Réunion general GDR NBODY
(Nearly in) Toulouse, Janvier 2022

Many thanks Titou, Pina, Julien, Michel.

cnrs GDR Groupement de recherche
NBODY Quantum n-body problem in chemistry and physics

All-electron space-time formalism for efficient cubic scaling *GW* calculations

X. Blase

Institut Néel, CNRS and University Grenoble Alpes, France.

NÉEL institut

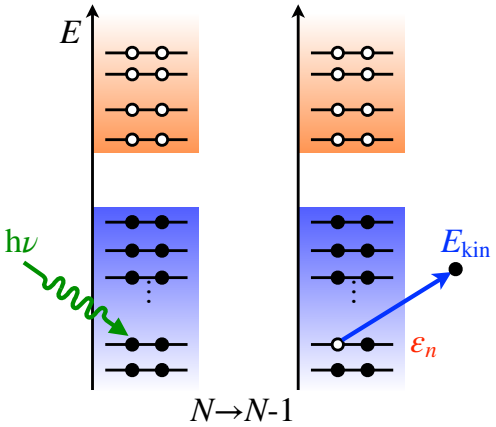
Ivan Duchemin
(LSIM/IRIG/CEA-Grenoble)

The GW approximation in a nutshell

Lars Hedin, Phys. Rev. 139, A796 (1965)

- GW calculations tackle the calculation of electronic energy levels as given by a photoemission experiment (electron addition /removal energies)

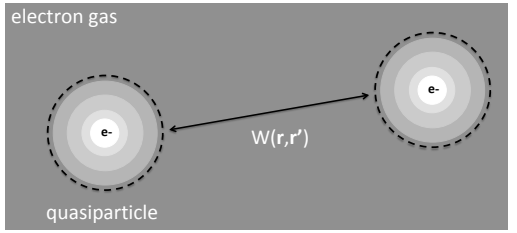
- G is the time-ordered one-body Green's function with poles at the proper electron addition/removal energies



$$G(\mathbf{r},\mathbf{r}';\omega)=\sum_n \frac{f_n(\mathbf{r})f_n^*(\mathbf{r}')}{\hbar\omega-\varepsilon_n+i\eta\hbar\,\mathrm{sgn}(\varepsilon_n-\mu)} \quad \Longleftrightarrow \quad \begin{aligned}\varepsilon_n &= (E_n^{N+1}-E_0^N) \qquad (\varepsilon_n>\mu)\\ \varepsilon_n &= (E_0^N-E_n^{N-1}) \qquad (\varepsilon_n>\mu)\end{aligned}$$

- W is the screened Coulomb potential

$$W(1,2)=v(1,2)+\int d34\; v(1,3)\tilde{\chi}(3,4)W(4,2)$$



- GW : lowest approximation in *W* for the exchange-correlation self-energy

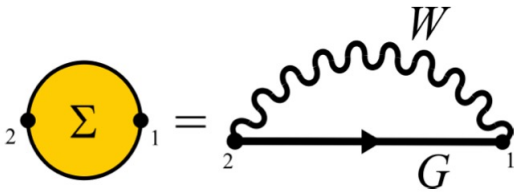
$$\Sigma(1,2)=i\int d34\; G(1,3)W(4,1^+)\tilde{\Gamma}(4,2;3)$$

Vertex $\quad \tilde{\Gamma}(1,2,3)=\delta(1,2)\delta(1,3)+\dfrac{\partial\Sigma(1,2)}{\partial V(3)}$

$(V=U+V^H)$

U = infinitesimal external perturbation

The GW in practice : frequency representation and Kohn-Sham starting point



A dynamical XC operator

Exchange and Correlation

$$\Sigma(\mathbf{r},\mathbf{r}';\varepsilon) = \frac{i}{2\pi}\int d\omega\, e^{i\omega 0^+} G(\mathbf{r},\mathbf{r}';\varepsilon+\omega)W(\mathbf{r},\mathbf{r}';\omega)$$

Energy convolution between G and W

First-order like perturbation to correct Kohn-Sham energies

$$\varepsilon_n^{GW} = \varepsilon_n^{KS} + \langle\phi_n^{KS}|\Sigma_{XC}(\varepsilon_n^{GW}) - V_{XC}^{DFT}|\phi_n^{KS}\rangle$$

Build the « best » G and W available : start with Kohn-Sham eigenstates possibly with hybrid functional

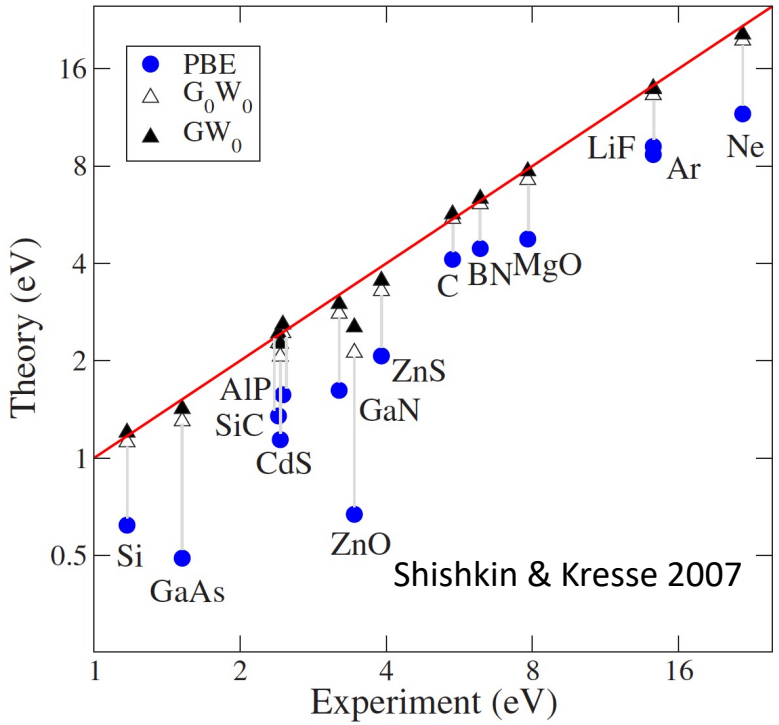
$$G^{KS}(\mathbf{r},\mathbf{r}';\omega) = \sum_n \frac{\phi_n^{KS}(\mathbf{r})[\phi_n^{KS}]^*(\mathbf{r}')}{\hbar\omega - \varepsilon_n^{KS} + i0^+\hbar\,\mathrm{sgn}(\varepsilon_n^{KS}-\mu)}$$

$$\chi_0^{KS}(\mathbf{r},\mathbf{r}';\omega) = \sum_{mn}(f_m - f_n)\frac{[\phi_m^{KS}(\mathbf{r})]^*\phi_n^{KS}(\mathbf{r})[\phi_n^{KS}(\mathbf{r}')]^*\phi_m^{KS}(\mathbf{r}')}{\omega - (\varepsilon_m^{KS}-\varepsilon_n^{KS}) + i0^+}$$

The construction of the susceptibility scales as $O(N^4)$

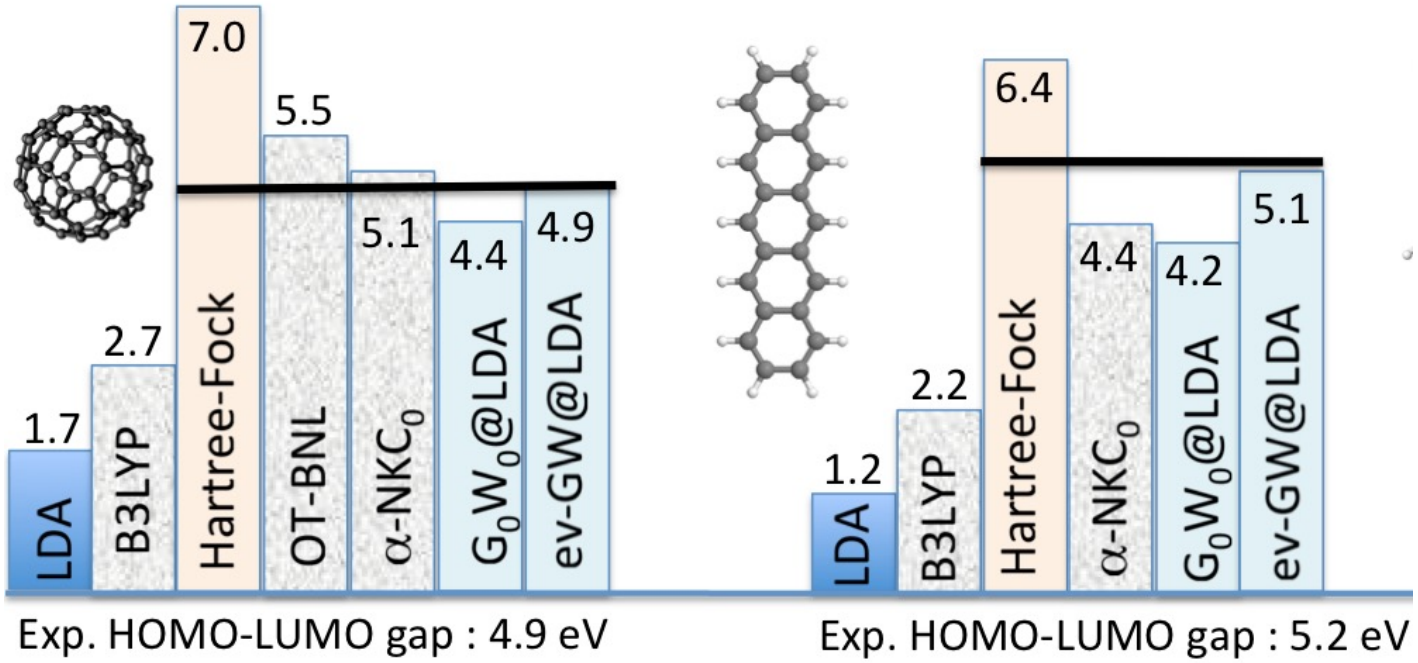
Standard results in condensed matter physics and chemistry

First ab initio *GW* calculation in solids : mid-80s



Shishkin & Kresse 2007

Blase, Attaccalite, Olevano, 2011

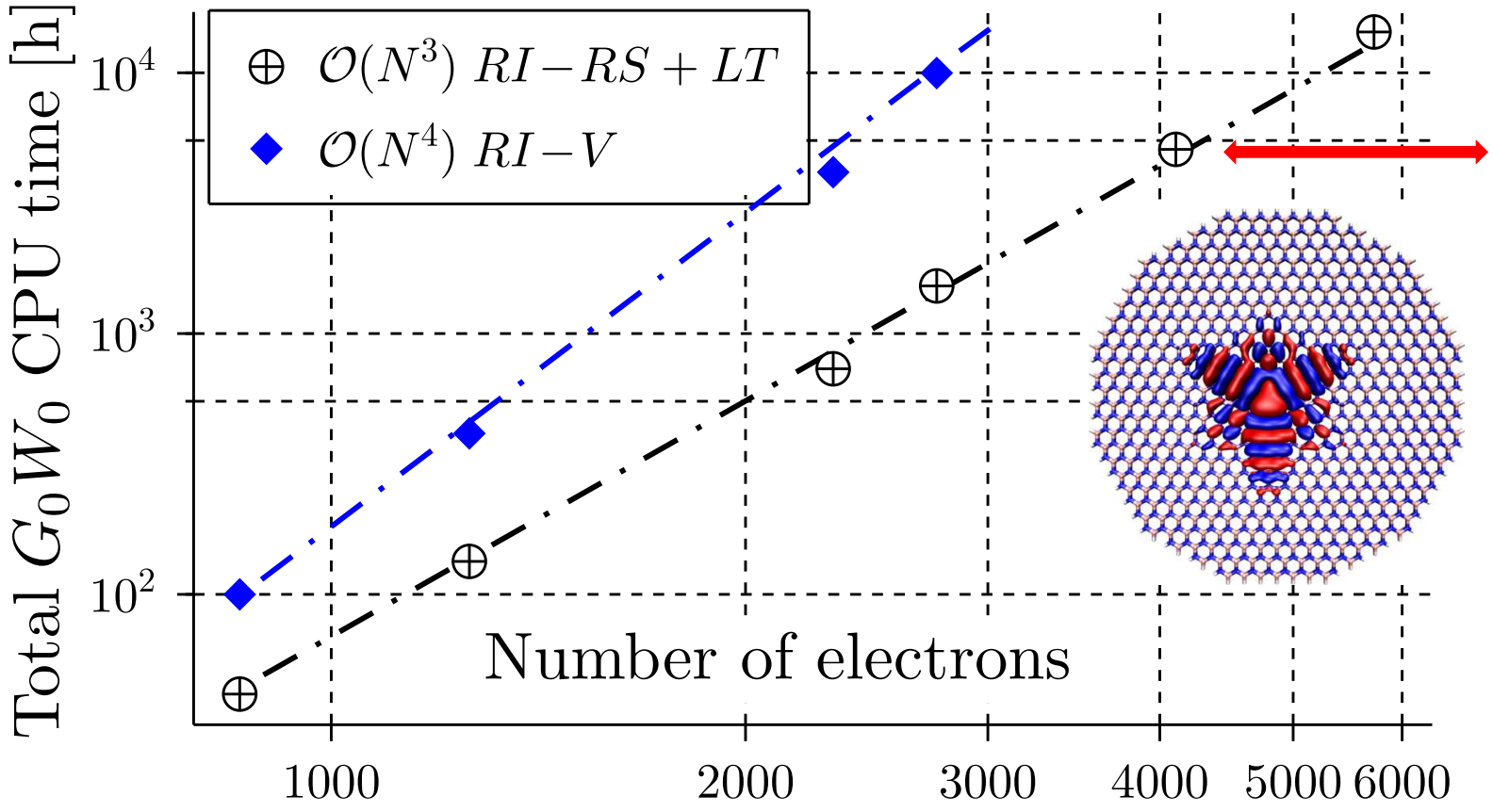


Partially self-consistent *GW* : insert self-consistently corrected energy levels to (re)build G and W => much more accurate in particular when starting with functionals with no exact exchange.

Fully self-consistent GW : update as well the eigenstates (or spectral weights in G)
=> several flavours of self-consistency and much more expensive.

From quartic to cubic scaling GW formalism

Duchemin & Blase JCTC 2021



Cubic-scaling algorithm:
4000 electrons (667 C atoms)
in less than 5000 hours
(with meV accuracy).

Standard *O(N⁴)* approach : 3000 electrons (500 carbon atoms) with 6-311G* basis and resolution-of-identity (RI-V) within 20000 hours !

=> *GW* very popular post-DFT approach for electronic structure properties implemented in many physics and chemistry codes (VASP, ABINIT, QuantumEspresso, BerkeleyGW, WEST, Turbomole, CP2K, MolGW, Fiesta, QuAcK, etc.)

Cubic-scaling space-time approach (Rojas, Godby, Needs, PRL 1995)

$$\chi_0(\mathbf{r},\mathbf{r}';\omega) = 2\sum_{ja} \frac{\phi_j^*(\mathbf{r})\phi_a(\mathbf{r})\phi_a^*(\mathbf{r}')\phi_j(\mathbf{r}')}{\omega-(\varepsilon_a-\varepsilon_i)} + c.c.$$

Imaginary time representation

$$\chi_0(\mathbf{r},\mathbf{r}';i\tau) = -iG(\mathbf{r},\mathbf{r}';i\tau)G(\mathbf{r}',\mathbf{r};-i\tau)$$

$$G(\mathbf{r},\mathbf{r}';i\tau) = -i\sum_{j}^{\text{occ}} \phi_j(\mathbf{r})\phi_j(\mathbf{r}')e^{\varepsilon_j\tau} \quad (\tau>0)$$

$$-i\sum_{a}^{\text{vir}} \phi_a(\mathbf{r})\phi_a(\mathbf{r}')e^{\varepsilon_a\tau} \quad (\tau<0)$$

Almlöf, Häser (1991,1992) :
Laplace transform idea for MP2 calculations

$$E^{(2)} = -\frac{1}{4}\sum_{ijab}\frac{\langle ab||ij\rangle^2}{\varepsilon_a+\varepsilon_b-\varepsilon_i-\varepsilon_j}$$

Laplace transform

$$(\varepsilon_a+\varepsilon_b-\varepsilon_i-\varepsilon_j)^{-1} = \int_0^{+\infty} dt\, e^{-(\varepsilon_a+\varepsilon_b-\varepsilon_i-\varepsilon_j)t}$$

Disentangles energies in denominator

The decoupling of the sums between occupied and unoccupied states reduce the calculation of the susceptibility to O(N^3). Such an approach has hardly been used in practice but is regaining popularity under various forms.

See Kresse *et al*. PRB 2016.

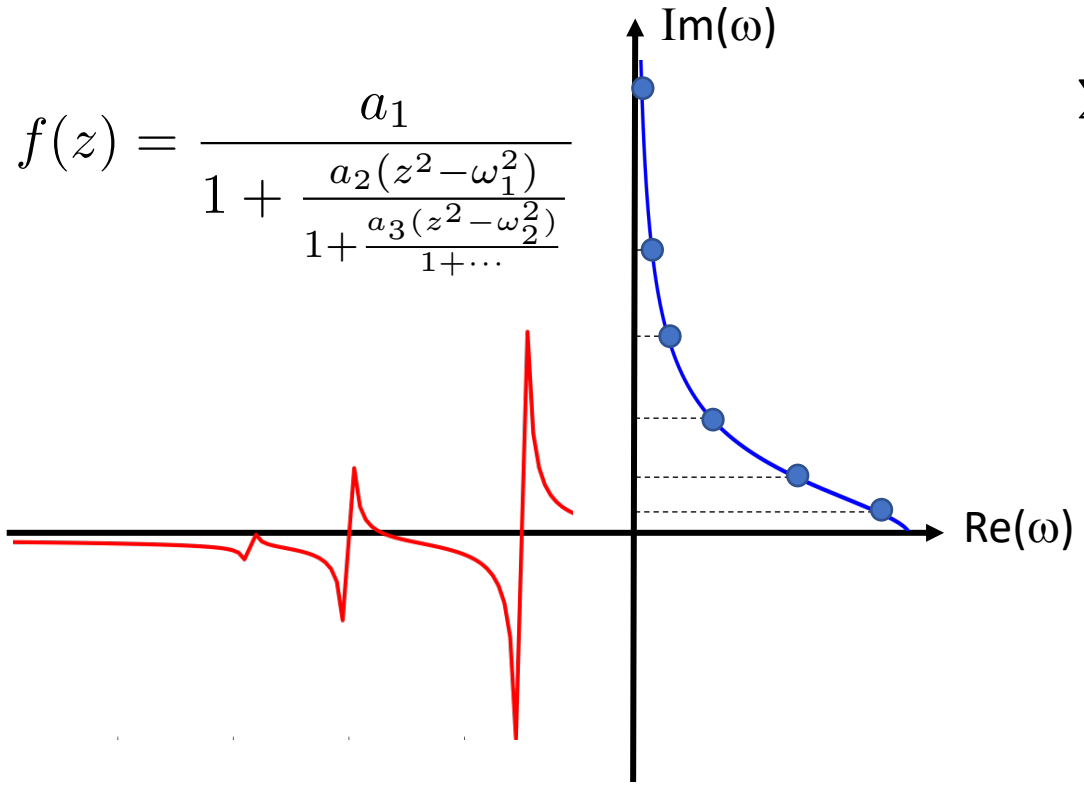
A first issue : The GW self-energy analytic continuation problem

$$\chi_0(\mathbf{r},\mathbf{r}';i\tau) \Longrightarrow \chi_0(\mathbf{r},\mathbf{r}';i\omega) \Longrightarrow W(\mathbf{r},\mathbf{r}';i\omega) \Longrightarrow \Sigma(\mathbf{r},\mathbf{r}';i\omega)$$

Safe and stable procedure : the self-energy is smooth along the imaginary axis

BUT: the analytic continuation to the real-axis is a well-known problem

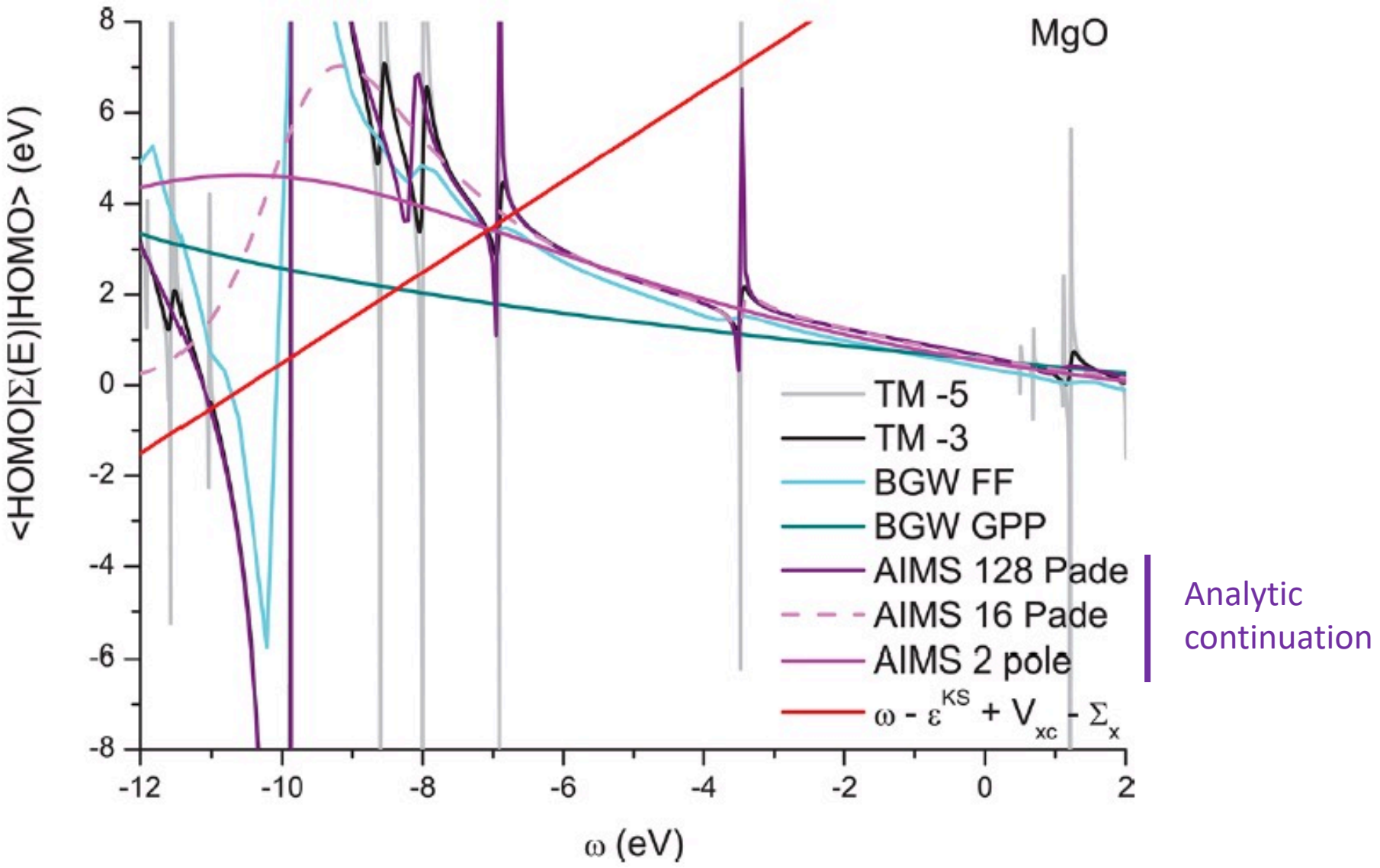
$$f(z) = \cfrac{a_1}{1+\cfrac{a_2(z^2-\omega_1^2)}{1+\cfrac{a_3(z^2-\omega_2^2)}{1+\cdots}}}$$



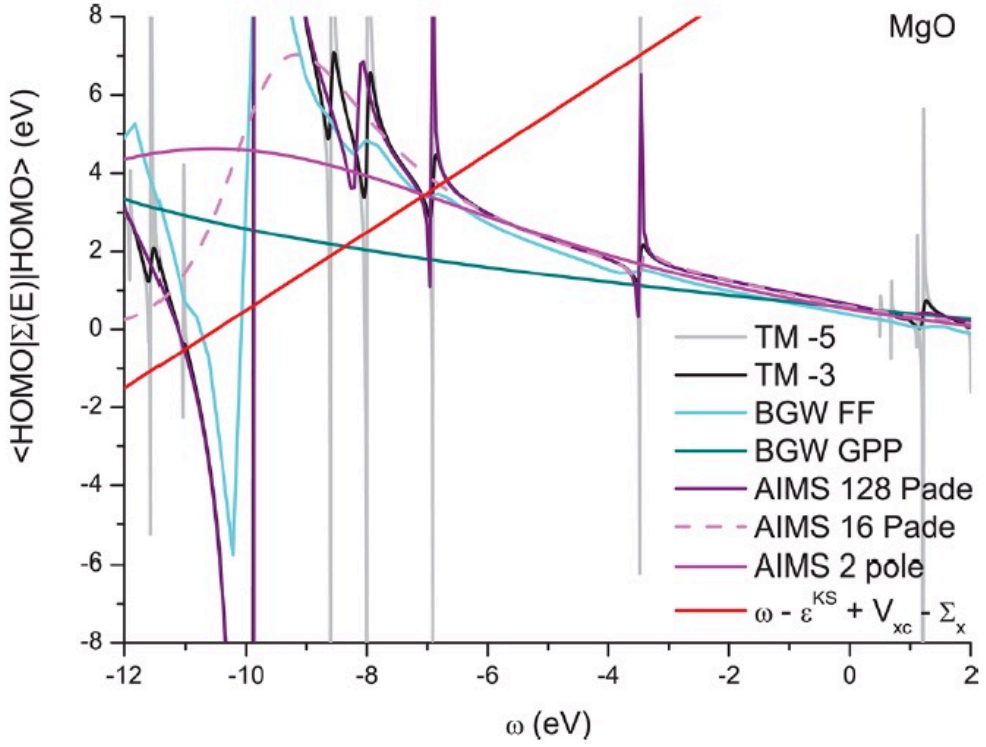
$$\Sigma(\mathbf{r},\mathbf{r}';\varepsilon) = \frac{i}{2\pi}\int d\omega\; e^{i\omega 0^+} G(\mathbf{r},\mathbf{r}';\varepsilon+\omega)W(\mathbf{r},\mathbf{r}';\omega)$$

The number of pole of the self-energies are the poles of G multiplied by the poles of W !

A difficult case (among others)



van Setten *et al*. "GW100: benchmarking…", JCTC 2015



$$\varepsilon_n^{GW} = \varepsilon_n^{KS} + \langle \phi_n^{KS} | \Sigma_{XC}(\varepsilon_n^{GW}) - V_{XC}^{DFT} | \phi_n^{KS} \rangle$$

Potentially large errors, including missing states !
(see "GW100" paper, JCTC 2015)

Solution : G and its poles on the real axis are known ! The only thing that needs to be continued is the screened Coulomb potential that has much less poles than the self-energy !

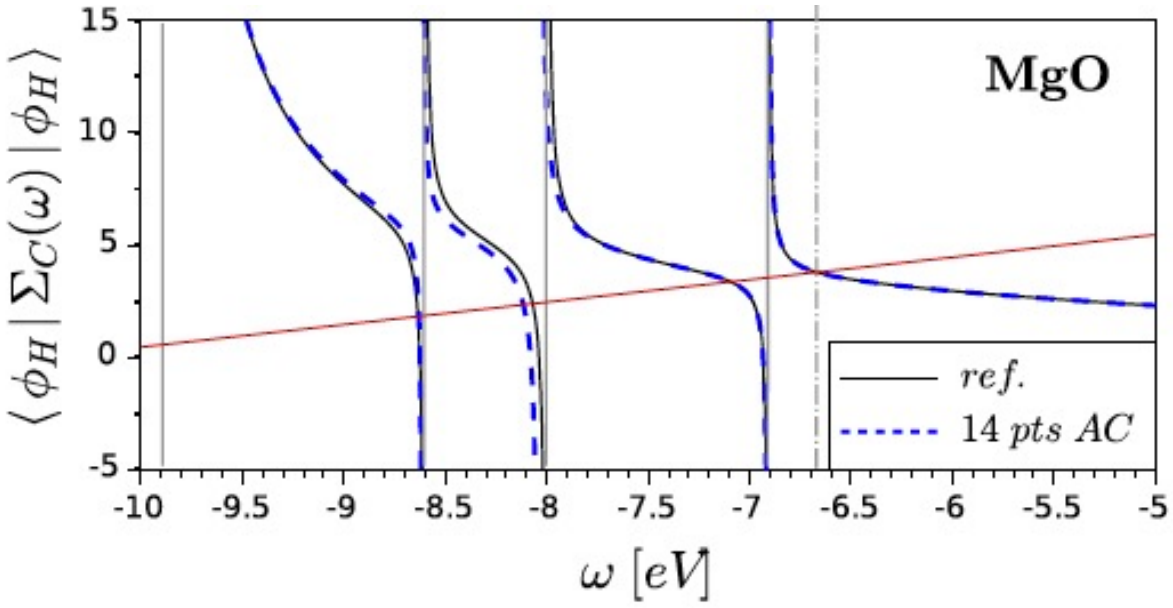
$$G(\mathbf{r}, \mathbf{r}'; \omega) = \sum_n \frac{\phi_n(\mathbf{r}) \phi_n^*(\mathbf{r}')}{\omega - \varepsilon_n + i\eta \times \mathrm{sgn}(\varepsilon_n - \mu)}$$

$$\chi_0(\mathbf{r}, \mathbf{r}'; i\tau) \Longrightarrow \chi_0(\mathbf{r}, \mathbf{r}'; i\omega) \Longrightarrow W(\mathbf{r}, \mathbf{r}'; i\omega) \Longrightarrow W(\mathbf{r}, \mathbf{r}'; \omega)$$

Form $\Sigma(\omega)$

A stable analytic continuation approach

Duchemin & Blase, JCTC 2020



Quasiparticle energy at the crossing of the red line and the correlation self-energy $\Sigma_C(\omega)$

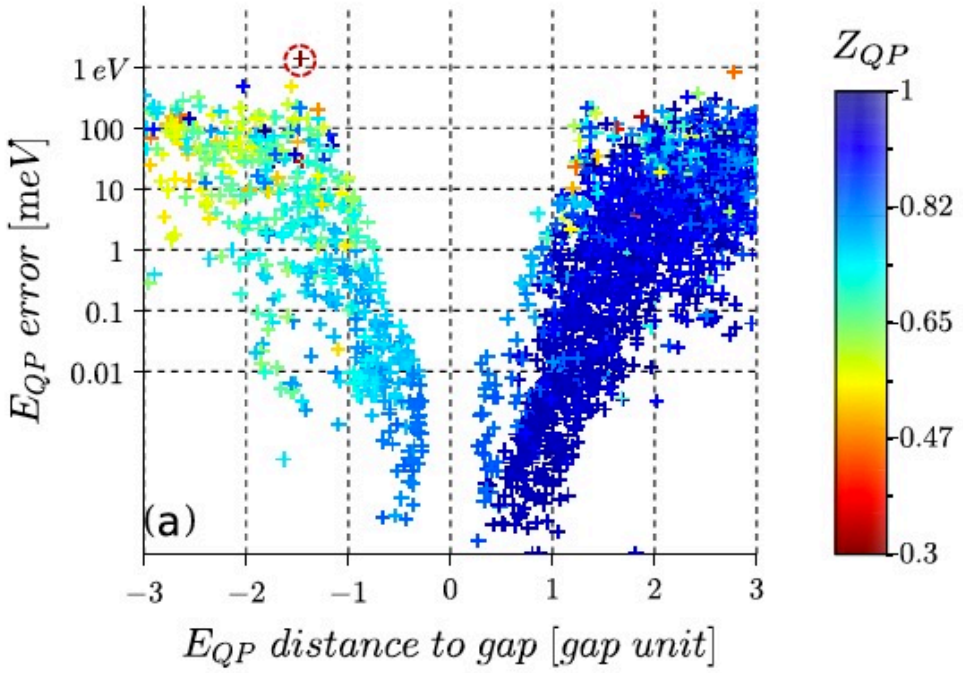
$$\varepsilon_n^{GW} = \varepsilon_n^{KS} + \langle \phi_n^{KS} | \Sigma_{XC}(\varepsilon_n^{GW}) - V_{XC}^{DFT} | \phi_n^{KS} \rangle$$

Crossing of 2 functions: $\varepsilon_n^{GW} = \omega$

$$f(\omega) = \omega - \varepsilon_n^{KS} + \langle \phi_n^{KS} | V_{XC}^{DFT} | \phi_n^{KS} \rangle$$

$$g(\omega) = \langle \phi_n^{KS} | \Sigma_{XC}(\omega) | \phi_n^{KS} \rangle$$

Analytic continuation errors (GW100 test set)



Analytic continuation extremely safe close to the gap with few points (10-14) along the imaginary axis.

A second issue : Disentangling the molecular orbitals with separable resolution-of-the-identity (RI)

In physics (planewaves) or quantum chemistry (atomic basis) susceptibility not calculated on a grid :

$$\chi_0(\mathbf{r},\mathbf{r}';i\omega) = 2\sum_{ja} \frac{\phi_j^*(\mathbf{r})\phi_a(\mathbf{r})\phi_a^*(\mathbf{r}')\phi_j(\mathbf{r}')}{i\omega - (\varepsilon_a - \varepsilon_i)} + c.c.$$

but using an auxiliary basis (planewaves, atomic orbitals, etc.) that projects on codensities (MOs products) :

$$\langle P_\mu|\chi_0(\mathbf{r},\mathbf{r}';i\omega)|P_\nu\rangle = 2\sum_{ja} \frac{\langle\phi_j\phi_a|P_\mu\rangle\langle P_\nu|\phi_a\phi_j\rangle}{i\omega - (\varepsilon_a - \varepsilon_i)} + c.c.$$

MOs are entangled (not separable) in RI fitting coefficients !

Standard resolution-of-the-identity (RI) approach

This amounts in quantum chemistry to a resolution-of-the-identity (RI) technique :

$$\phi_n(\mathbf{r})\phi_m(\mathbf{r}) = \sum_\mu \mathcal{F}_\mu(\phi_n\phi_m)P_\mu(\mathbf{r})$$

$$\mathcal{F}_\mu^V(\phi_n\phi_m) = \sum_\nu [V^{-1}]_{\mu\nu}(P_\nu|\phi_n\phi_m)$$

Coulomb fitting

$$[\chi_0^{RI}(i\omega)]_{\mu\nu} = 2\sum_{ja} \frac{\mathcal{F}_\mu(\phi_j\phi_a)\mathcal{F}_\nu(\phi_j\phi_a)}{i\omega - (\varepsilon_a - \varepsilon_j)} + c.c.$$

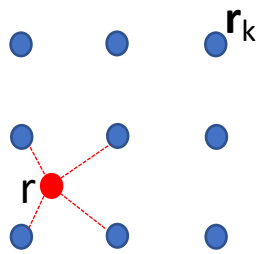
or

$$[\chi_0^{RI}(i\tau)]_{\mu\nu} = -2i\sum_{ja} \mathcal{F}_\mu(\phi_j\phi_a)\mathcal{F}_\nu(\phi_j\phi_a)e^{\varepsilon_j\tau}e^{-\varepsilon_a\tau}$$

MOs are entangled (not separable) in RI fitting coefficients !

Disentangling the molecular orbitals with separable resolution-of-the-identity (RI)

$$\phi_n(\mathbf{r})\phi_m(\mathbf{r}) = \sum_\mu \mathcal{F}_\mu^V(\phi_n\phi_m)P_\mu(\mathbf{r})$$

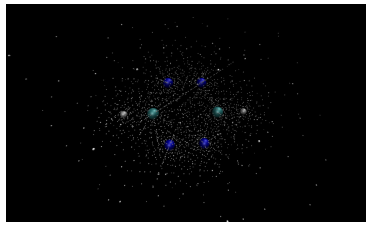


"Interpolation"

Fitting coefficient by interpolation (quadrature)

$$\mathcal{F}_\mu^{RS}(\phi_j\phi_a) = \sum_k M_{\mu k}\ \phi_j(\mathbf{r}_k)\phi_a(\mathbf{r}_k)$$

Build $\{M_{\mu k}, \mathbf{r}_k\}$



$$\underset{\{M_{\mu k},\mathbf{r}_k\}}{\text{argmin}} \sum_{\mu,ja} \left| \left(\mathcal{F}_\mu^{RS}(\phi_j\phi_a) - \mathcal{F}_\mu^V(\phi_j\phi_a)\right) P_\mu \right|_V^2$$

Least-square minimization of optimal quadrature

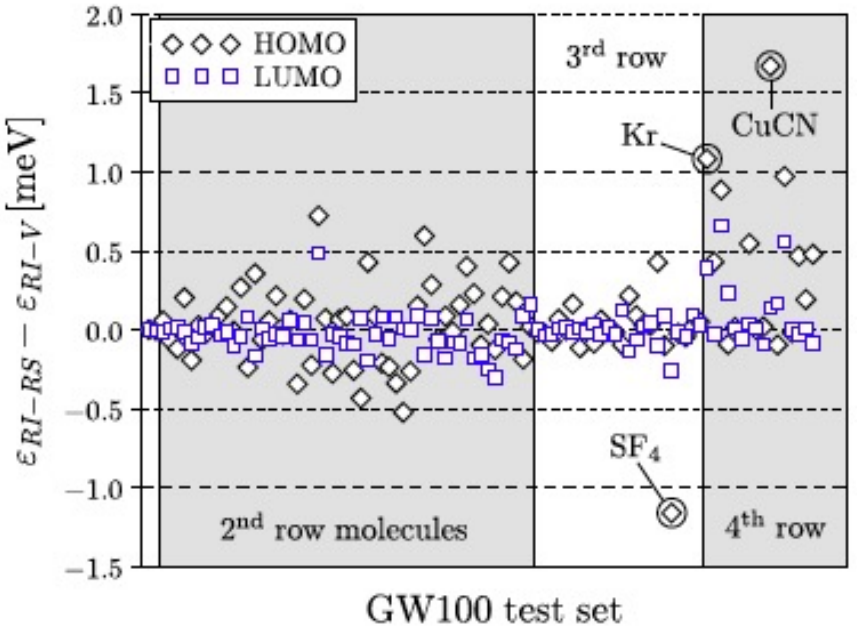
Build minimal grid that reproduces a RI-V calculation for a given auxiliary basis and a given accuracy (meV)

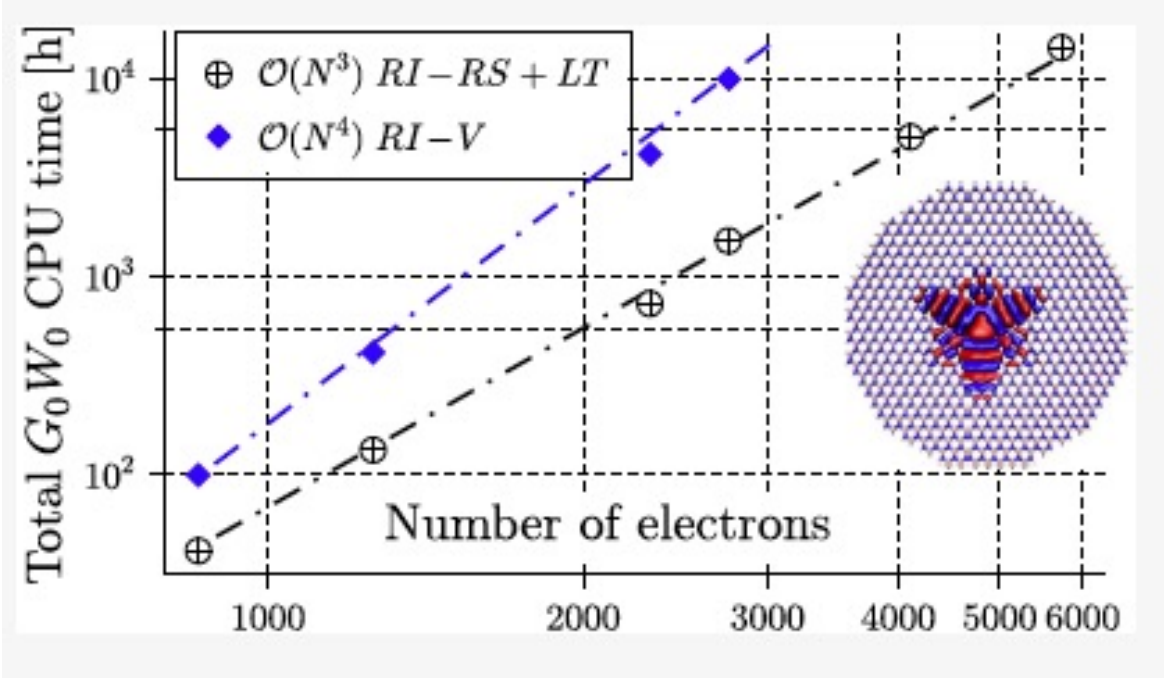
=> Typical grid size needed: 300-500 points per atom (3-4 times larger than auxiliary Gaussian basis set)

See also: ISDF (Interpolative Separable Density Fitting), Lu and Thicke, J. Comput. Phys. 2017

Self-adaptative minimal grid : a few hundred grid points per atom for meV accuracy !

Standard RI-V results can be obtained with much reduced CPU (cubic scaling) and meV accuracy !
Crossover with standard quartic scaling approach for systems containing a few dozens of atoms.





Duchemin & Blase, JCP 2019; JCTC 2021

- The same technique(s) can be used for cubic-scaling RPA correlation energies (Duchemin & Blase, JCP 2019)

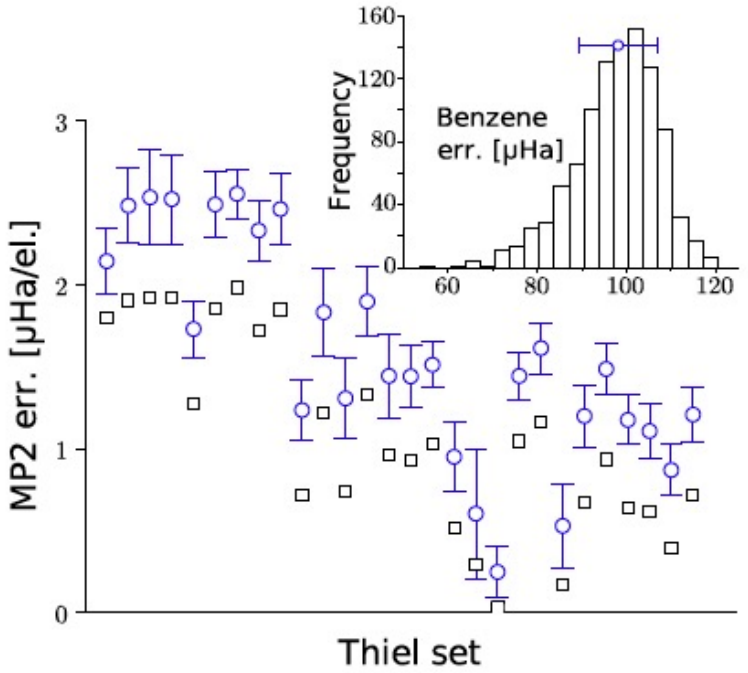
$$E_C^{RPA} = \frac{1}{2\pi}\int_0^\infty d\omega\, Tr\left[\ln(1-\chi_0(i\omega)\cdot v) + \chi_0(i\omega)\cdot v\right]$$

- The same technique can be « simply » used to calculate 2-electron Coulomb integrals (applications to HF, MP2, etc.)

$$(ia|jb) = \sum_{\mu\nu} \mathcal{F}_\mu(ia)\cdot(P_\mu|P_\nu)\cdot\mathcal{F}_\mu(jb)$$

$$\downarrow$$

$$E_C^{MP2} = -\sum_{ij}\sum_{ab}\frac{(ia|jb)[2(ia|jb)-(ib|ja)]}{\varepsilon_a+\varepsilon_b-\varepsilon_i-\varepsilon_j}$$



For Coulomb integrals, similar in spirit to other techniques in quantum chemistry such as chain-of-sphere COS (Neese et al, 2011), Tensor Hyper Contraction techniques (Sherrill et al. 2016), etc.

Towards large scale GW (and Bethe-Salpeter) calculations …

Conclusions : cubic scaling all-electron *GW* calculations can be performed on hundreds of atoms with modest CPU time requirements (reducing the memory needs as well).

- No use of sparcity in AO basis set overlap
- No use of localization of 2-body operators such as $G(\mathbf{r},\mathbf{r}';\omega), \chi_0(\mathbf{r},\mathbf{r}';\omega), etc.$

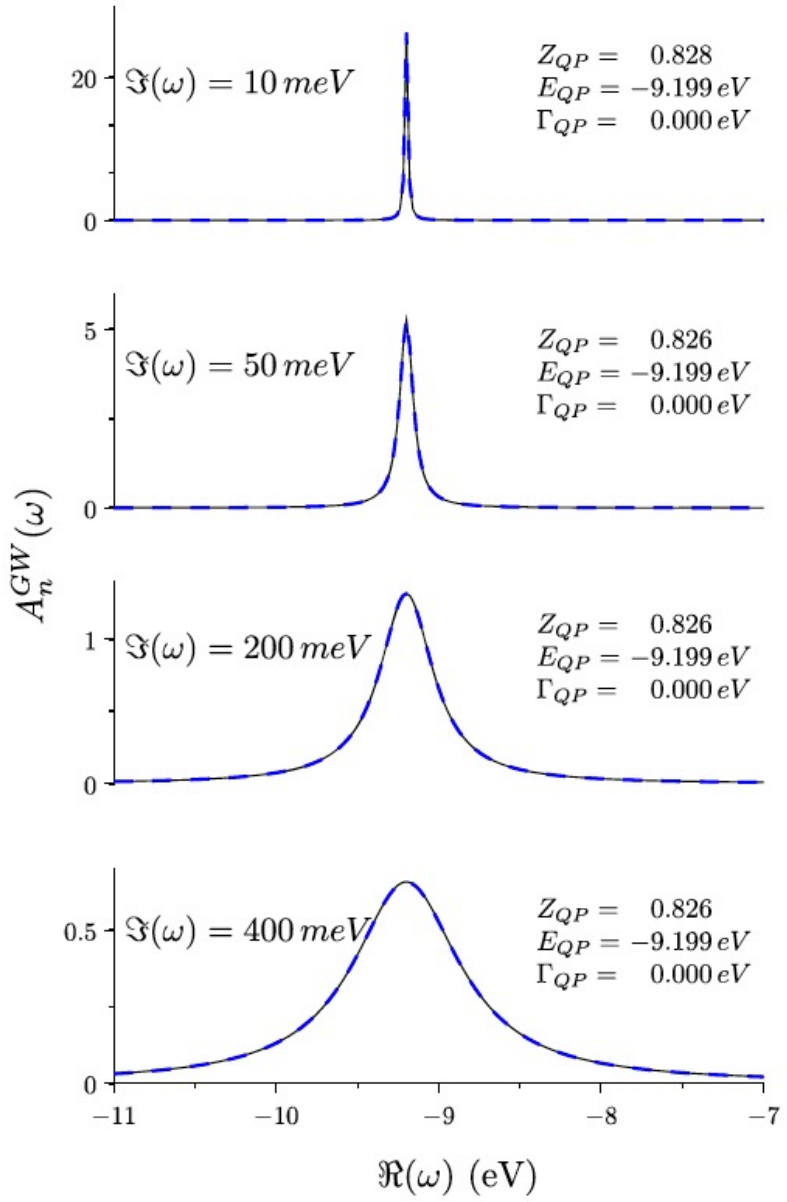
THANKS

Many thanks again Titou, Pina, Julien, Michel.

anr agence nationale de la recherche AU SERVICE DE LA SCIENCE

We thank the ANR for financial support in the framework of the BSE-forces grant (ANR-20-CE29-0005)

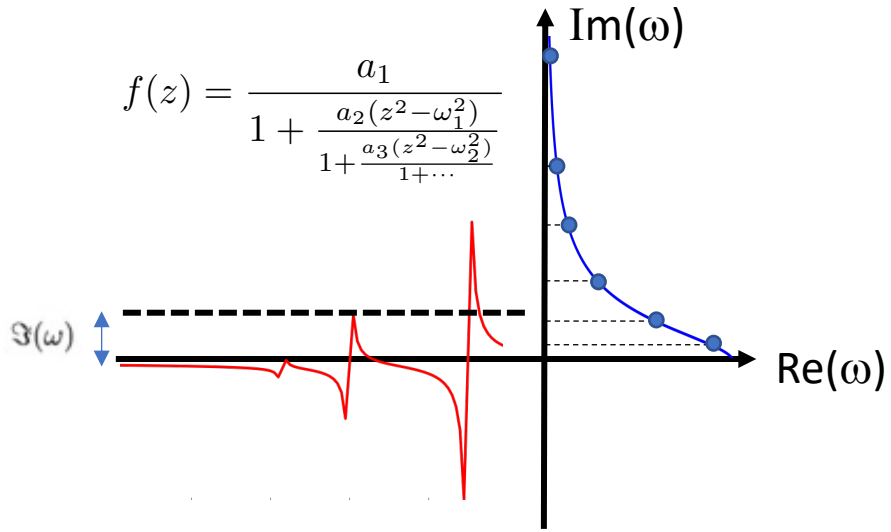
GW spectral function and quasiparticle peak



$$A_n^{GW} = \frac{1}{\pi} |Im\langle\phi_n|G(\omega)|\phi_n\rangle|$$

$$G = G^{KS} + G^{KS}(\Sigma - V^{XC})G$$

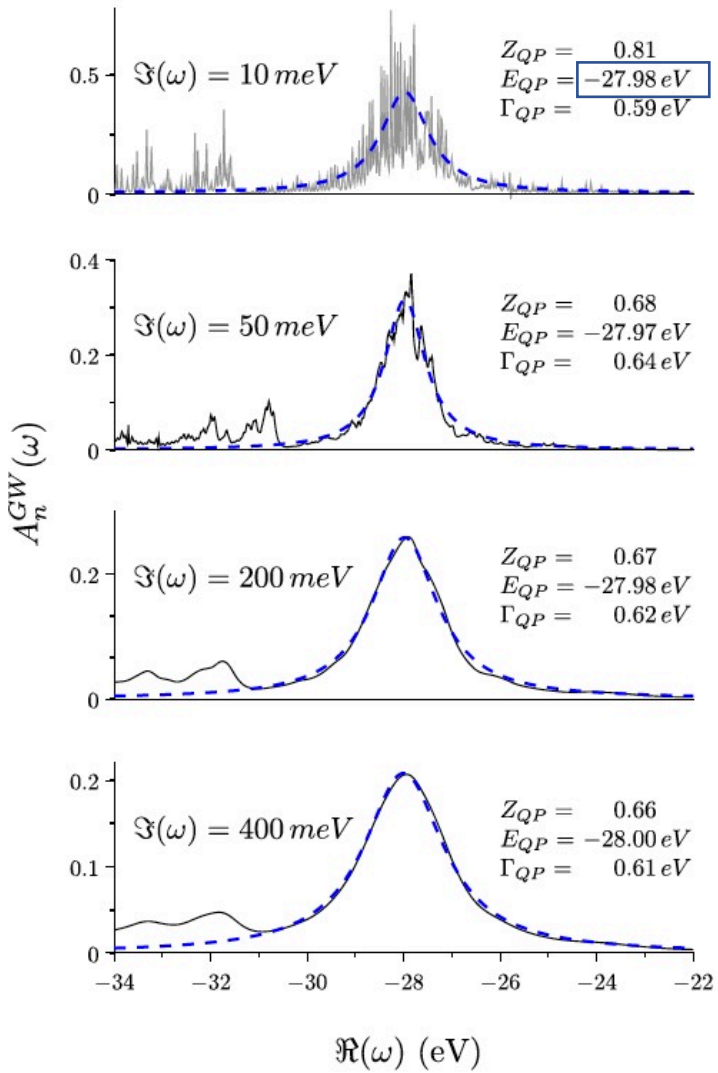
$$f(z) = \frac{a_1}{1 + \frac{a_2(z^2-\omega_1^2)}{1+\frac{a_3(z^2-\omega_2^2)}{1+\cdots}}}$$



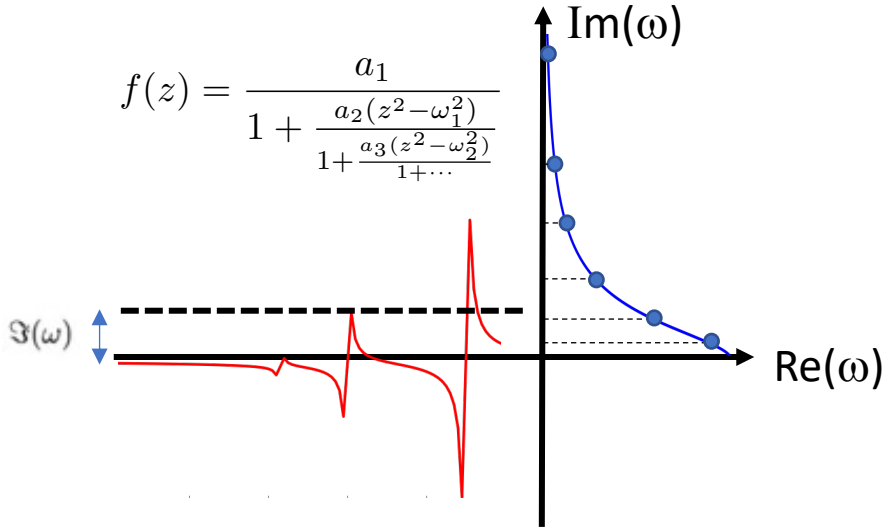
TCNQ HOMO

GW spectral function and quasiparticle peak

$$A_n^{GW}(\omega) = \frac{1}{\pi}|\Im(\langle\phi_n|G(\omega)|\phi_n\rangle)|$$

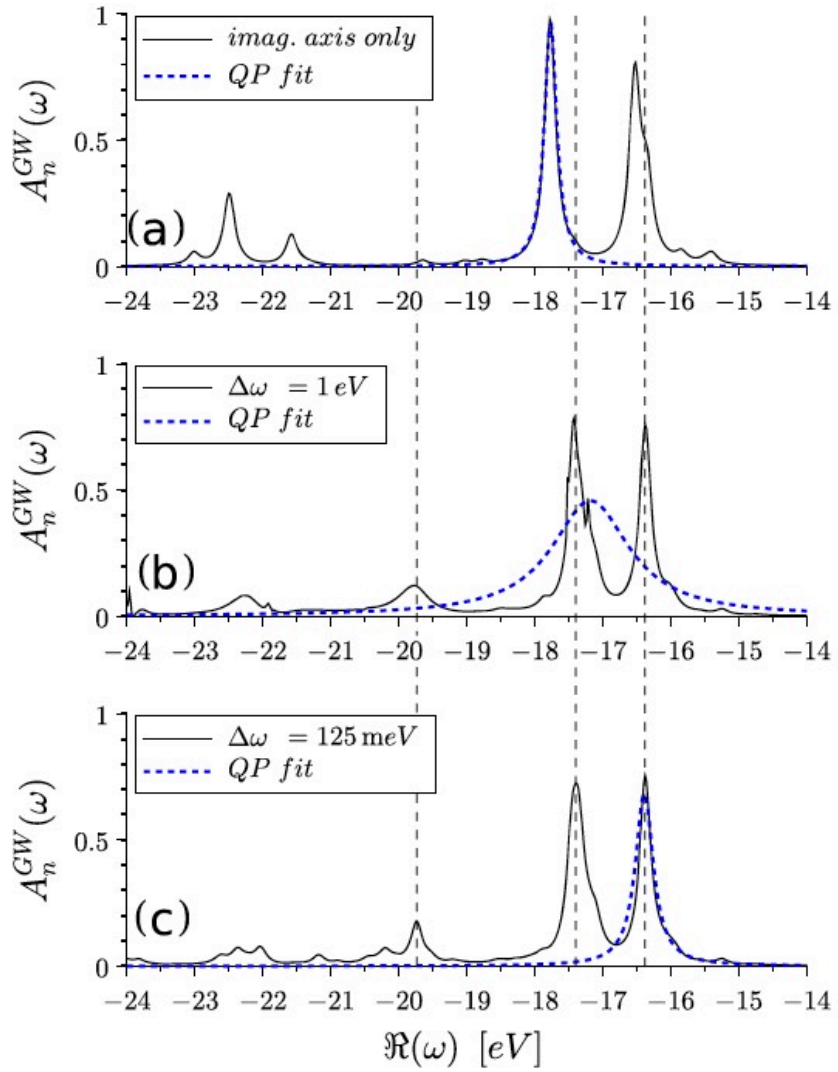


$$f(z) = \frac{a_1}{1 + \frac{a_2(z^2-\omega_1^2)}{1+\frac{a_3(z^2-\omega_2^2)}{1+\cdots}}}$$



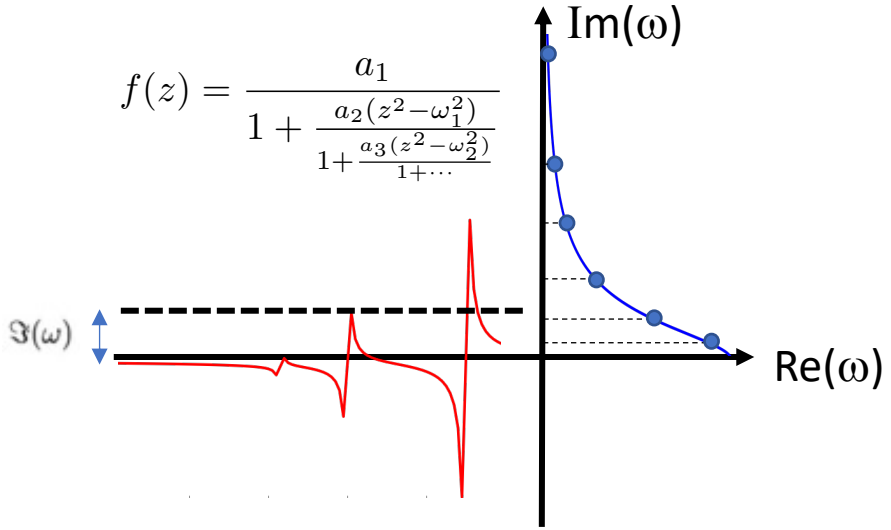
GW spectral function and quasiparticle peak

Multiple quasiparticle (?) peaks



nitrobenzonitrile

$$f(z) = \frac{a_1}{1 + \frac{a_2(z^2-\omega_1^2)}{1+\frac{a_3(z^2-\omega_2^2)}{1+\cdots}}}$$



Tensor Hypercontraction Technique

R. M. Parrish, E. G. Hohenstein, T. J. Martinez, and C. D. Sherrill, J. Chem. Phys. 137, 224106 (2012)

$$\underset{Z}{\text{argmin}} \sum_{\rho,\rho'} ||(\rho|\rho') - \sum_{kk'} \rho(\mathbf{r}_k) \cdot Z_{kk'} \cdot \rho'(\mathbf{r}_{k'})||^2$$

Similar idea but fit directly 2-electrons Coulomb integrals => *O(N⁴)* setting of quadrature instead of *O(N³)* in our case.

Chain-of-sphere approach (COS)

Neese et al. 2009 - present

$$(\mu\kappa|\nu\lambda) \simeq \sum_k W_k \mu(\mathbf{r}_k)\kappa(\mathbf{r}_k) \int d\mathbf{r} \frac{\nu(\mathbf{r})\lambda(\mathbf{r})}{|\mathbf{r}-\mathbf{r}_k|}$$

Quadrature on one codensity only

One-electron integrals in Gaussian basis

Review : Elisa Rebolini *et al*. JCTC 2016

Imaginary frequency grid points (RPA)

Need to evaluate integrals of the form

E ranges from E_{gap} to $\max(\varepsilon_a - \varepsilon_i)$

$$\int_0^\infty d\omega \left(\frac{1}{E - i\omega} - \frac{1}{E + i\omega} \right) = \pi$$

Minimize

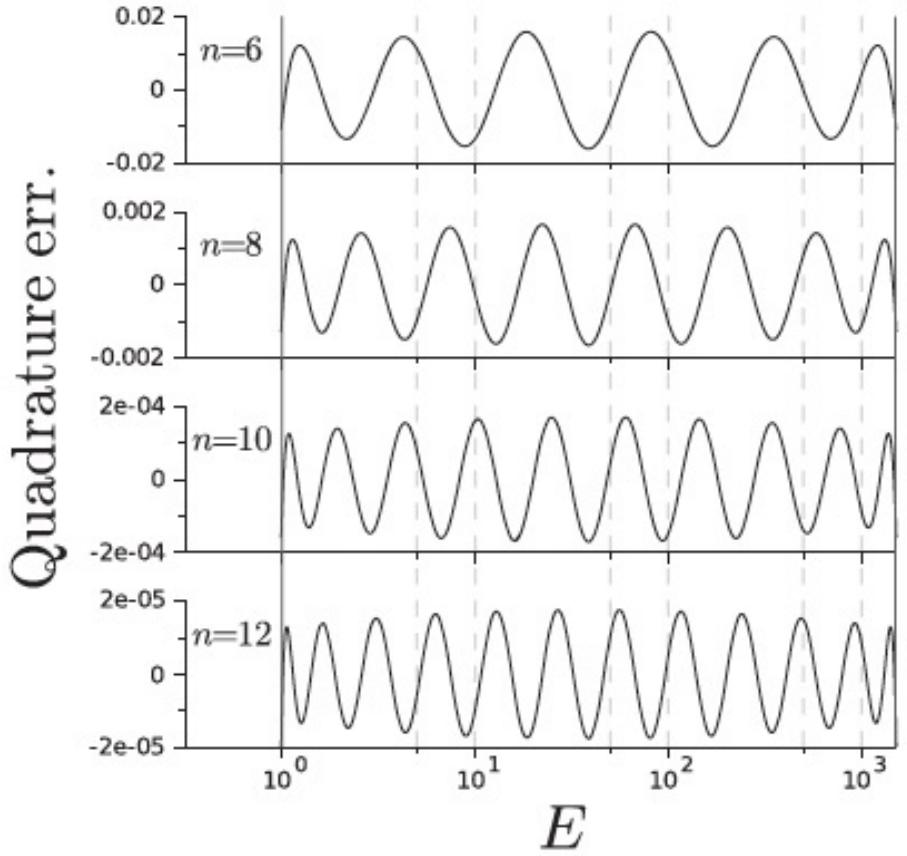
Discretized integral

$$\underset{w_k, z_k}{\operatorname{argmin}} \left[\int_{\ln(E_{min})}^{\ln(E_{max})} du \left\| \sum_k w_k \left[\frac{1}{e^u - i z_k} + \frac{1}{e^u + i z_k} \right] - \pi \right\|^2 \right].$$

weight

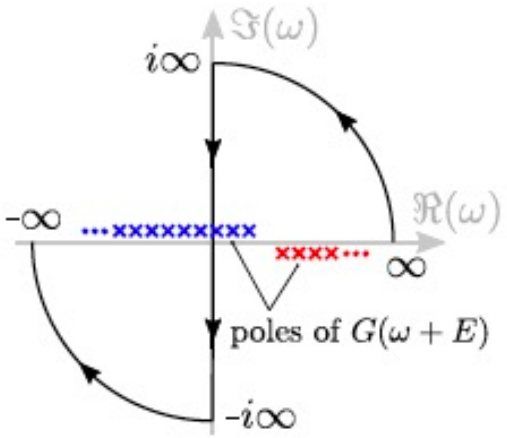
Discrete frequencies

Typically 10 to 14 frequencies



Contour deformation with analytic continuation

$$\Sigma_C^{GW}(\mathbf{r},\mathbf{r}';E) = \frac{-1}{2\pi}\int_{-\infty}^{\infty} d\omega\, G(\mathbf{r},\mathbf{r}';E+i\omega)(W-v)(\mathbf{r},\mathbf{r}';i\omega)$$
$$- \sum_i \phi_i(\mathbf{r})\phi_i(\mathbf{r}')(W-v)(\mathbf{r},\mathbf{r}';\varepsilon_i - E)\theta(\varepsilon_i - E)$$
$$+ \sum_a \phi_a(\mathbf{r})\phi_a(\mathbf{r}')(W-v)(\mathbf{r},\mathbf{r}';E-\varepsilon_a)\theta(E-\varepsilon_a)$$



Poles of W shifted by Green's function poles