Quasiparticle Self-Consistent GW Approximation: Strengths and Weaknesses

Mark van Schilfgaarde
Arizona State University

The Quasiparticle self-consistent GW approximation–QS*GW*

- PRL93, 126406; PRL 96, 226402; PRB76, 165106.

❖ What it is, how it differs from standard sc-GW

❖ Range of applicability, and limits to precision

❖ How well does QS*GW* work in complex systems?

"Complex" can refer to

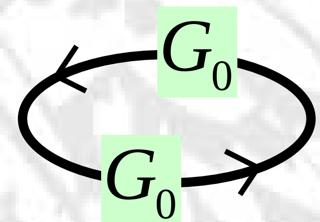
- ➢ Many-atom, inhomogeneous structures, e.g surfaces
  - Is success in simple systems replicated?
  - Limited by algorithm efficiency and computer power
- ➢ Complexities originating from electron correlations.
  - Depends on "smallness" of approximations in QS*GW*.

GW: A Perturbation theory

Start from some non-interacting hamiltonian H_0.

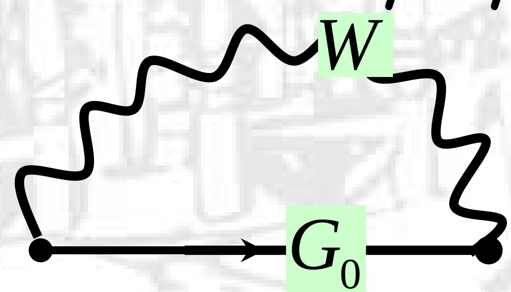
1. $$H_0 = -\frac{\nabla^2}{2} + V_{eff}(\mathbf{r},\mathbf{r}') \Rightarrow G_0 = \frac{1}{\omega - H_0}$$ Example: $= H^{\text{LDA}}$

2. $\Pi = -i\,G_0 \times G_0$ RPA Polarization function



3. $$W = \varepsilon^{-1} v = (1 - \Pi v)^{-1} v$$
$$v(\mathbf{r},\mathbf{r}') = |\mathbf{r} - \mathbf{r}'|^{-1}$$
Dynamically screened exchange
(Recover HF theory by $\varepsilon \to 1$)

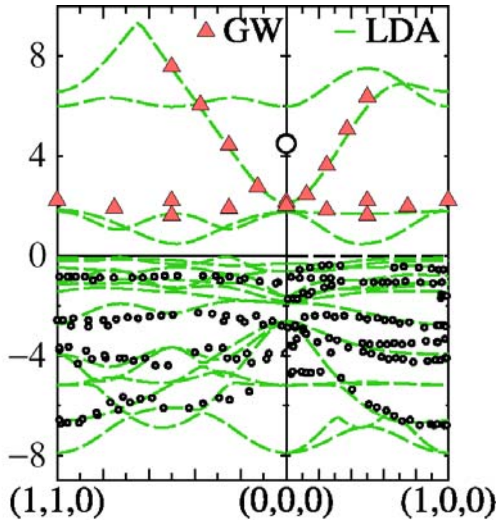
4. $\Sigma = i\,G_0 W$ Self-energy $\Sigma =$



$$H(\mathbf{r},\mathbf{r}',\omega) = -\frac{\nabla^2}{2} + V^H(\mathbf{r}) + V^{ext}(\mathbf{r}) + \Sigma(\mathbf{r},\mathbf{r}',\omega)$$

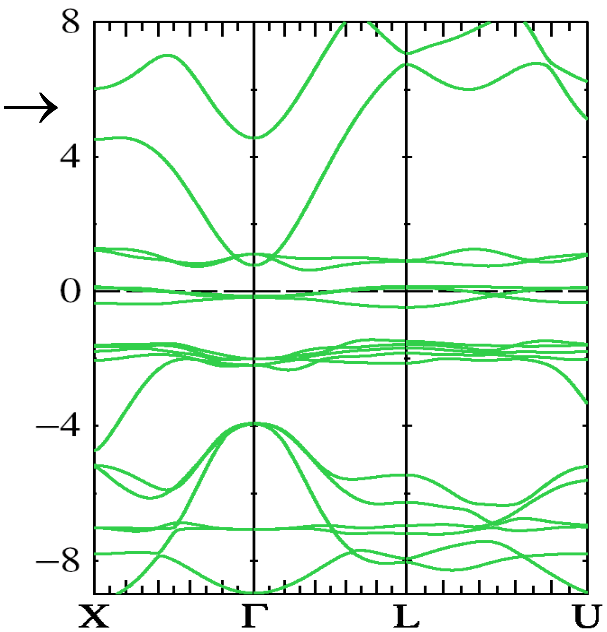
LDA-based GW Approximation

GW is a perturbation theory around some non-interacting hamiltonian H_0. Usually $H_0 = H^{\mathrm{LDA}}$. Then $GW \rightarrow G^{\mathrm{LDA}}W^{\mathrm{LDA}}$

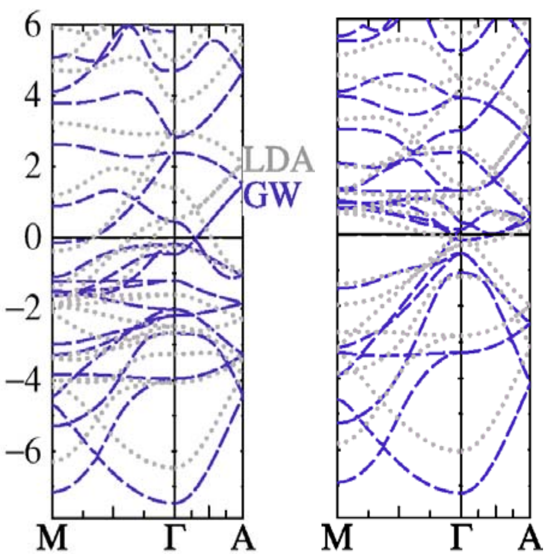


← NiO only slightly improved over LDA

CoO is still a metal→



Most $4f$ systems similarly have a narrow $4f$ band at E_F



← Bands, magnetic moments in MnAs are poor.

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

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: a self-consistent perturbation theory

Partition H into $H_0 + \Delta V$ and (noninteracting + residual) in such a way as to minimize ΔV :

$$G_0 = \frac{1}{\omega - H_0} \xrightarrow{GWA} G = \frac{1}{\omega - \left(H_0 + \Delta V(\omega)\right)}$$

$$\left(\omega - (H_0 + \Delta V(\omega))\right) G(\omega) = \delta(\mathbf{r} - \mathbf{r}')$$

We seek the $G_0(\omega)$ that most closely satisfies Eqn. of motion

$$\left(\omega - (H_0 + \Delta V(\omega))\right) G_0(\omega) \approx \delta(\mathbf{r} - \mathbf{r}')$$

$$\rightarrow \Delta V(\omega) G_0(\omega) \approx 0$$

If the GWA is meaningful, $G_0 \approx G$

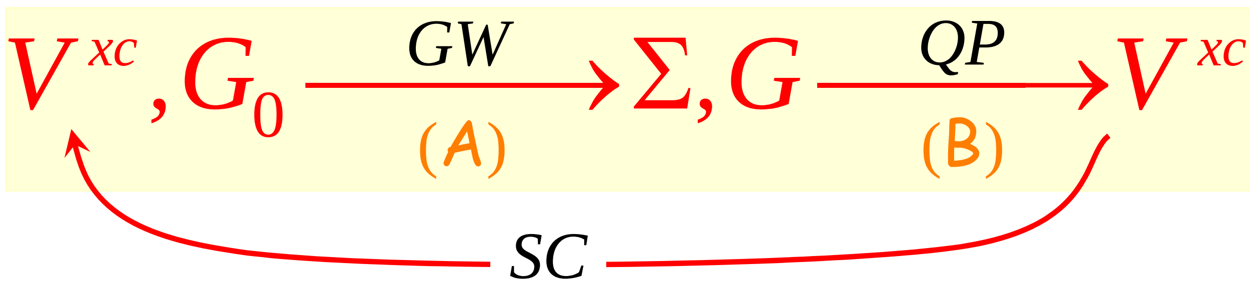
Q: How to find G_0 that minimizes $\Delta V\, G_0$?

QSGW cycle

A: Define a norm functional N that is a measure of the difference between $\psi[H]$ and $\psi[H_0]$

$$N = \frac{1}{2}\sum_{ij}\left|\langle\psi_j|\Delta V(\varepsilon_{\rm i})|\psi_i\rangle\right|^2 + \left|\langle\psi_j|\Delta V^{\dagger}(\varepsilon_{\rm i})|\psi_i\rangle\right|^2$$

Step 0: Generate trial $V^{\rm xc}$ from LDA, LDA+U, or ...



Step A: Generate $\Sigma(\omega)$ from $V^{\rm xc}$ using the GWA.

Step B: Find a static and hermitian $V^{\rm xc}$ as close as possible to $\Sigma(\omega)$, by minimizing N (next slide)

Use $V^{\rm xc}$ as trial $V^{\rm xc}$ and iterate A,B until self-consistency

Should be independent of starting point (not guaranteed)

$$G_0 \xrightarrow[\text{(A)}]{GW} G \xrightarrow[\text{(B)}]{QP} G_0$$

Minimize N (approximately) by choosing

$$V^{\text{xc}} = \frac{1}{2}\sum_{ij} \langle \psi_i | \mathrm{Re}\left(\Sigma(E_i) + \Sigma(E_j) \right) | \psi_j \rangle$$

Defines a noninteracting effective potential with Hartree-Fock structure: $\left\{ -\nabla^2 + V^{\text{ext}} + V^{\text{H}} + V^{\text{xc}} \right\} \psi_i = \varepsilon_i \psi_i$

At self-consistency, ε_i of G matches ε_i of G_0 (real parts)

Self-consistency is thus a means to determine the best possible starting hamiltonian H_0 (within the GWA).

See PRB76, 165106 (2007).

Shishkin, Marsman, and Kresse: improved W by adding (approximate) ladder diagrams (PRL99, 246403 (2007))

QSGW is not true self-consistent GW

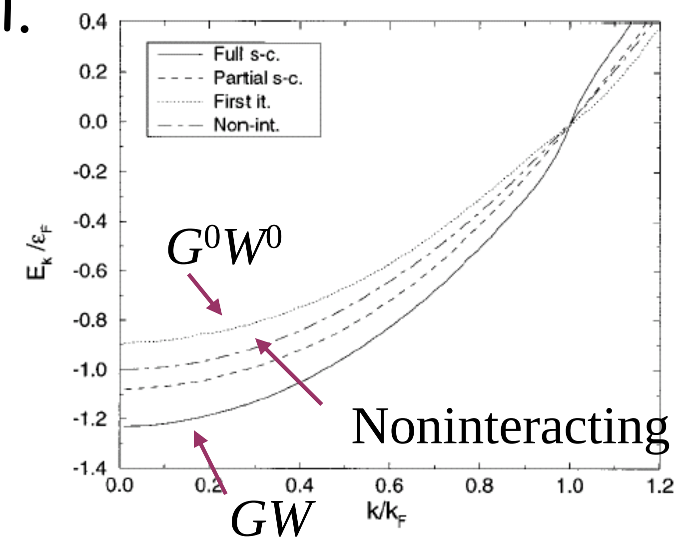
True self-consistent GW (scGW)

$$G \Rightarrow \Pi = -iGG \Rightarrow W = \varepsilon^{-1}v \Rightarrow \Sigma = iGW \Rightarrow G = \frac{1}{\omega - (T + V^H + V^{ext} + \Sigma)}$$

True self-consistent GW looks good as formal theory:

→ Based on Luttinger-Ward functional.
→ Keeps symmetry for G
→ Conserving approximation

But poor in practice, even for the electron gas



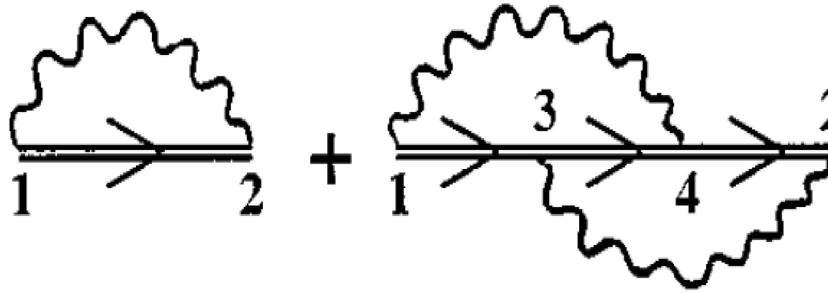
"Z-factor cancellation" is not satisfied (next slides)

B. Holm and U. von Barth, PRB57, 2108 (1998)

Higher order terms in Jellium

E. Shirley compared sc-*GWGWG* to sc-*GW* in Jellium: (Phys. Rev. B 54, 7758 (1996))

$$-i\Sigma(12) =$$



"While a non-self-consistent … *GW* treatment reduces occupied bandwidths by 10–30% …, selfconsistency leads to overall increased bandwidths. Subsequent inclusion of the next-order term in *GWGWG* restores reduced bandwidths, which agree well with experiment."

Z-factor cancellation in Σ

Exact $\Sigma = iGW\Gamma$. Suppose W is exact. Then

$$G_0 = \frac{1}{\omega - H_0 + i\delta}$$

$$G = \frac{1}{\omega - H_0 - \left[-V^{xc} + \Sigma(\omega_0) + (\partial\Sigma/\partial\omega)_{\omega_0} (\omega - \omega_0) \right] + i\delta}$$

$$Z = (1 - \partial\Sigma/\partial\omega)^{-1}$$

Residual of this pole (loss of QP weight) is reduced by Z

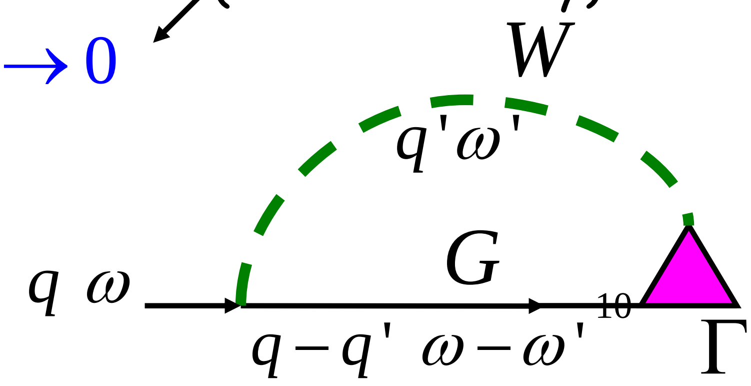
Write G as $G = ZG^0 + (\text{incoherent part})$

(Ward identity)

Also, $\Gamma = 1 - \partial\Sigma/\partial\omega = Z^{-1}$ for $q', \omega' \to 0$

Therefore,

$$GW\Gamma \approx G^0 W + (\text{incoherent part})$$



Z-factor cancellation in Π

$W=(1-\Pi v)^{-1}v$ is not exact, either.

A similar analysis for proper polarization Π .

$$\Pi = -iGG\Gamma \approx -iG_0G_0 + (\text{incoherent part})$$

(See Appendix A in PRB76, 165106 (2007)).

In the exact fully self-consistent theory, Z-factors cancel QP-like contribution in complicated ways.

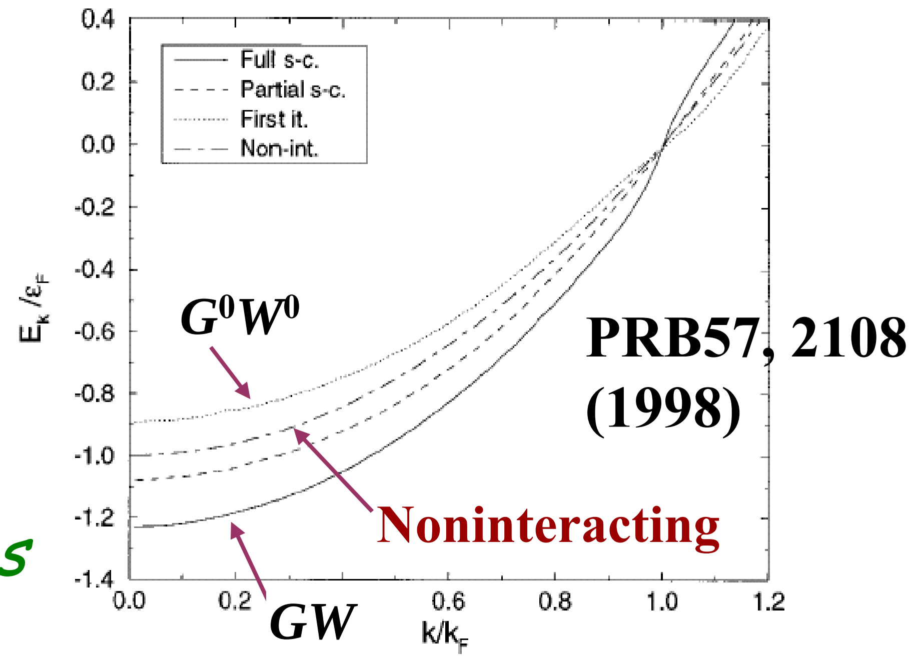
Self-consistent GW neglects Γ, so no Z-factor cancellation $\Rightarrow$ results rather poor. Higher order diagrams required to restore Z-factor cancellation.

Complexity avoided by doing perturbation theory around a noninteracting H_0 : convergence more rapid for a given level of approximation.

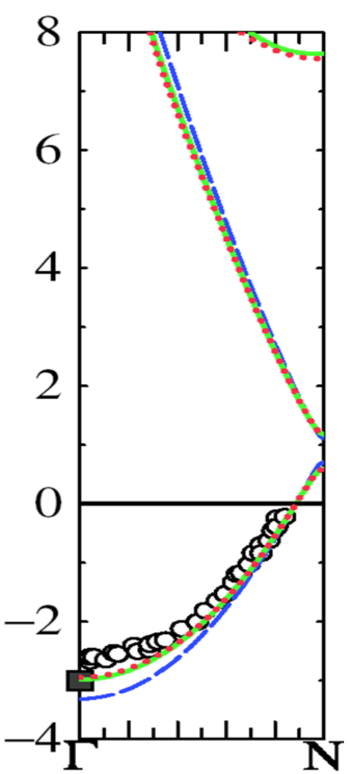
Na as approximate realization of HEG

Holm and von Barth compared scGW to G^0W^0 in the homogeneous electron gas.

The G^0W^0 bandwidth *narrows* by ~10%.



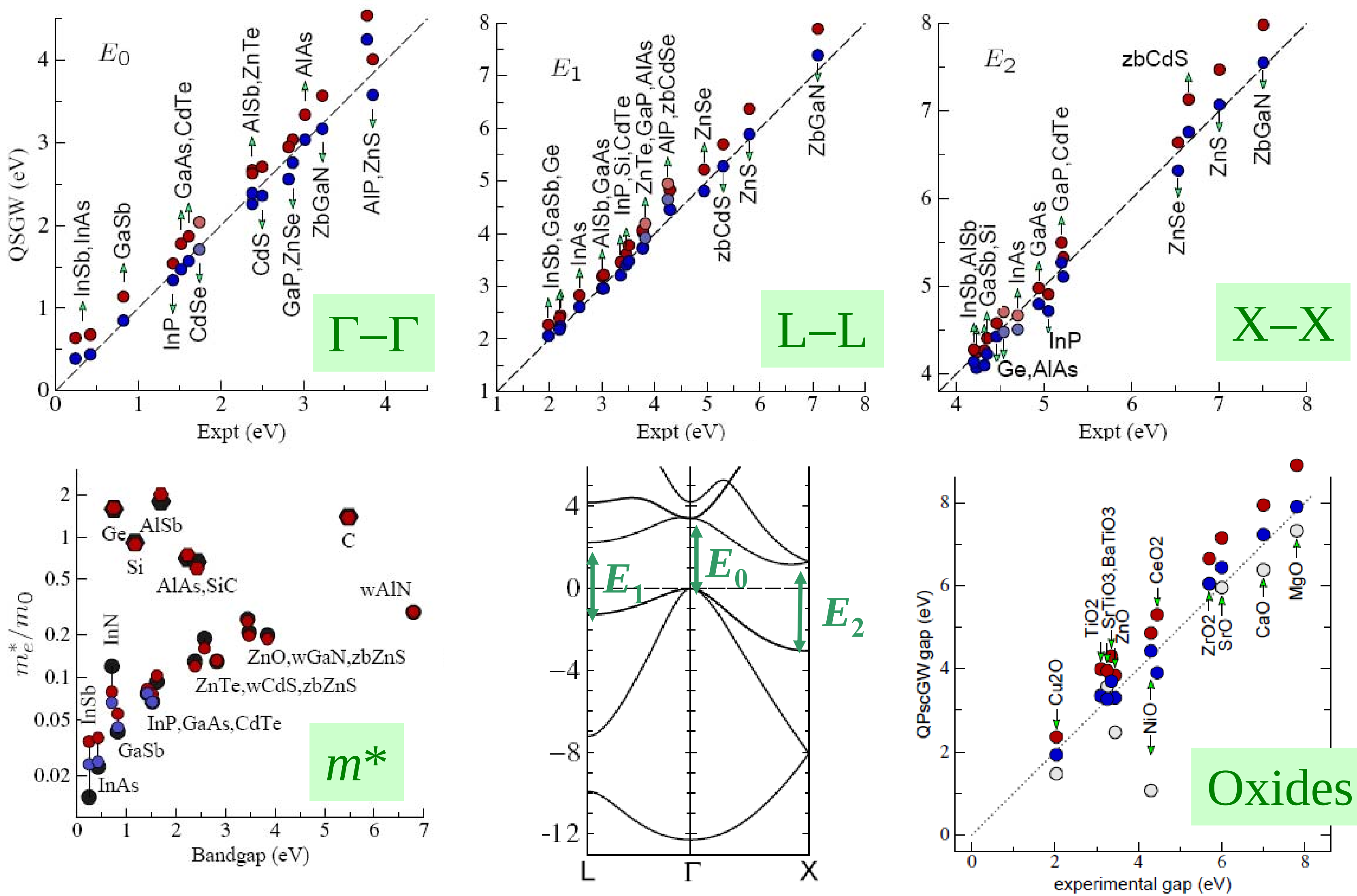
The scGW bandwidth *widens* by ~20%.



Shirley showed that the next order term, sc:GW+GWGWG essentially restores the G^0W^0 bandwidth PRB 54, 7758 (1996) in true scGW.

QSGW predicts the Na bandwidth to *narrow* relative to LDA by ~10%, in agreement with PE and standing wave measurements.

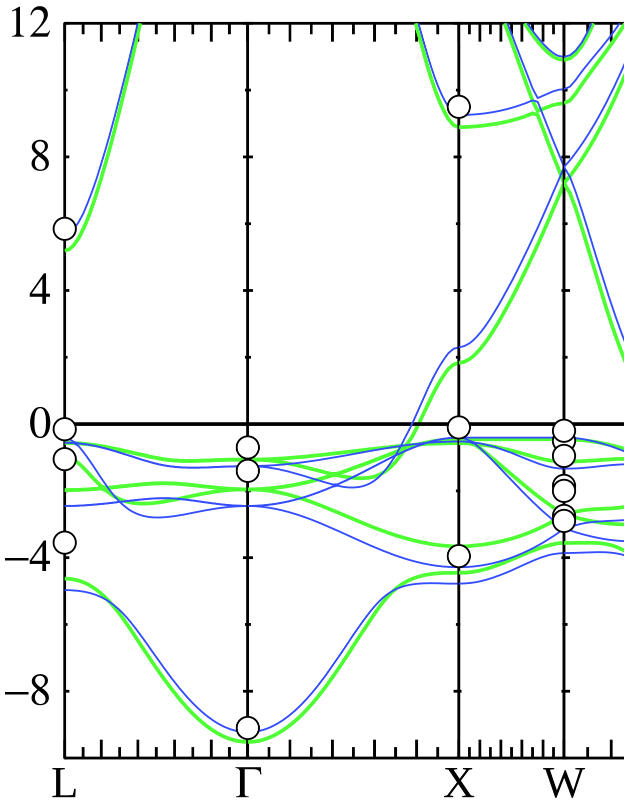
Critical points, m* in *sp* bonded systems



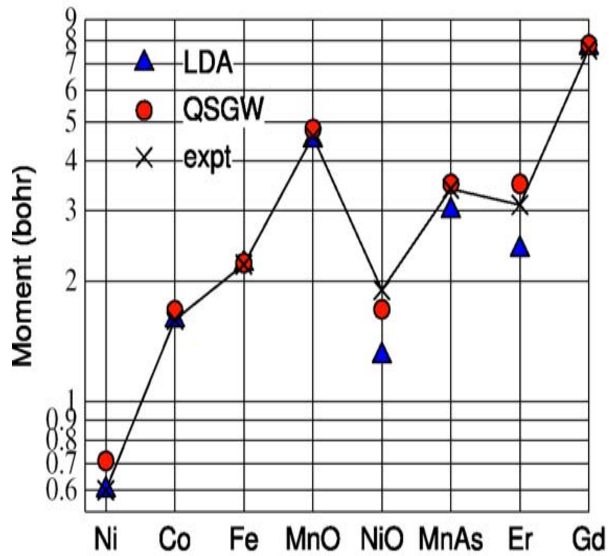
CP's always slightly overestimated (●); m* mostly quite good

QSGW in elemental *d* systems (mostly)

Ni (majority)

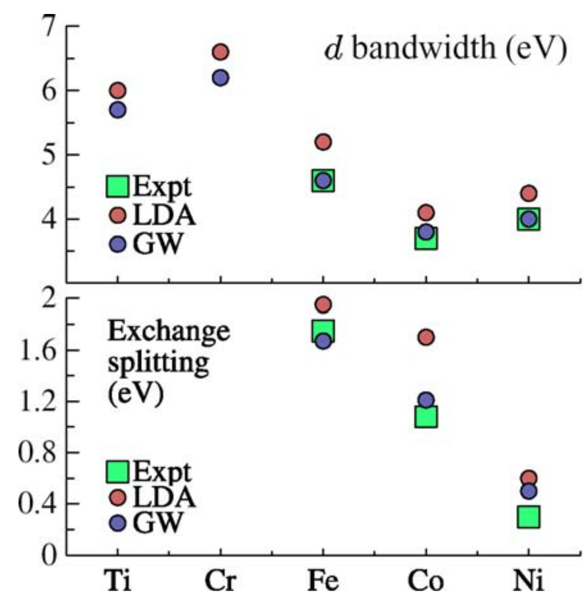


* Generally good agreement with photoemission O



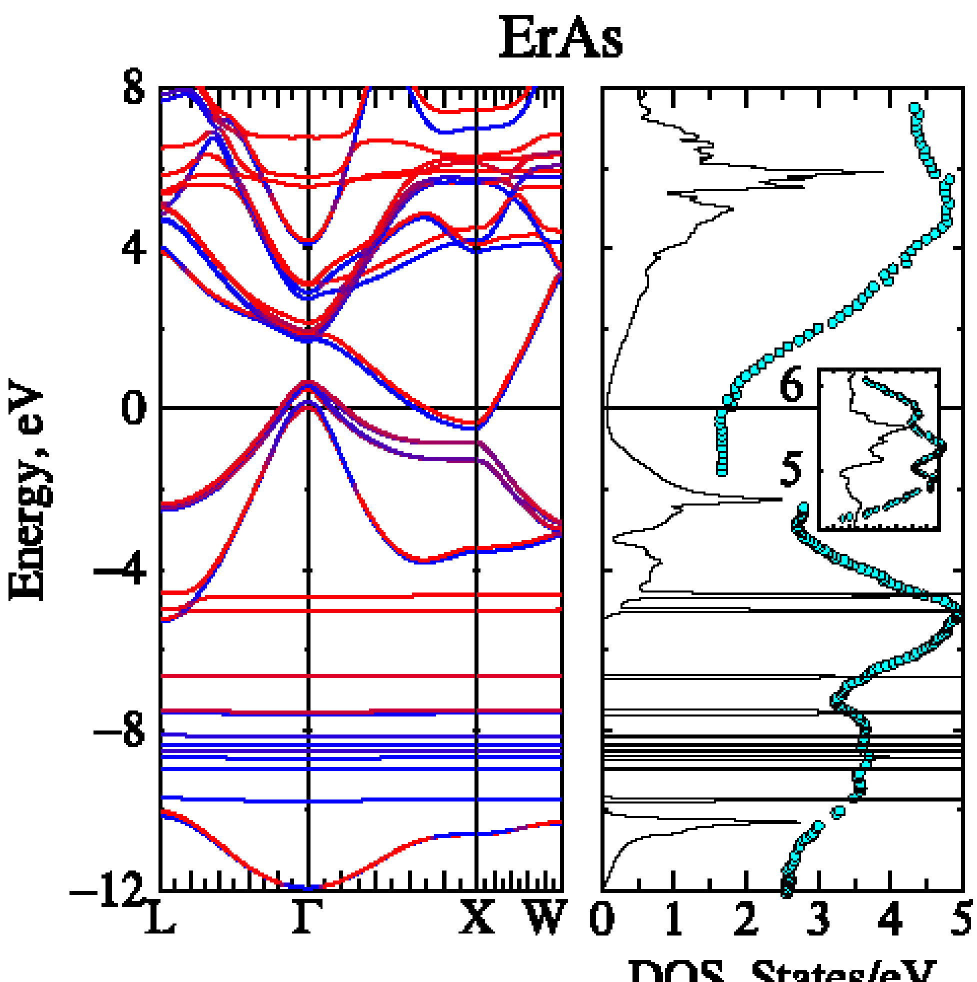
* magnetic moments: small systematic errors (slightly overestimated)

* *d* band exchange splitting and bandwidths are systematically improved relative to LDA.



QSGW theory in 4*f* systems

PRB 76, 165126 (2007)



f subsystem reasonably well described.
Errors very systematic:

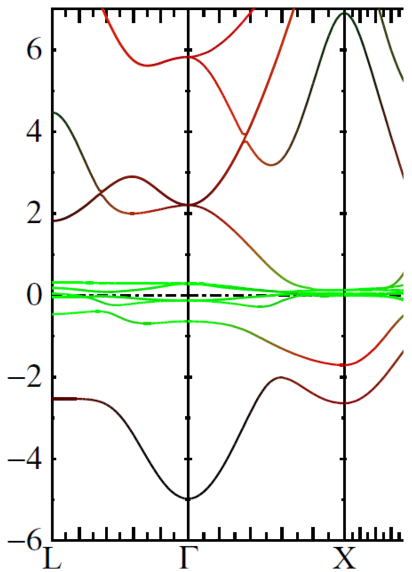
Occupied f states reasonably close to photoemission (missing multiplet structure)

Unoccupied f states systematically too high. Generally true in 4*f* systems.

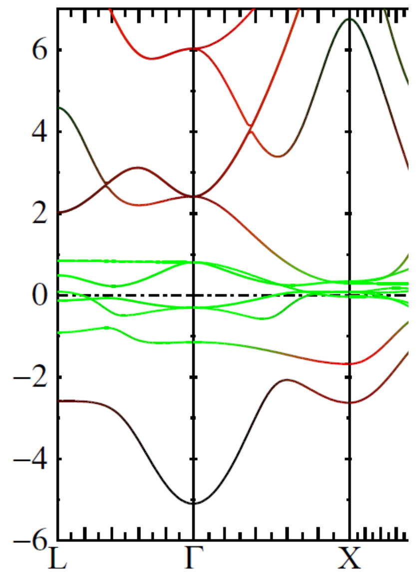
spd subsystem also well described: hole concentrations, masses

QSGW applied to Pu

5*f* bandwidth renormalized by ~2x. Implies one-body, noninteracting hamiltonian quite different than LDA…

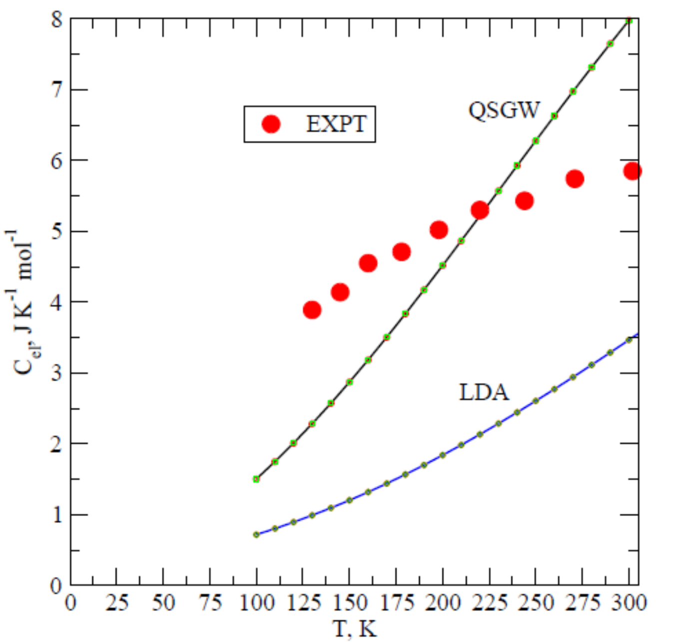


QSGW



LDA

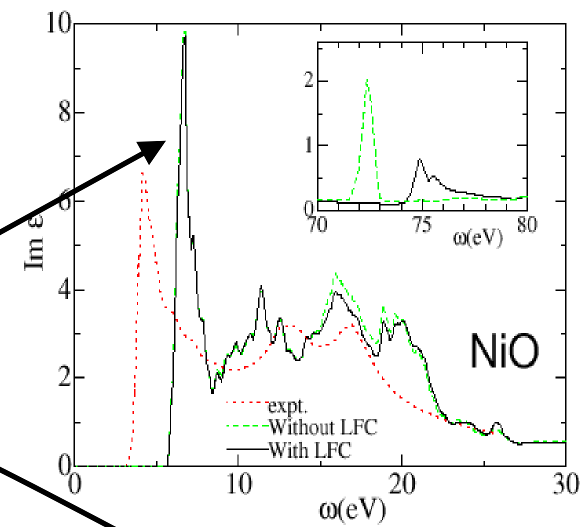
Important implications for LDA+U, LDA+DMFT



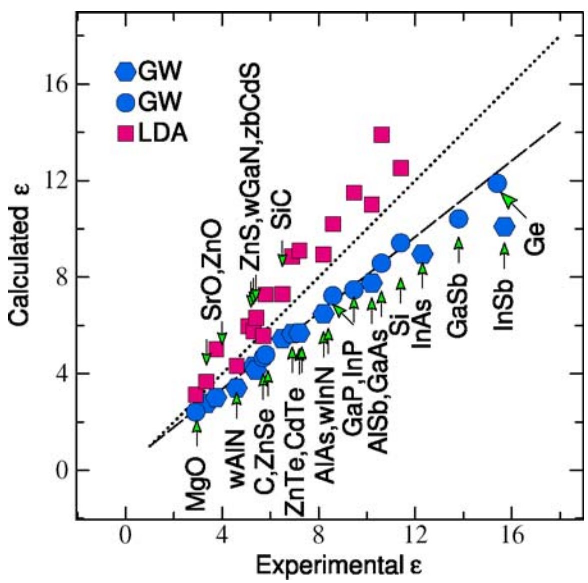
Low-temperature specific heat much changed from LDA. still poor agreement w/ expt. Outside 1-body? (spin fluct)

Systematics of Errors

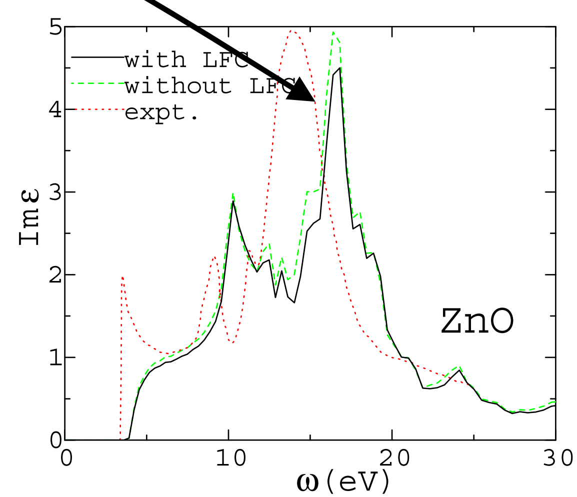
- ✓Unoccupied states universally too high
  - ✓~0.2 eV for *sp* semicond;
  - ✓<~1eV for itinerant *d* $SrTiO_3$, TiO_2
  - ✓>~1eV for less itinerant *d* NiO
  - ✓>~3 eV for *f* Gd,Er,Yb
- ✓Peaks in Im $\varepsilon(\omega)$ also too high



- ✓ε_∞ universally
- ✓20% too small



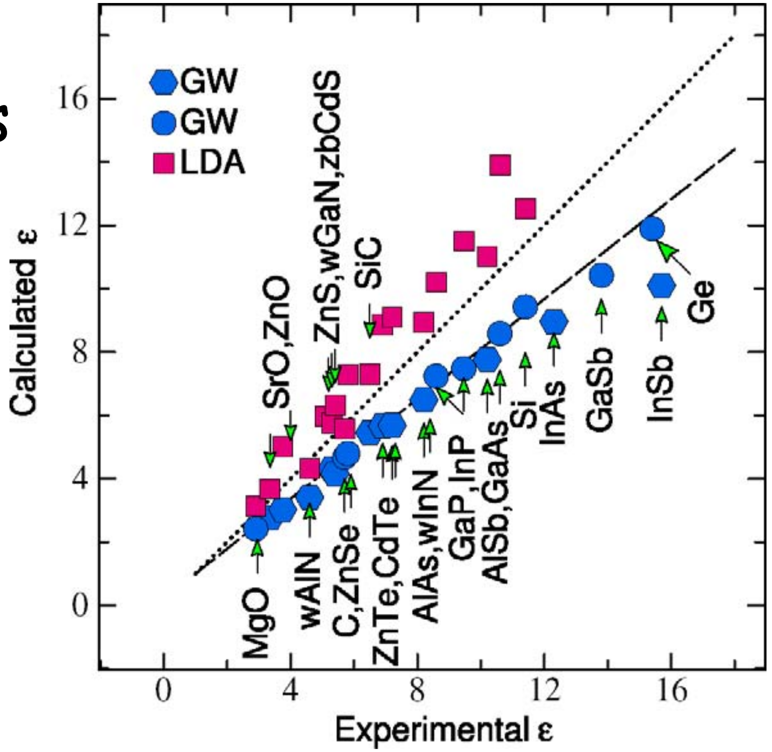
- ✓Magnetic moments slightly overestimated



Likely origin of Errors

Exact theory: $\Sigma = iGW\Gamma$. Requires that *both* Γ *and* W be exact.
Two sources of error:

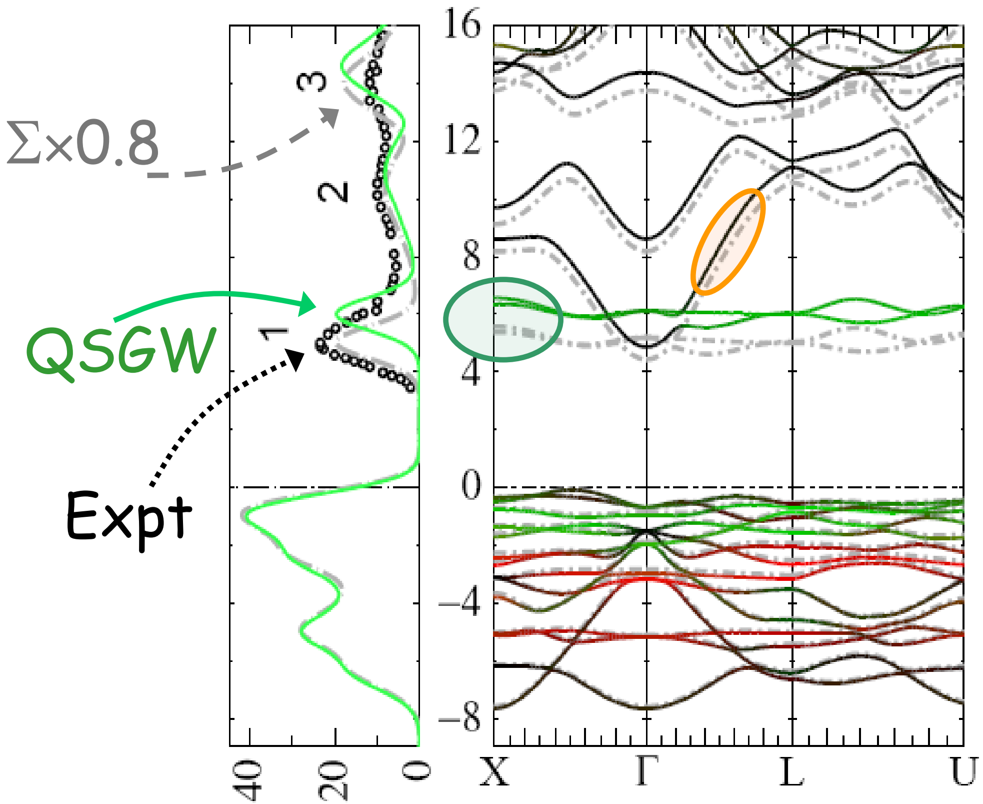
1. Main error: originates from RPA approximation to $\Pi \cong G_0 G_0$:
   ε_∞ is underestimated in insulators by a universal factor 0.8. Thus, $W(\omega=0)$ is too large, roughly by a factor 1/0.8.
   - Accounts for most errors in QP levels, e.g. semiconductor gaps (see Shiskin et al, PRL 99, 246403)

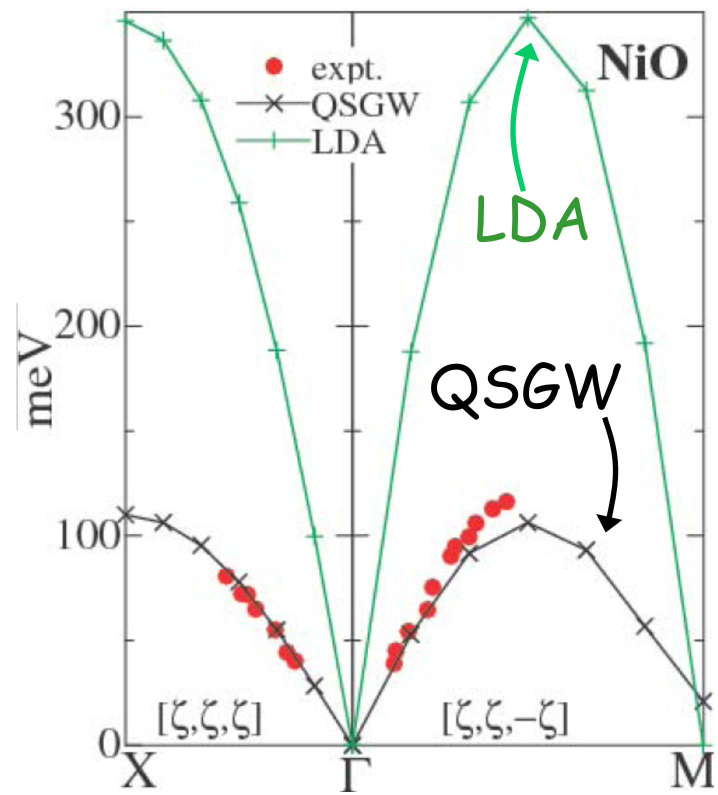


2. Secondary: missing vertex corrections Γ.

NiO: illustration of errors in polarization Π

Bands of both *sp* and *d* character are present
Scaling Σ by 0.8 shifts *sp*- and *d*- characters differently.



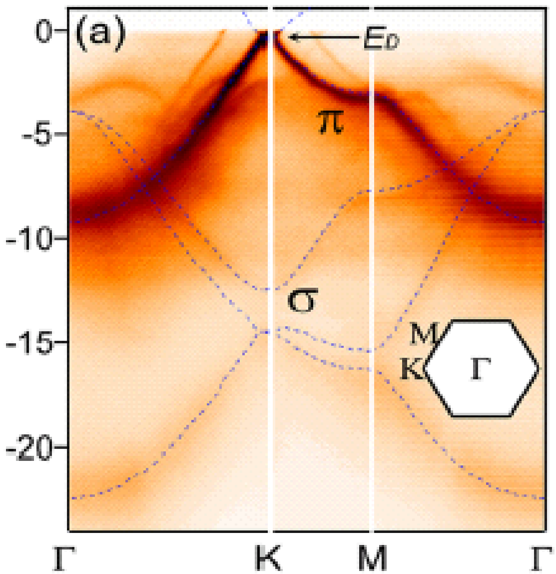


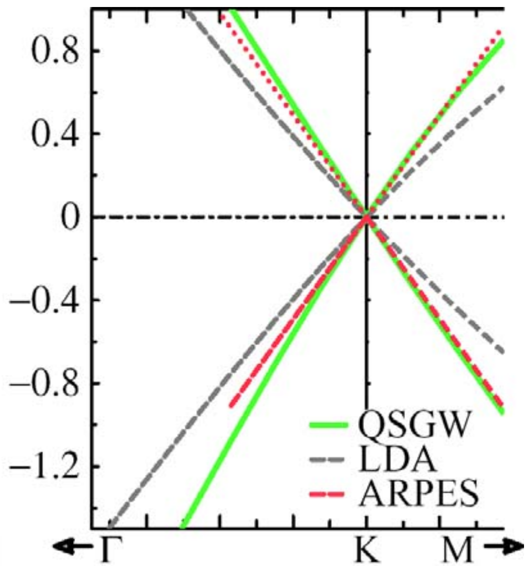
SW spectra from poles of transverse susceptibility are in good agreement with experiment.

skip

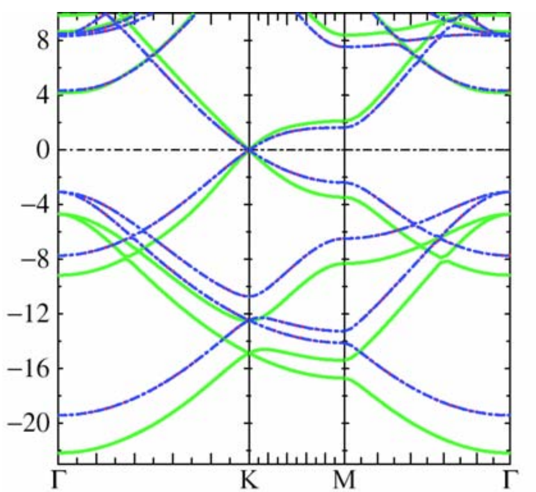
Graphene

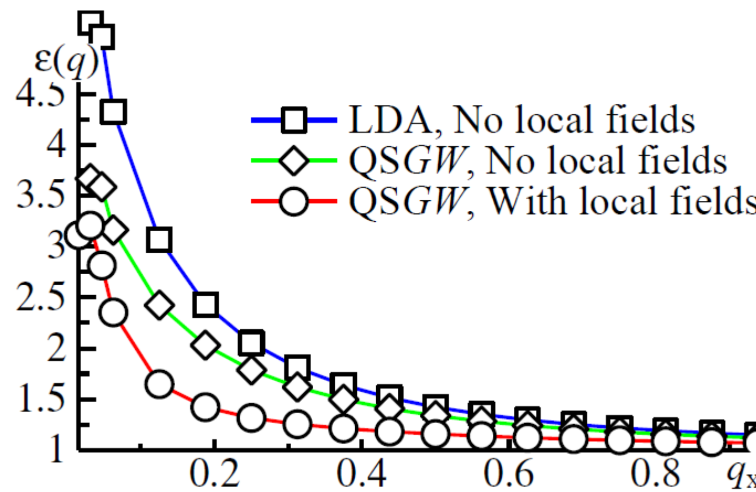
(Ohta *et al*, PRL **98,** 206802 (2007))



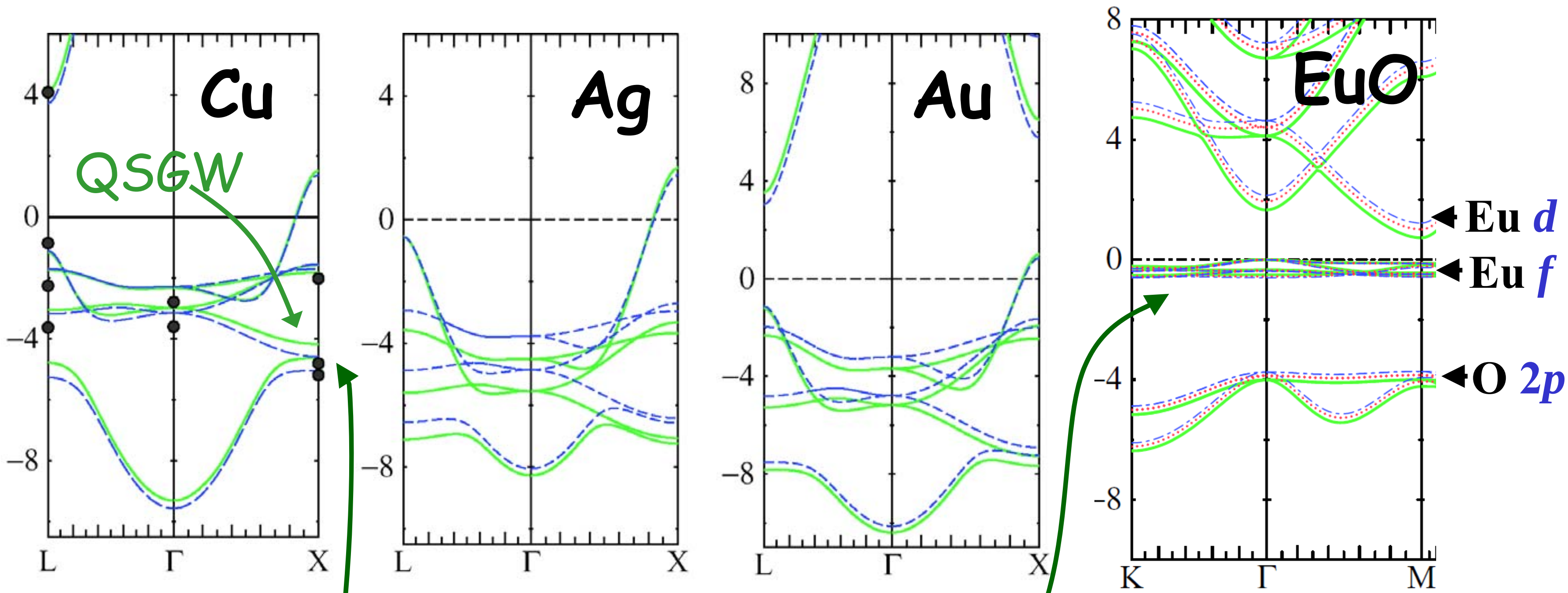


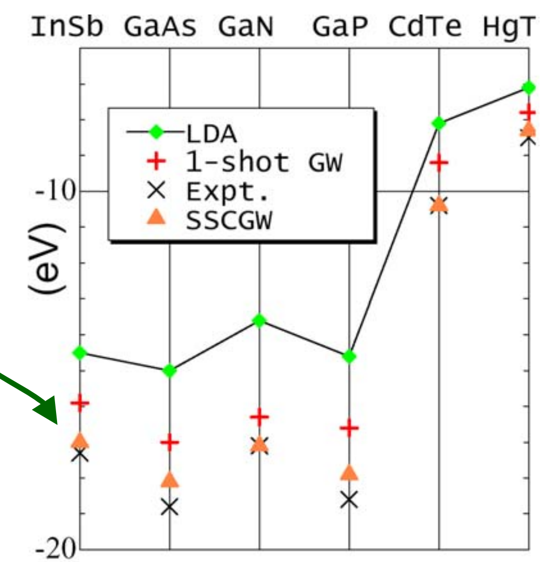
LDA dispersions at K much softer than ARPES.
e-ph renormalization worsens agreement.
QSGW a little too steep: correction of Π should resolve





Errors caused by missing vertex Γ





Localized *d* bands *consistently* ~0.4 eV higher than expt.
Eu $4f$ ~0.7 eV too high
Cannot be explained in terms of errors in $\Pi(\mathbf{q},\omega)$.

Γ At the Si/SiO_2 Interface

Band Offsets at the Si/SiO_2 Interface from Many-Body Perturbation Theory

R. Shaltaf,[1] G.-M. Rignanese,[1] X. Gonze,[1] Feliciano Giustino,[2,3,*] and Alfredo Pasquarello[2,3]

PRL **100**, 186401

GW, *GW*Γ and QS*GW* applied to Si, SiO_2, and junction. Look at bulk compounds first.

−0.24
+0.35
+0.60

−0.21
+0.51
+0.72

	Si			c-SiO_2			s-SiO_2		
	GW	*GW*Γ	QS*GW*	*GW*	*GW*Γ	QS*GW*	*GW*	*GW*Γ	QS*GW*
δE_v	−0.4	+0.1	−0.6	−1.9	−1.3	−2.8	−1.9	−1.3	−2.8
δE_c	+0.2	+0.7	+0.2	+1.5	+1.8	+1.3	+1.4	+1.8	+1.1
δE_g	+0.6	+0.6	+0.8	+3.4	+3.1	+4.1	+3.3	+3.1	+3.9

Authors show effect of Γ on δE_v, δE_c separately not small.
Approximately similar for Si, SiO_2 … is it general?
Γ may be important in correcting *GW* offsets.
Caveat: our own all-electron *GW* and QS*GW* calculations show quite different δE_v, δE_c distribution in Si.

skip

The Si/SiO_2 Valence Band Offset

Authors found that δ(VBM)=(VBM)^{QP} – (VBM)^{DFT} calculated for bulk applies to interface: i.e. interface calculation not necessary to get QP correction to band offset,

PRL **100**, 186401

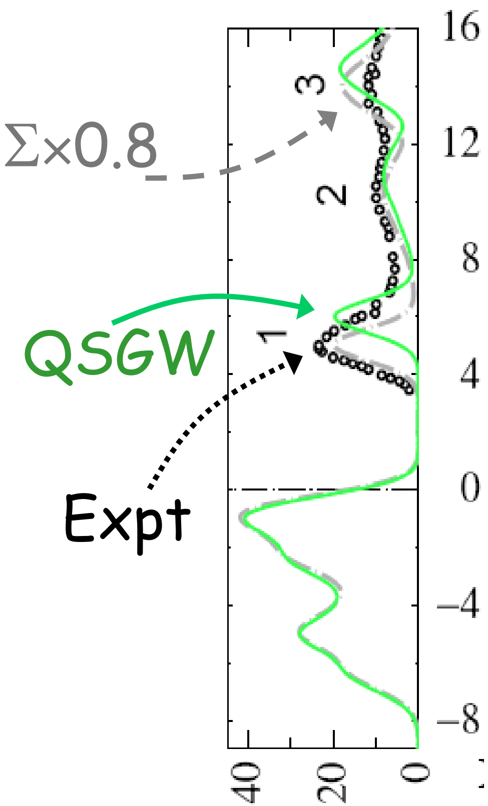
TABLE III. Quasiparticle band offsets (eV) for cubic and strained SiO_2 using *GW*, *GW*Γ, and QS*GW*.

		Cubic				Strained			
	Model	DFT	*GW*	*GW*Γ	QS*GW*	*GW*	*GW*Γ	QS*GW*	Expt.
VBO	I	2.6	4.1	4.0	4.8	4.1	4.0	4.8	4.3
	II	2.5	4.0	3.9	4.7	4.0	3.9	4.7	
CBO	I	1.6	2.9	2.7	2.7	2.8	2.7	2.5	3.1
	II	1.8	3.1	2.9	2.9	3.0	2.9	2.7	

Their *GW* and *GW*Γ results are very similar, rather good.
QS*GW* VBM a little worse: VBM(QS*GW*) = VBM(Expt) + 0.5 eV
Caveat: all electron results certain to be different (cf Si).
Known problems with PP-based *GW* [Gómez-Abal, Li, Scheffler, Ambrosch-Draxl, Phys. Rev. Lett. 101, 106404]

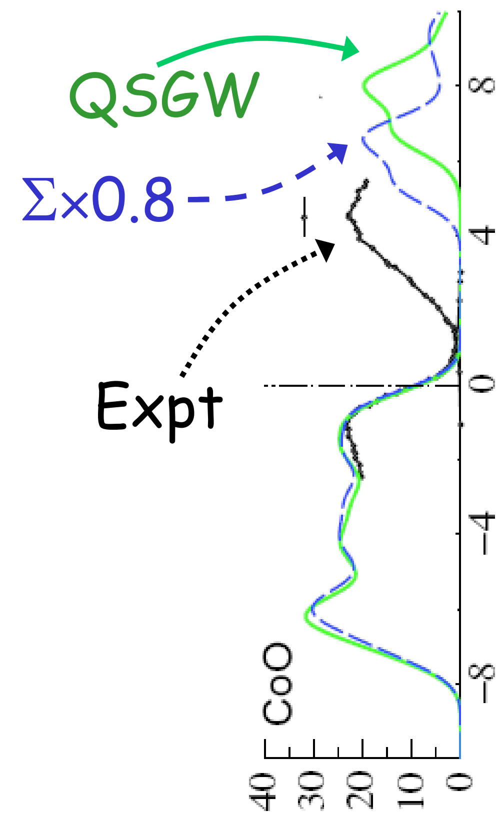
NiO vs CoO

NiO: QSGW misses satellites and subgap excitations arising from internal *dd* transitions.
But QP picture dominates electronic structure; these effects are small perturbations to QP picture.



NiO: Scaling Σ by 0.8 yields very good agreement with both PE and BIS measurements.

CoO, FeO, Ce_2O_3 :
situation less rosy. Substantial disagreement with BIS. Splitting within a single spin channel.



Conclusions

- The QSGW approximation
  - Self-consistent perturbation theory; self-consistency used to minimize the size of the (many-body) perturbation
    - optimum partitioning between H_0 and $\Delta V=H-H_0$.
  - QSGW has some formal justification and it works very well in practice! A true ab initio theory that does not depend on any scheme based on ansatz, e.g. LDA, LDA+U
  - Reliably treats variety of properties in a wide range of materials: The errors are systematic and understandable.

QSGW is well positioned to become a reliable framework, which can address both many-atom and correlated systems