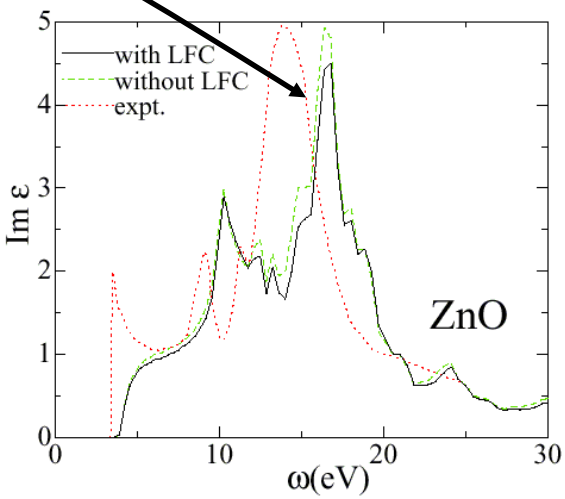
skip

Systematics of Errors

- ✓Unoccupied states universally too high
  - ✓~0.2 eV for *sp* semiconductors;
  - ✓<~1eV for itinerant *d* $SrTiO_3$, TiO_2
  - ✓>~1eV for less itinerant *d* NiO
  - ✓>~3 eV for *f* Gd,Er,Yb
- ✓Peaks in Im ε(ω) also too high
- ✓ε_∞ 20% too small
- ✓Magnetic moments slightly overestimated





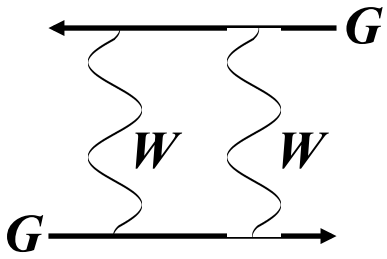


skip

Consequences of improving $\Pi(q,\omega)$

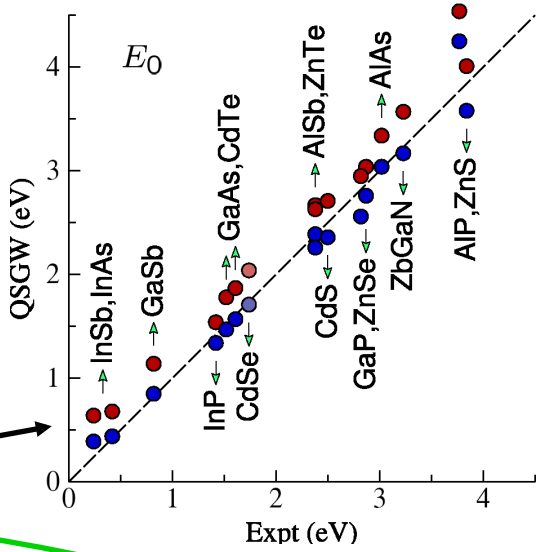
Errors are mostly consistent with missing electron-hole correlation in the dielectric function -- excitonic effects

Ladder diagrams seem to reliably correct Im $\varepsilon(\omega)$ starting from QP picture in many systems, e.g. Cu_2O (Reining et al, PRL 2006)

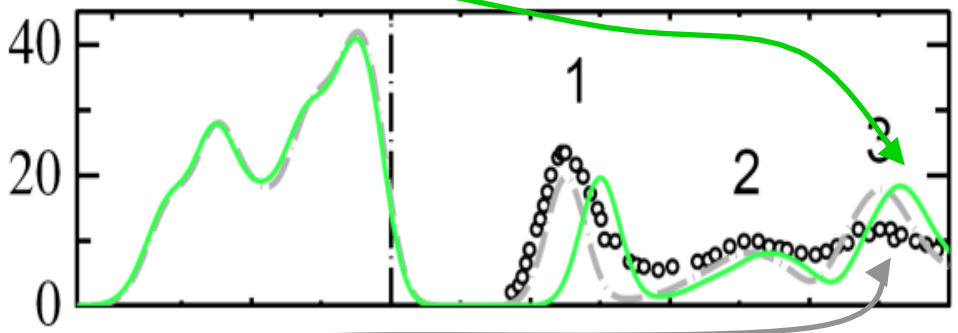


Shishkin, Marsman and Kresse (PRL 99, 246403): diagrams largely eliminate QS*GW* gap errors in semiconductors ...

Similar result just from scaling Σ by 0.8



Inverse photoemission in (highly correlated) NiO is similarly corrected ...

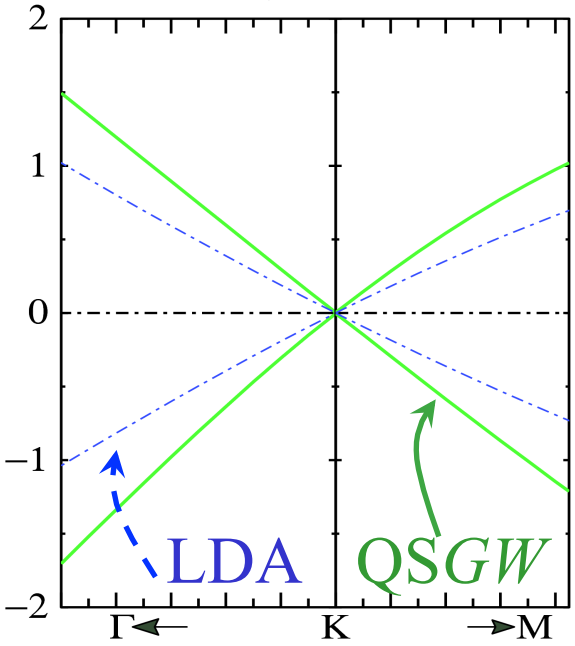


QS*GW* as a framework for H_0

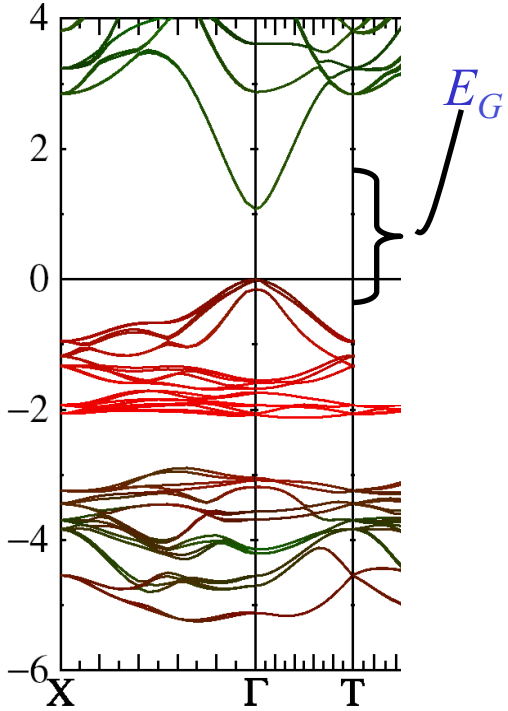
Except for specialized many-body effects, properties of interest are typically sufficiently described by H_0, e.g. semiconductor band offsets, magnetic moments, transport.

QS*GW* (QSGW+BSE) generates a nearly optimal H_0 for many kinds of materials classes …

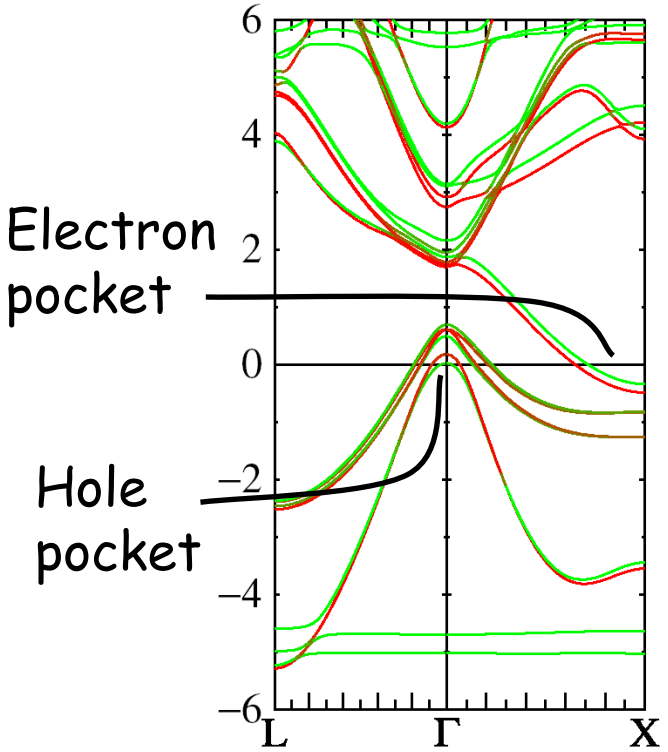
graphene



$CuInSe_2$



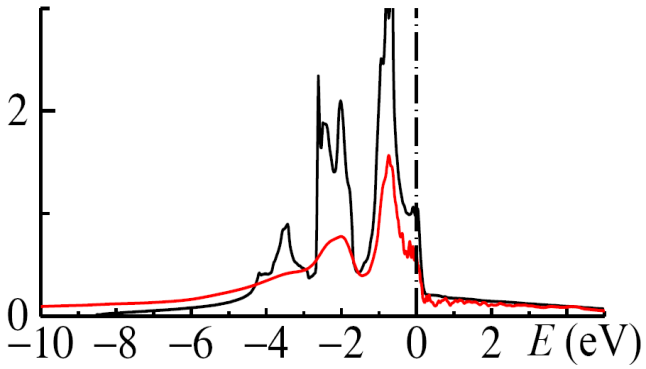
ErAs



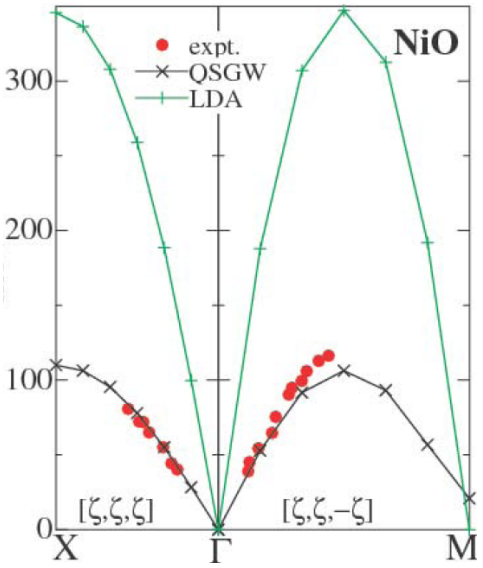
Dual Nature of QSGW

QSGW generates both an optimal H_0 or G_0 and an interacting G that contains dynamical, many-body effects.

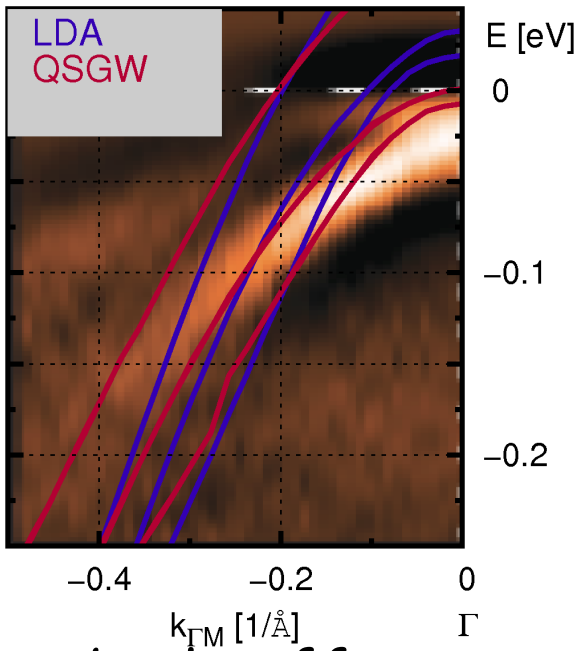
DOS, Fe



Spin waves, NiO



ARPES, $BaFe_2As_2$



QSGW is relatively simple … important many-body effects (superconductivity, Kondo physics, …) not captured. But it provides an ideal framework to handle such correlations.

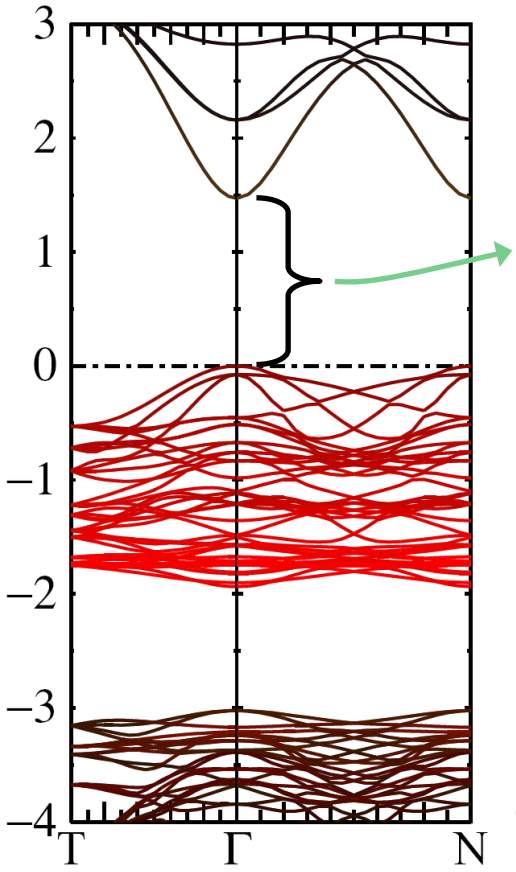
Key point: what "correlations" are depends on reference!

Limitations to QSGW: cost

QSGW is expensive: as now implemented scales as N^4:

32 atoms feasible on 12 processor cluster ⇒ 100 atoms on large facility

$(Cu_2ZnSnS_4)_4$



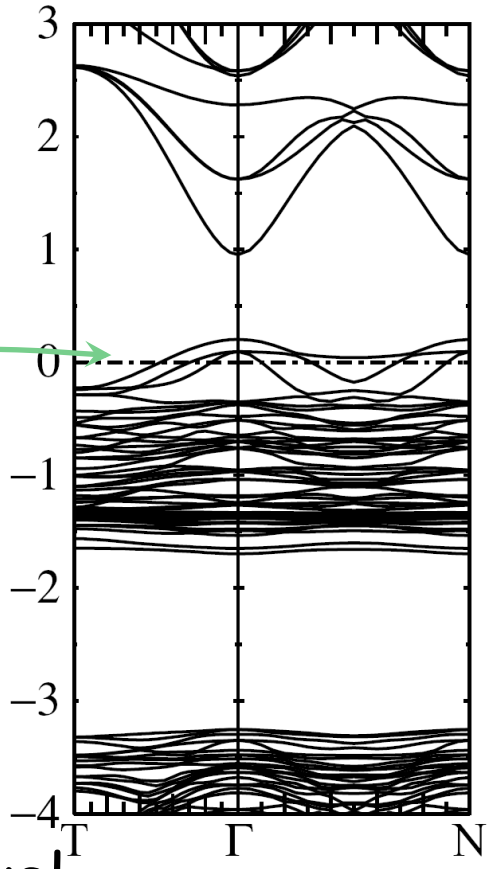
Example: 32 atom supercell of Kesterite

Bulk $E_G = 1.5$ eV (expt)

Thermodynamic calc predict: Kesterite has numerous antisite defects, e.g. Cu_{Zn} .

Model as $Cu_9Zn_3Sn_4S_{16}$.
Cu_{Zn}: shallow acceptor but cell too small to pinpoint level

… + Cu_{Zn}



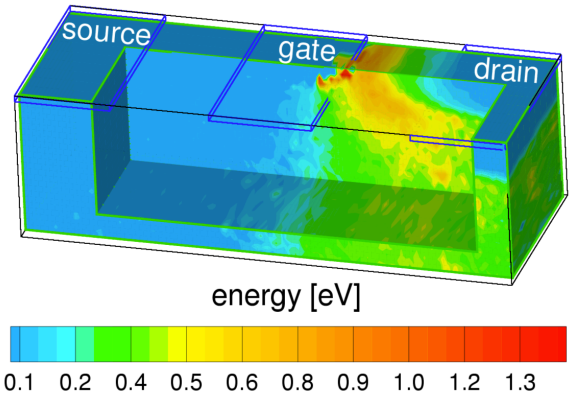
Bridging Length Scales

Algorithmic improvements ⇒~50-100 atoms with efficient use of parallel architectures. Sufficient for many key properties at the nanoscale level, e.g. band offsets, energy levels of defects.

Alternate strategies: Use QS*GW* H_0 to:

1) Map onto reduced classical H_0, or quantum H_0, (e.g. Wannier functions, many-body context)

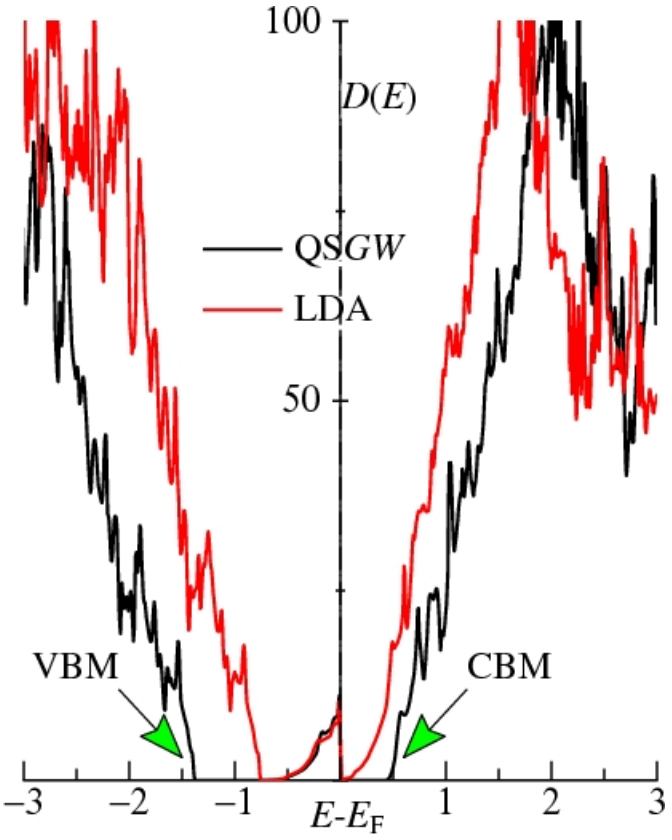
2) Use QS*GW* as a parameter generator for empirical hamiltonians, classical or quantum type. An electronic device simulator requires energy bands, scattering matrix elements.

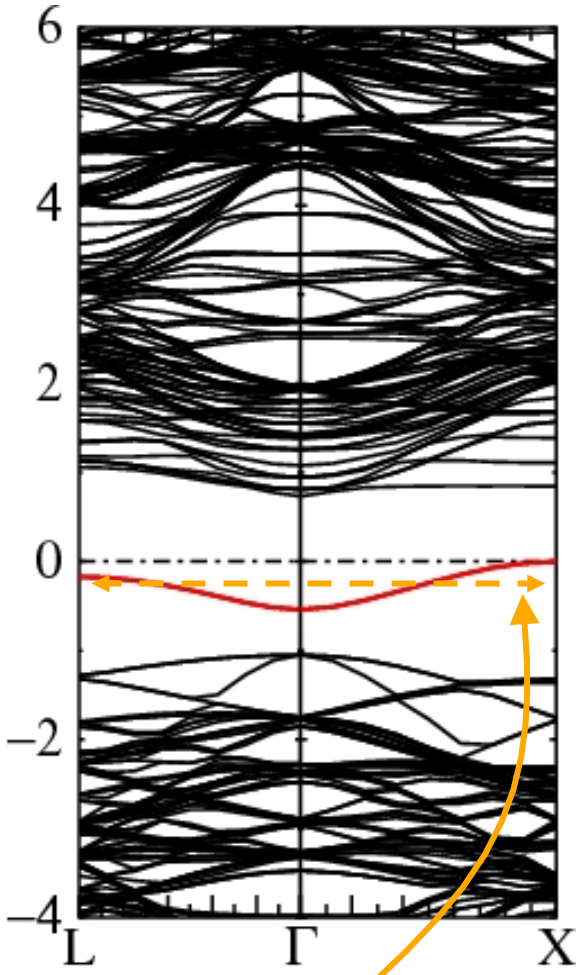


As Antisite in GaAs: QSGW vs LDA

Localized level at VBM + 0.75 eV.
LDA puts As_{Ga} at ~0.35eV w/ E_G = 0.3eV

Different extensions to LDA may predict the same bandgap, yet place the deep level at different positions within the gap.





QSGW-derived : band CG comes out near observed position

QSGW as an engine for Classical Simulations

Classical transport:

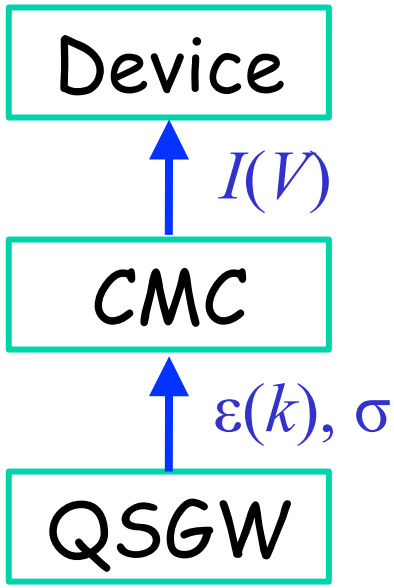
Sophisticated techniques developed to solve Boltzmann transport equation, which can model real electron devices.

Example: band Cellular Monte Carlo (CMC) (Saraniti, ASU).

CMC simulators can feed (in principle) into higher-level simulators, e.g. circuit simulators, or solar cell simulators.

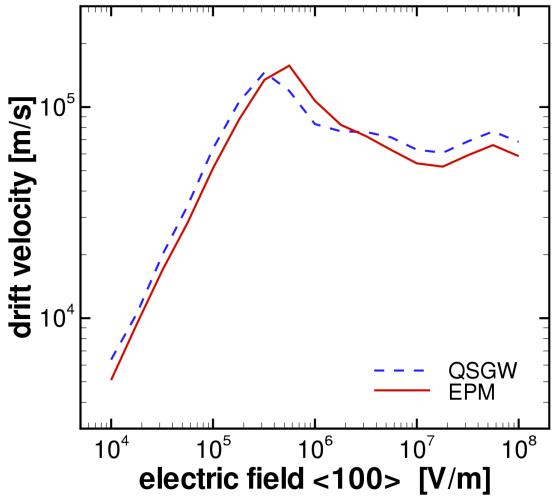
CMC requires as input: energy bands, scattering matrix elements (electron-phonon, impurity scattering)

These can be supplied by QSGW. Makes feasible an *ab initio* device simulator. (Future area of research)

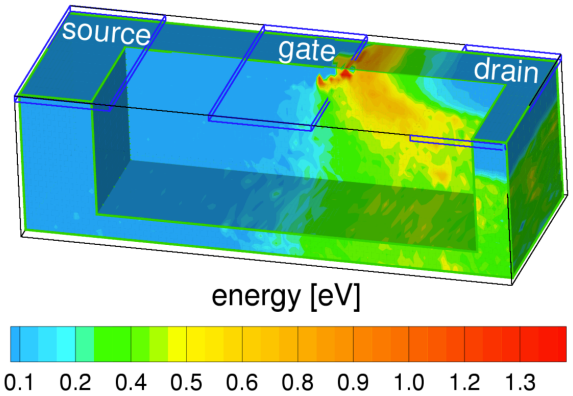


Bridging Length Scales: feasibility demo

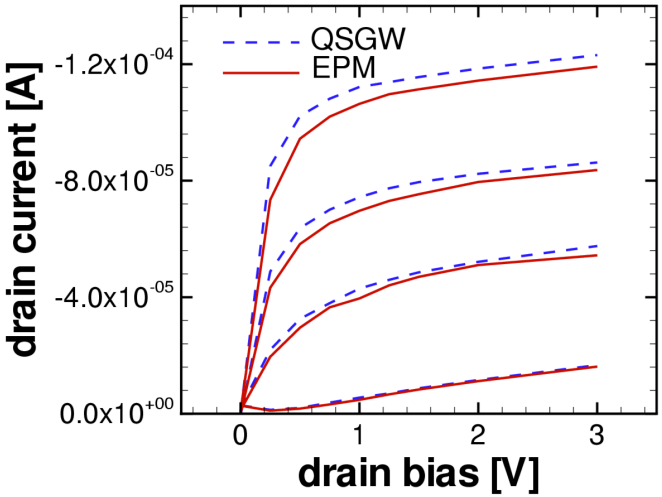
First step: feed QS*GW* bands into Cellular MC simulator.
Test case: GaAs (bands very well known from experiment)



Velocity-field characteristics of bulk GaAs, computed by CMC QSGW and empirical PP energy bands.



Electron energy distribution function in a 3D CMC simulation of GaAs MESFET



Corresponding current-voltage characteristics. Slight differences with a calculation using an EPM band structure.

Conclusions

- The *QSGW* approximation
  - has some formal justification.
- Unique features:
  - Reliably treats variety of properties in a wide range of materials in a true ab initio manner. A kind of gold standard at the 1-particle level
  - The errors are systematically improvable.
  - Truly predictive when correlations are not strong.

Limitations:

- Does not handle strong correlations properly
  - ✓ Include extra diagrams, or combine with DMFT.
- Cost: N^4 scaling
  - ✓ Build reduced or model hamiltonians for defect studies, *ab initio* device simulation
  - ✓ Redesign algorithms.