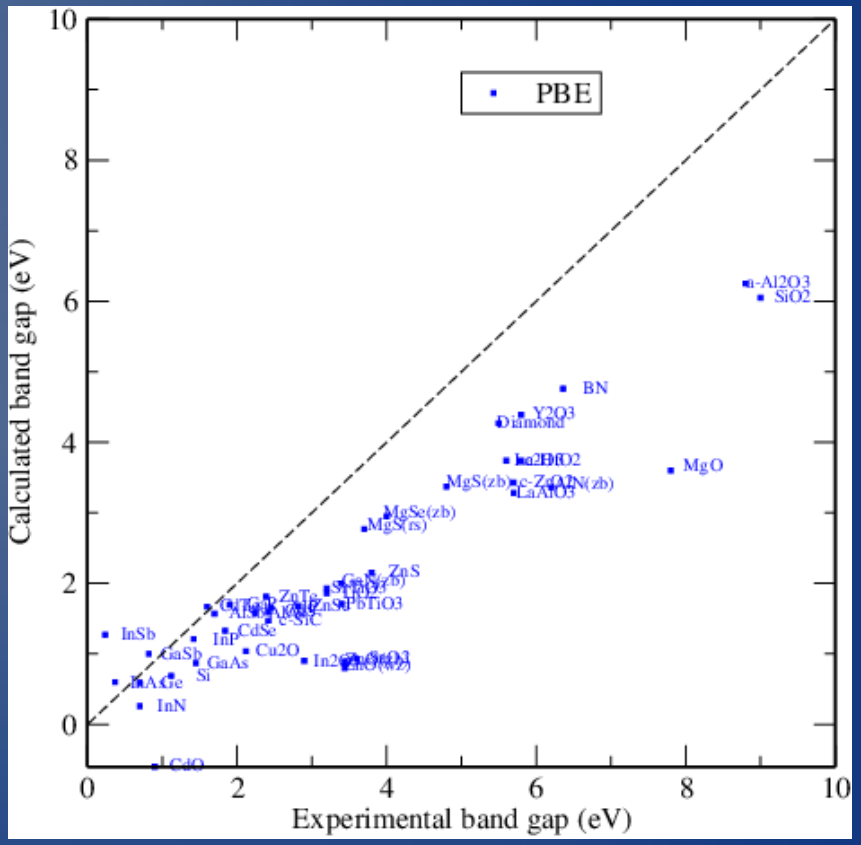
Electronic band structure, sX-LDA, Hybrid DFT, LDA+U and all that

Keith Refson
STFC Rutherford Appleton Laboratory

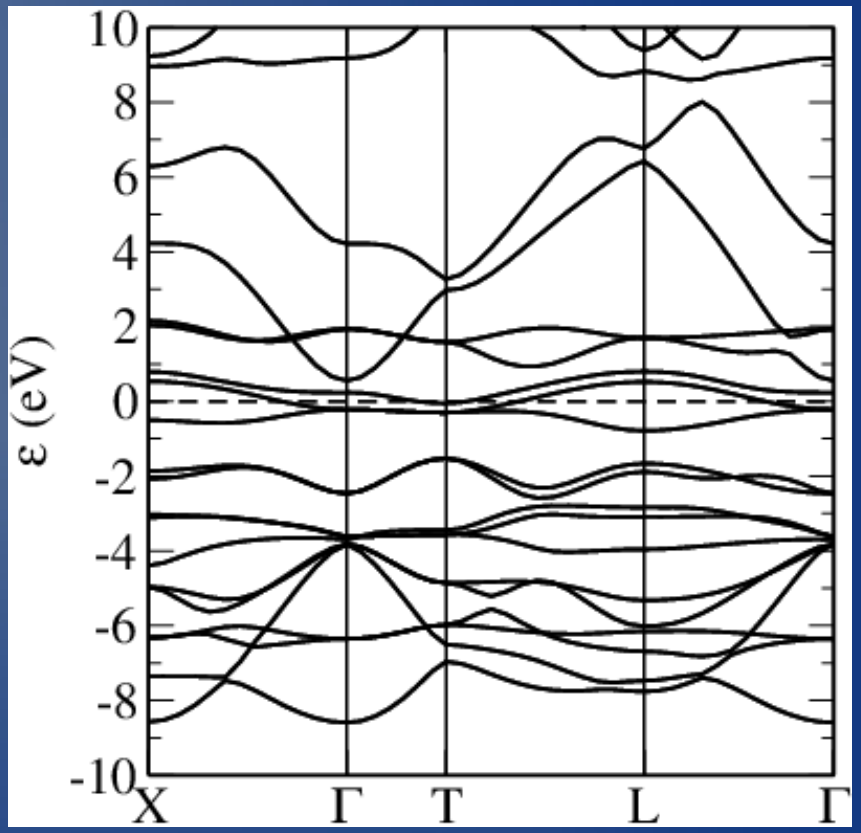
LDA/GGA DFT is good but ...

- Naive LDA/GGA calculation severely underestimates band-gaps.
- Almost universal effect in wide-gap insulators and semiconductors.

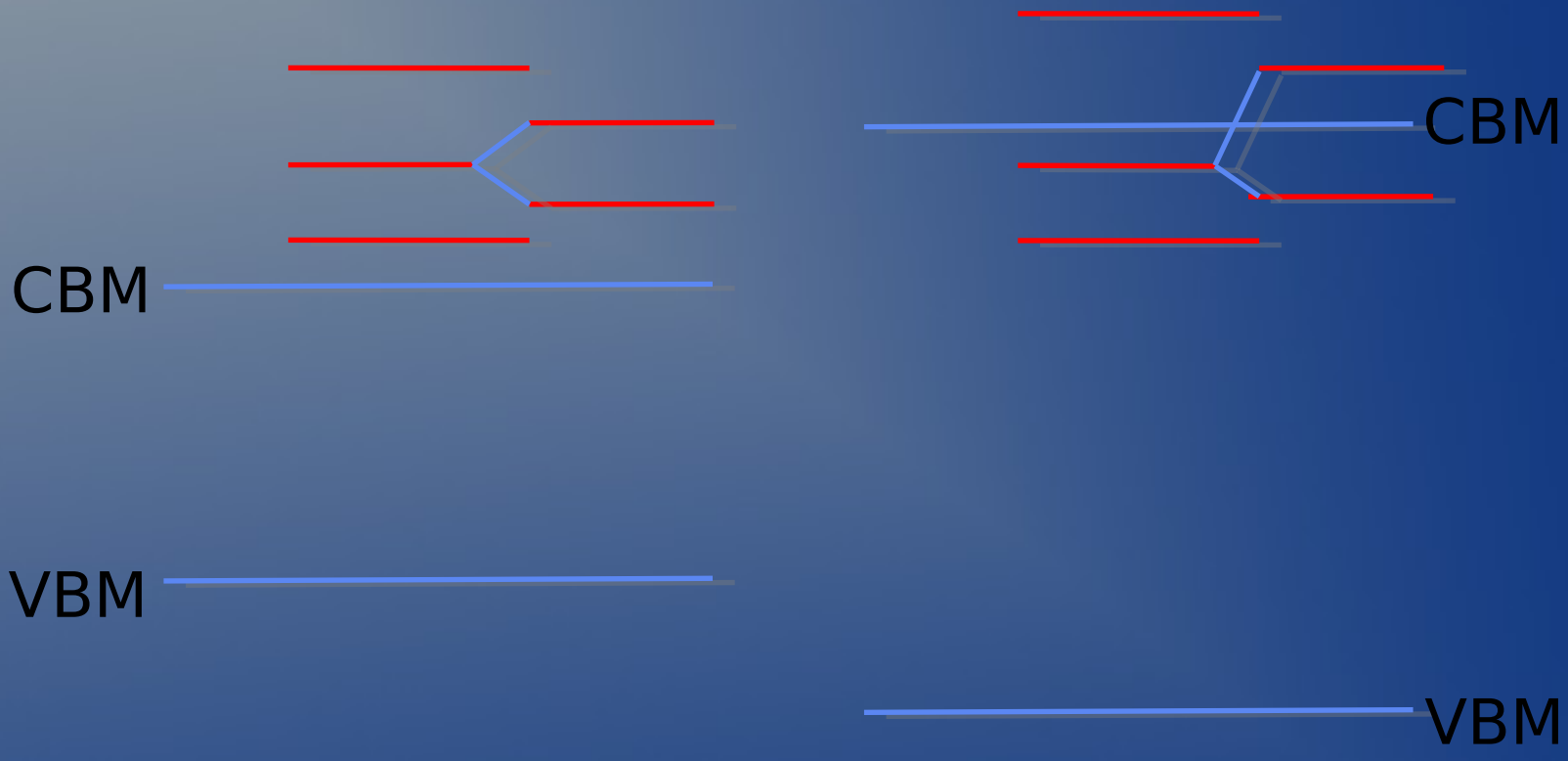


Transition-metal Oxides

- TMOs are usually insulators
- FeO and many others predicted to be metallic in LDA/GGA
- Many so-called “strongly-correlated” oxides falsely predicted to be metallic



Energy Levels of Defects



Local Functional (PBE) Correct levels

Origin of band-gap problem

- DFT is only a ground state theory?

  But band gap is a ground state property

$$E_{gap} = I - A = \{E(N-1) - E(N)\} - \{E(N) - E(N+1)\}$$

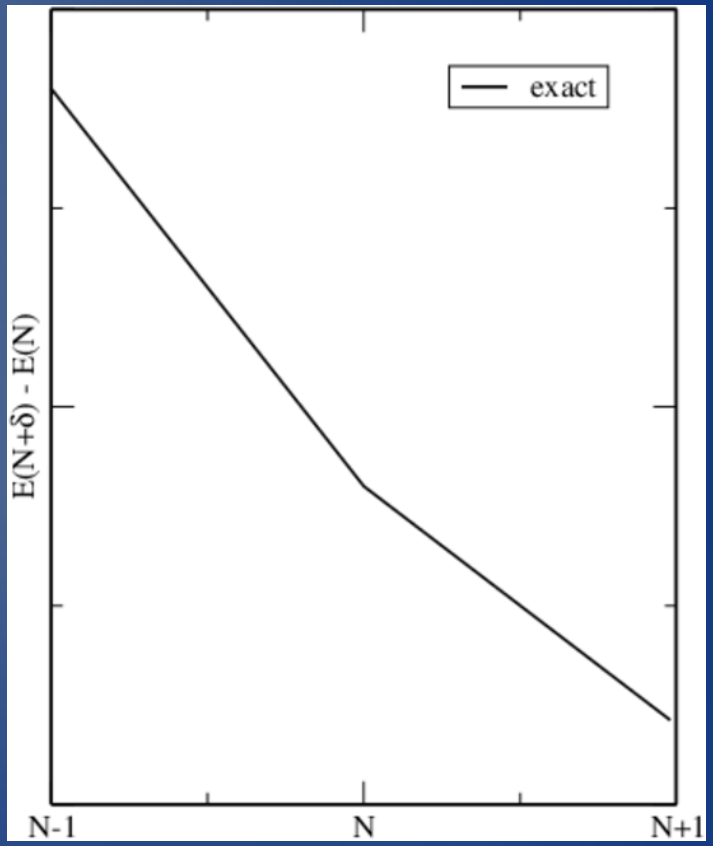
- Kohn-Sham states are not physical excitations – introduced to approximate K.E. by that K.E. of non-interacting system?
- XC potential V_{XC} has a fundamental discontinuity as function of occupancy not reproduced by LDA/GGA.

The Fundamental Gap

- The fundamental gap is the IP – EA
- $E_{gap}^{N} = I - A = \{E(N-1) - E(N)\} - \{E(N) - E(N+1)\}$
- Or alternatively
- $E_{gap}^{d} = \left[\frac{\partial E}{\partial N}\right|_{N+\delta} - \left[\frac{\partial E}{\partial N}\right|_{N-\delta}$
- In the exact case these are equal.
- Therefore E(N) is composed of straight line segments
- For LDA/GGA $E_{gap} = \epsilon_{KS}^{gap} = \epsilon_{HOMO}^{KS} - \epsilon_{LUMO}^{KS}$
- Therefore for LDA/GGA $\Delta_{XC} = 0$
- If Exc has orbital dependence
- $E_{gap}^{N} = \epsilon_{KS}^{gap} + \left[\frac{\partial E}{\partial N}\right|_{N+\delta} - \left[\frac{\partial E}{\partial N}\right|_{N-\delta}$

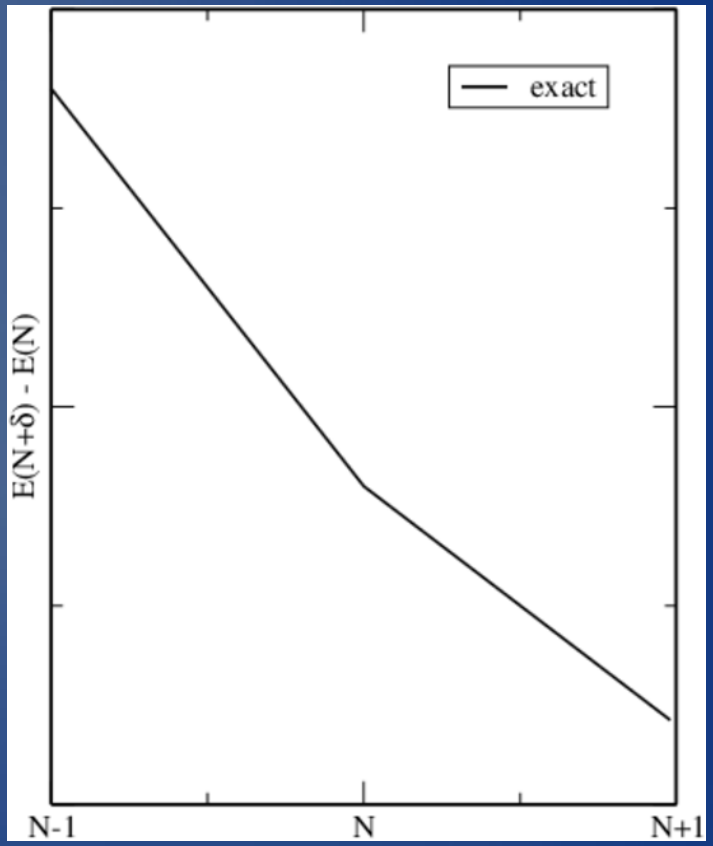
Derivative Discontinuity

- Bandgap error related to energy/fractional occupancy curve (Perdew PRL 49, 1691 (1982)



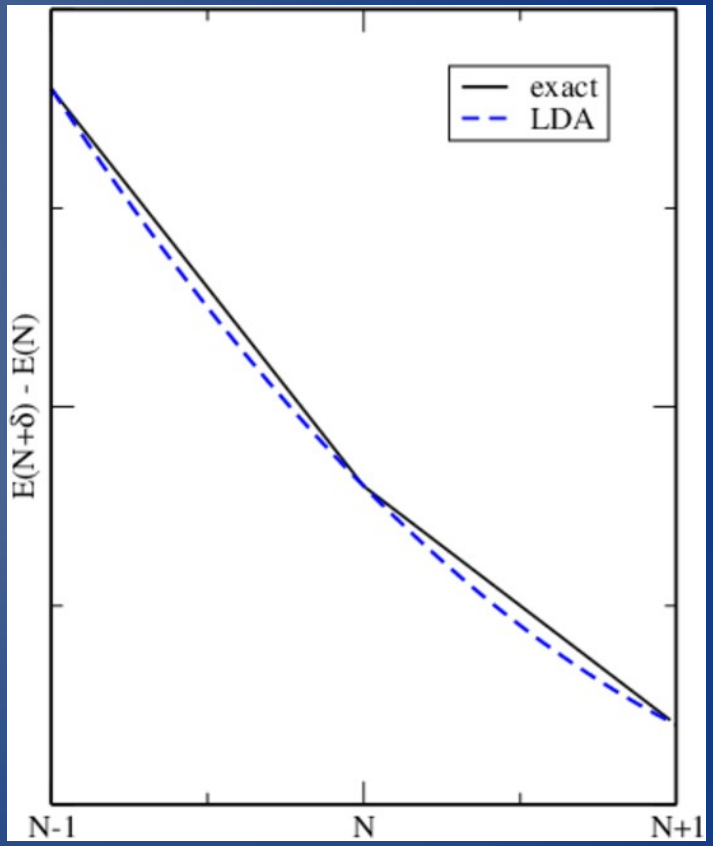
Derivative Discontinuity

- Bandgap error related to energy/fractional occupancy curve (Perdew PRL 49, 1691 (1982)
- exact energy/occupancy curve linear with +ve derivative discontinuity $\Delta_{XC}>0$



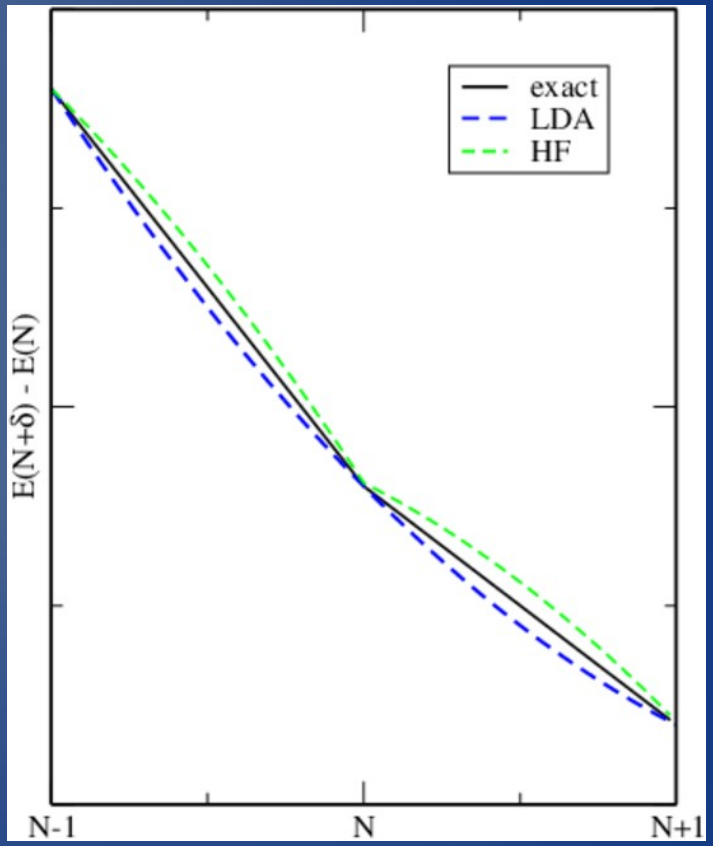
Derivative Discontinuity

- Bandgap error related to energy/fractional occupancy curve (Perdew PRL 49, 1691 (1982)
- exact energy/occupancy curve linear with +ve derivative discontinuity $\Delta_{XC} > 0$
- LDA/GGA form is convex with zero derivative discontinuity. $\Delta_{XC} = 0$



Derivative Discontinuity

- Bandgap error related to energy/fractional occupancy curve (Perdew PRL 49, 1691 (1982)
- exact energy/occupancy curve linear with +ve derivative discontinuity $\Delta_{XC}>0$
- LDA/GGA form is convex with zero derivative discontinuity. $\Delta_{XC}=0$
- HF form is concave with -ve discontinuity. $\Delta_{XC}<0$



Self-interaction error

- Consider Hartree (Coulomb) energy term in Hamiltonian

$$E_H = \frac{1}{2}\iint d^3\boldsymbol{r}\, d^3\boldsymbol{r}\,' \frac{n(\boldsymbol{r})n(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} = \frac{1}{2}\sum_{i,j}\iint d^3\boldsymbol{r}\, d^3\boldsymbol{r}\,' \frac{\psi_i^*(\boldsymbol{r})\psi_i(\boldsymbol{r})\psi_j^*(\boldsymbol{r}')\psi_j(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}$$

- Includes self-interaction contribution

$$E_H^{\text{self}} = \frac{1}{2}\sum_{i}\iint d^3\boldsymbol{r}\, d^3\boldsymbol{r}\,' \frac{\psi_i^*(\boldsymbol{r})\psi_i(\boldsymbol{r})\psi_i^*(\boldsymbol{r}')\psi_i(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}$$

- Exactly cancelled by i=j terms of Hartree Fock exchange

$$E_X = -\frac{1}{2}\sum_{i,j}\iint d^3\boldsymbol{r}\, d^3\boldsymbol{r}\,' \frac{\psi_i^*(\boldsymbol{r})\psi_j(\boldsymbol{r})\psi_i^*(\boldsymbol{r}')\psi_j(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}$$

- But only 93% cancelled by LDA exchange (in H atom)
- LDA electron “sees” extra Coulomb repulsion by own charge

  => can reduce energy by unphysical delocalisation.
- d-bands in TMs too wide, and too high in energy, closing gap

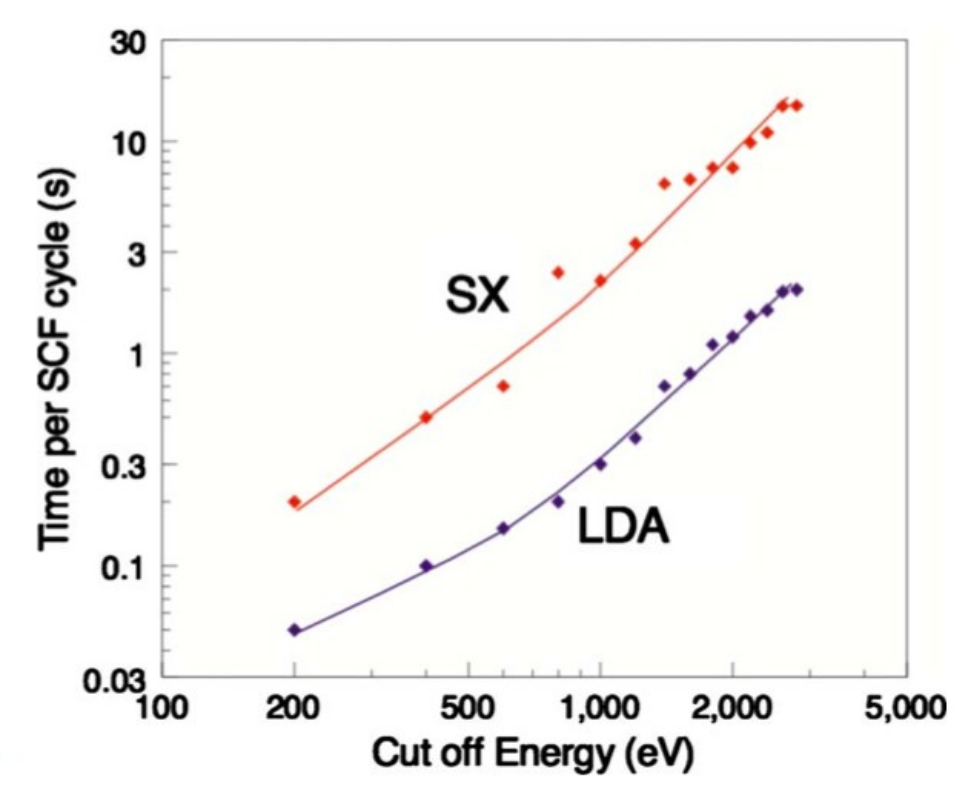
Non-local functionals

- Now well-known and commonly used zoo of non-local functionals (HSE, PBE0, B3LYP)
- Mix a fraction of non-local HF with local correlation
- Advantages:
  - They correct (more or less!) the band gap underestimate
  - They are energy functionals in the DFT sense
  - Can be used in the favoured energy minimisation techniques for SCF solution
  - Can calculate forces (important!)
  - Appear to correct the localisation error
- Disadvantages
  - Computationally expensive
  - Still investigating how well they perform (still rather empirical)

Computational Expense

$$-\frac{1}{2}\nabla^2\psi_i(\boldsymbol{r})+V_{loc}(\boldsymbol{r})\psi_i(\boldsymbol{r})+\int V_{XC}^{NL}(\boldsymbol{r},\boldsymbol{r}')\psi_i(\boldsymbol{r}')d^3\boldsymbol{r}'=\epsilon_i\psi_i(\boldsymbol{r})$$

- Require solutions of the non-local Schrödinger equation
- VNL incorporates HF-like terms
- Extra integral over r' multiplies computational cost with a plane-wave basis set
- Actually solve in recip space using FFT convolution.



Overall scaling $T \propto N_{PW} \times N_k^2 \times N_b^2$

Hybrid Functionals

PBEh (PBE0): simple mixture of exact exchange

$$E_{xc}^{PBEh} = \alpha E_x^{HF} + (1-\alpha) E_x^{PBE} + E_c^{PBE}; \alpha = 1/4$$

B3LYP: more complex mixture of exact exchange

HSE: range-separated version

$$\frac{1}{r} = \mathrm{SR} + \mathrm{LR} = \frac{erfc(\omega r)}{r} + \frac{erf(\omega r)}{r}$$

$$E_{xc}^{\omega PBEh} = \alpha E_x^{HF,SR}(\omega) + (1-\alpha) E_x^{PBE,SR}(\omega) + E_{LR}^{PBE}(\omega) + E_c^{PBE}$$

Screened Exchange

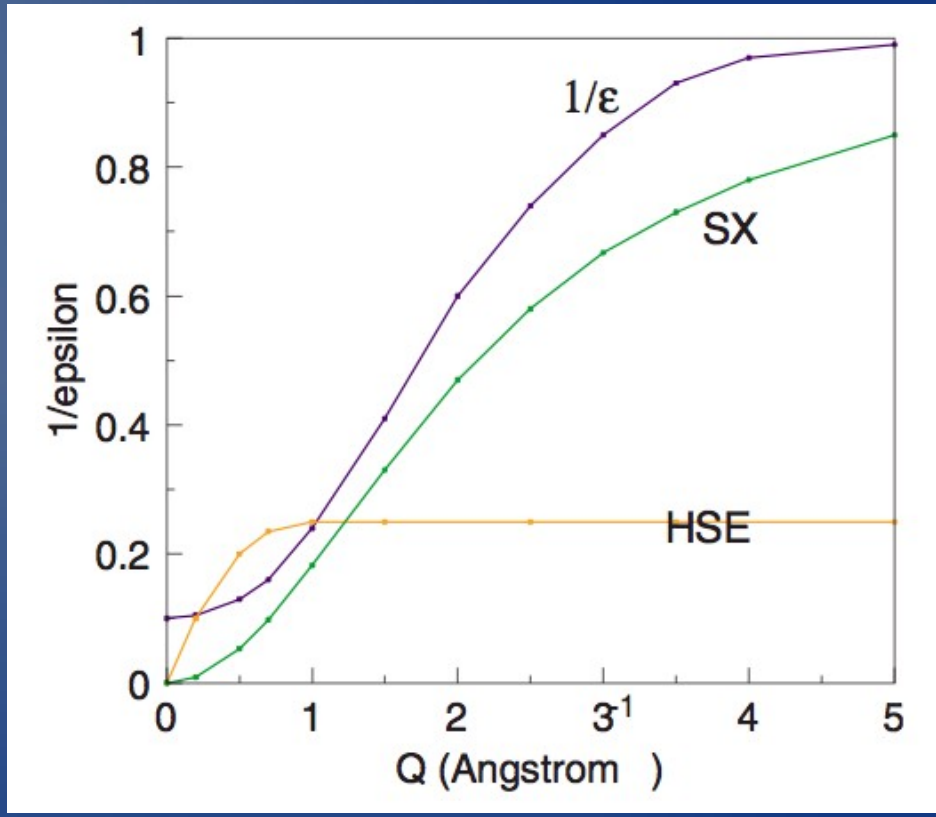
- The sX potential is non-local
- Replaces the LDA with short-ranged Thomas-Fermi screened Coulombic exchange potential

$$V_j^{sX}(\boldsymbol{r},\boldsymbol{r}') = -\sum_i \frac{\psi_i(\boldsymbol{r})\exp[-(k_s|\boldsymbol{r}-\boldsymbol{r}'|)]\psi_j^*(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} + \epsilon_{loc}^{LDA}(n) - \int V_X^{LDA}(n)F(n)n(\boldsymbol{r})d\boldsymbol{r}$$

- Here ks is the Thomas-Fermi screening length where the density of valence (non-d) electrons is screened
- See Clark and Robertson, PRB 82 085208 (2010) for details

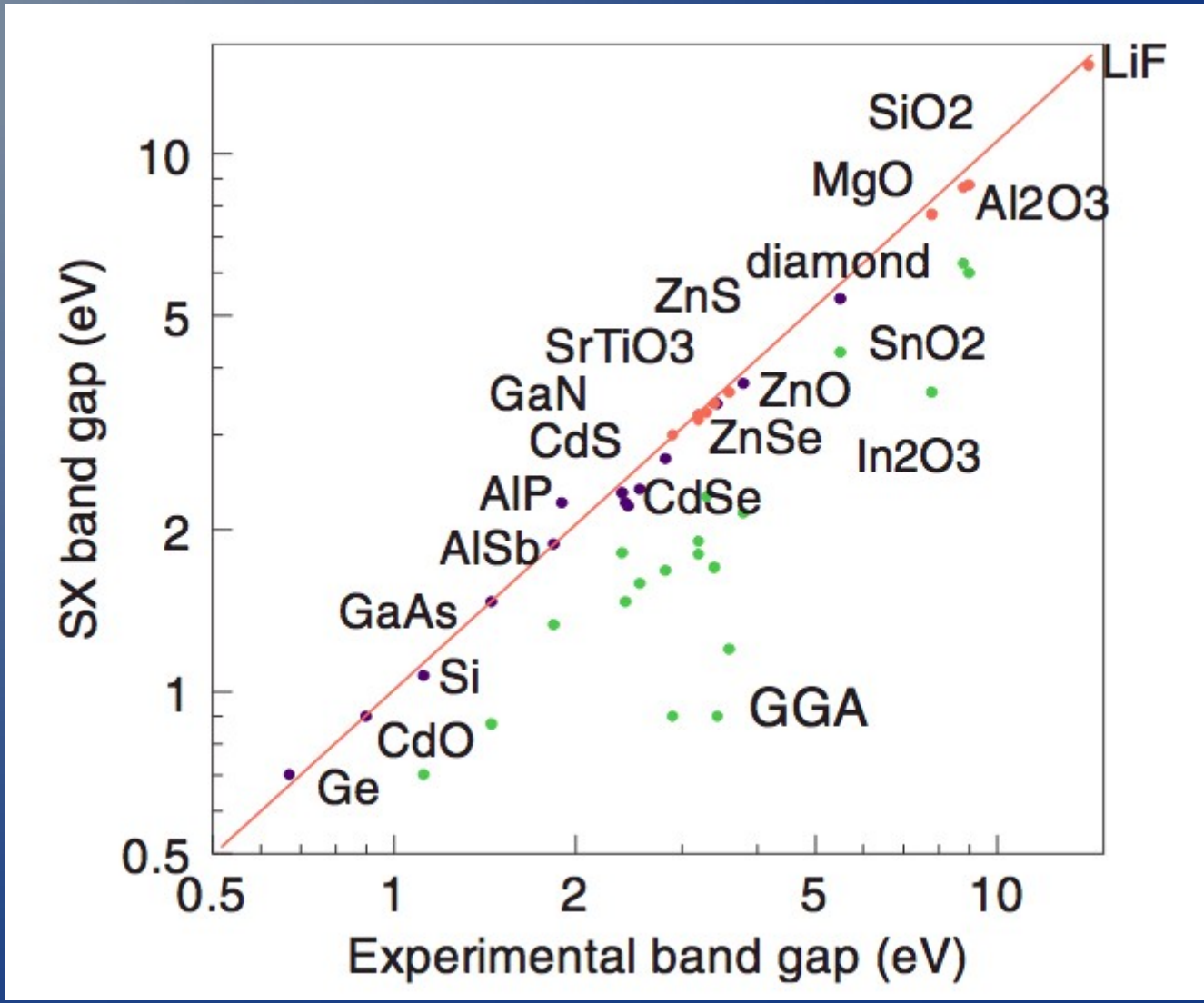
Screening

- Comparison of screening in HSE and sX
- sX has the correct asymptotic limits in the free-electron gas
- Note that the screening is similar to the inverse dielectric function (in this case GaAs)
- This is expected with TF screening (zeroth order approximation to semiconductor's screening)



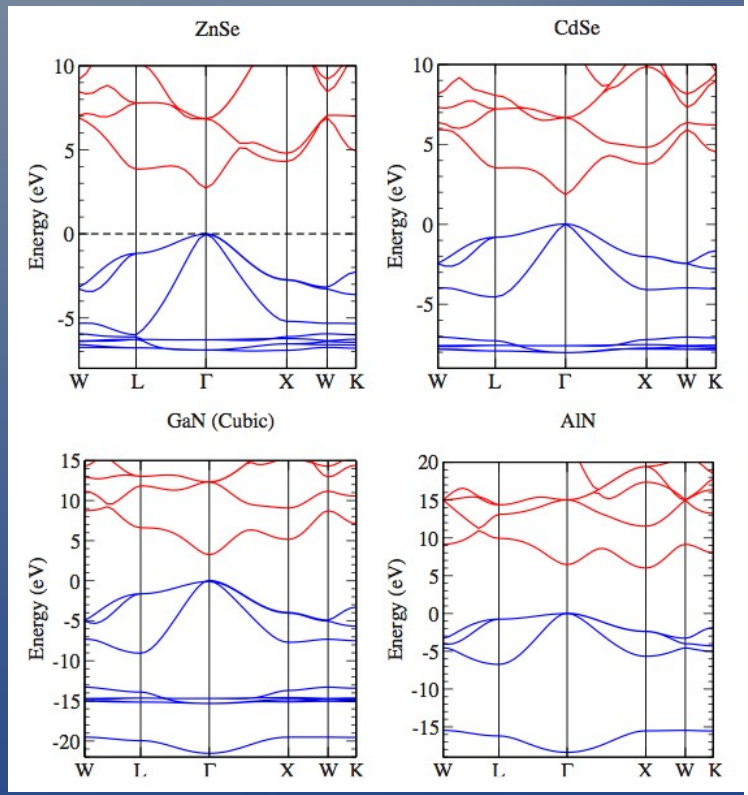
sX Band Gaps

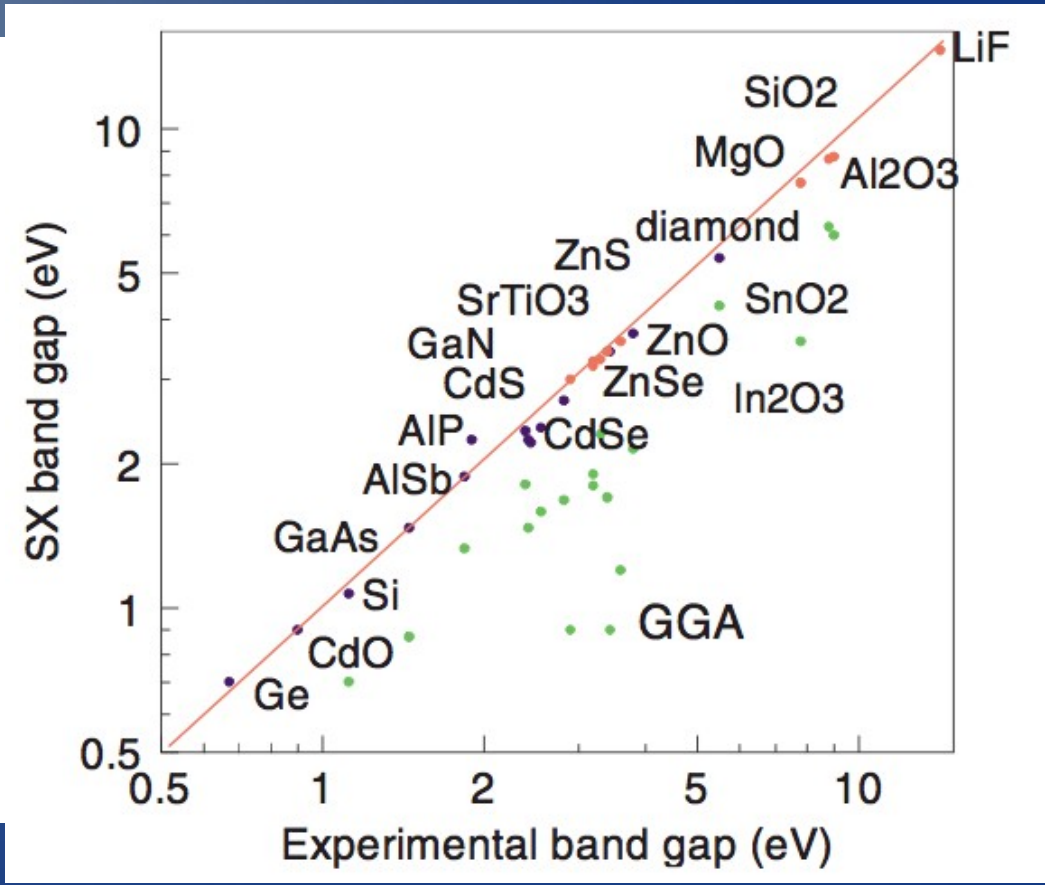
- Closely related HSE (HSE06) gaps are fairly similar
- Functional forms are different but both have a short-ranged screening of the HF component



SX bandgaps

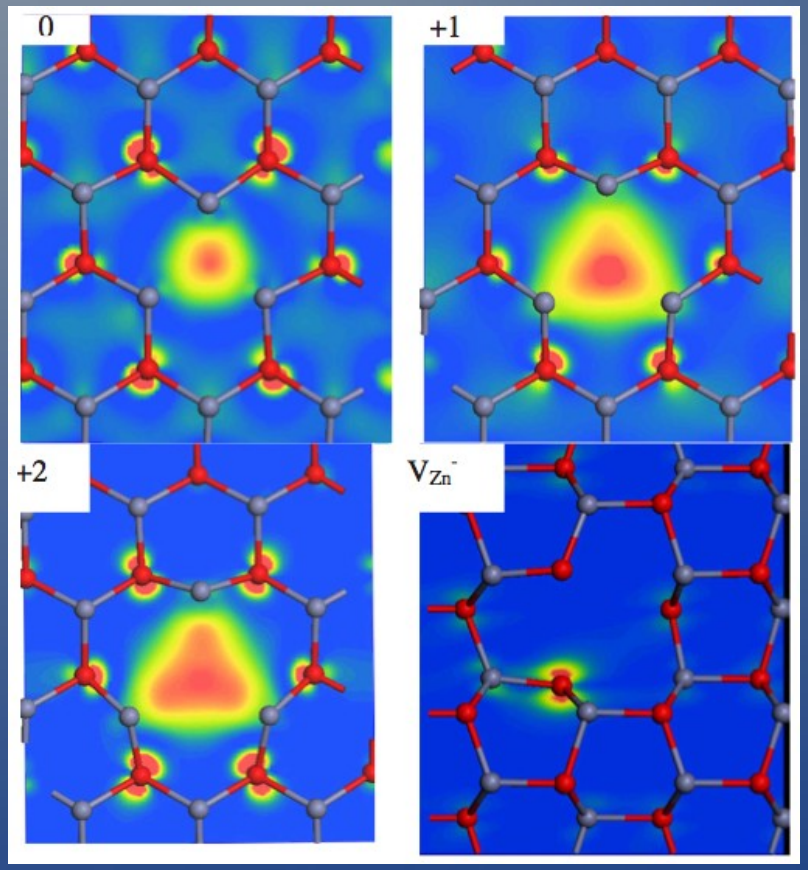
DFT and excitation energies are now a reality!

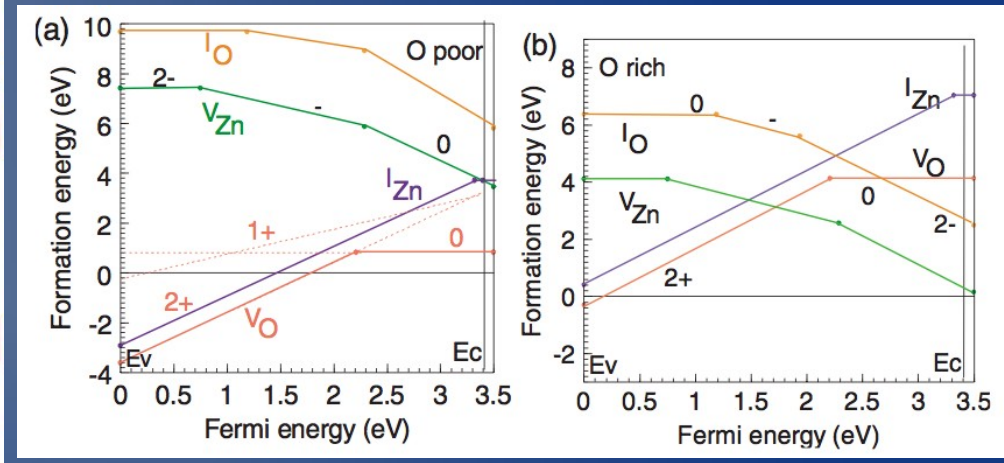




Clark & Robertson, PRB 82, 085208 (2010)

ZnO Defects





- Accurate prediction of:
  - Geometry of defects
  - Band Gap
  - Formation Energies

Clark, Zunger, et al., PRB 81, 115311 (2010)

CASTEP implementation

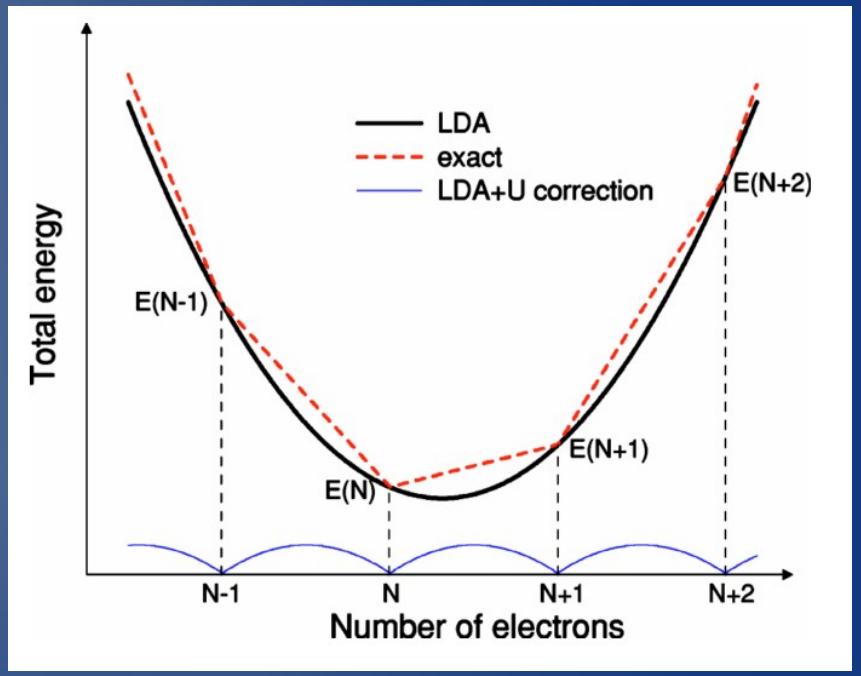
- CASTEP has implemented sX-LDA, PBE0, B3LYP
- Currently restricted to norm-conserving pseudopotentials.
- Only LDA or GGA pseudopotentials are available or can be generated.
- Density mixing scheme does not work: use FIX_OCCUPANCY mode, and allbands/EDFT SCF solver.
- To accelerate k-point convergence turn on divergence correction (Gygi and Baldereschi, PRB 34, 4405 (1986)
- Still very expensive compared to LDA/GGA

Adding an on-site Hubbard U

- Hubbard model favours integer occupancies
- Add term to Hamiltonian modelling on-site repulsion scaled by “U”

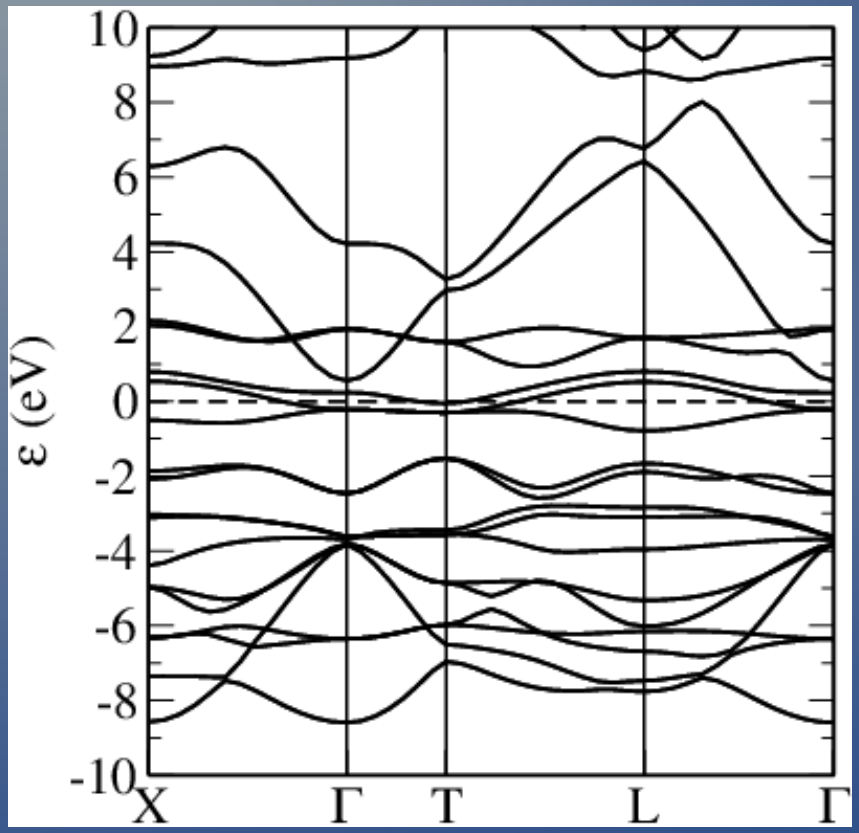
$$E_U = \frac{U}{2} \sum_{l,\sigma} Tr\left[n^{l,\sigma}\left(1 - n^{l,\sigma}\right)\right]$$

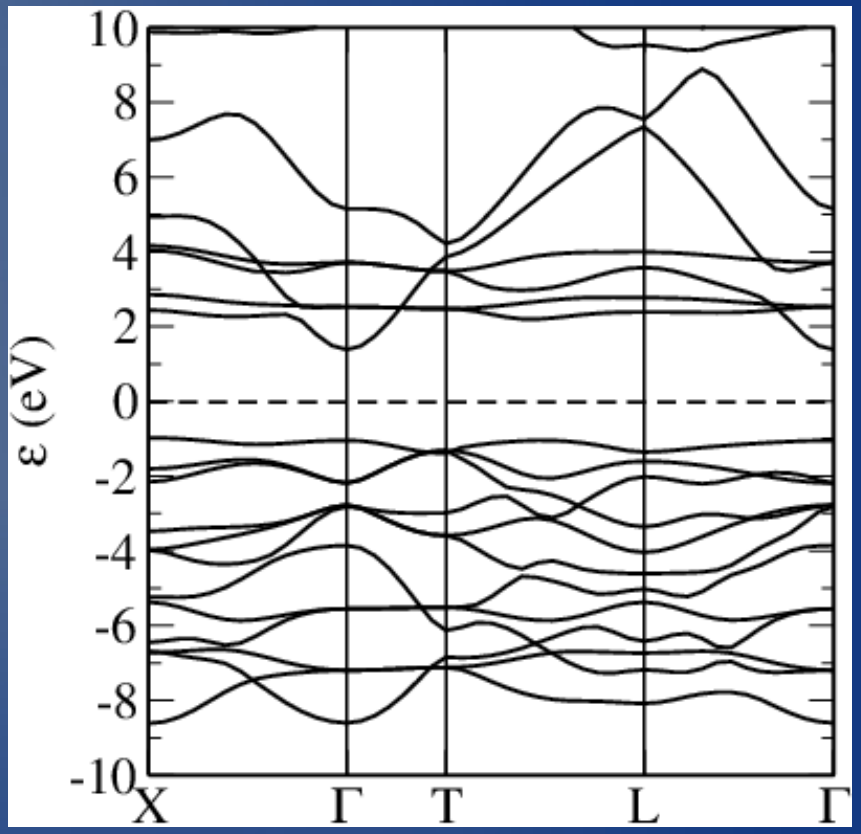
- U may be derived for bulk system by linear response
- No nonlocal integrals required - much cheaper than exact exchange.



Cococcioni et al, PRB 71 035105

Effect of U on FeO





- GGA+U opens gap between unoccupied and occupied d-states in open-shell systems
- Does not give good band gaps for closed-shell systems.

Hybrid, sX, LDA+U merits and demerits

- Hybrid functionals can correct band-gaps, but do fail to reproduce correct screening. Poor performance for metallic systems. Reasonable for TMOs. Very expensive.
- SX-LDA with TF screening length good for bulk, closed-shell systems. Screening model good for large Q. But what screening length to use for structured model systems? Expensive.
- LDA+U correctly splits partially occupied d- and f-shells in magnetic system. Some correction of self-interaction error, but worse bandgaps than sX and hybrids. Poor bandgaps for closed-shell systems with sane U.
- ALL these are MODEL methods (not ab initio) and all involve some parameterization.