

Hubbard, Hybrid and Nonlocal Functionals in QUANTUM ESPRESSO

P. Giannozzi

Università di Udine and CNR-IOM Democritos, Trieste, Italy

Summer School on Advanced Materials and Molecular Modelling
September 16-20, Ljubljana, Slovenija

1. Reminder on “plain” DFT: LDA and GGA, and its failures
2. Functionals for strongly correlated materials: DFT+U, hybrid functionals
3. Functionals for van der Waals-bonded materials



Performances of LDA and GGA: atoms, small molecules

LDA underestimates the stability of atoms; GGA is close to experiments.

A few binding energies of small molecules:

E_b (eV)	exp.	LDA	GGA	HF
H ₂	4.753	4.913	4.540	3.64
LiH	2.509	2.648	2.322	
O ₂	5.239	7.595	6.237	1.28
H ₂ O	10.078	11.567	10.165	
F ₂	1.66	3.32		1.37

Binding energy is too high in LDA. GGA is closer to experiments, but sometimes its binding is too weak. Hartree-Fock without corrections yields very bad results.

Lattice parameters in solids

a (Å)	exp.	LDA	Δ	GGA	Δ
Si	5.43	5.40	-0.50%	5.49	1.16%
Ge	5.65	5.62	-0.53%	5.47	1.59%
GaAs	5.65	5.62	-0.53%	5.73	1.42%
Al	4.03	3.98	-1.31%	4.09	1.57%
Cu	3.60	3.52	-2.35%	3.62	0.44%
Ag	4.07	4.00	-1.69%	4.17	2.47%
Ta	3.30	3.26	-1.12%	3.32	0.80%
W	3.16	3.14	-0.67%	3.18	0.67%
Pt	3.91	3.90	-0.41%	3.97	1.49%
Au	4.06	4.05	-0.13%	4.16	2.48%

LDA tends to “overbind”, GGA to “underbind”. GGA error more variable.

Bulk modulus in solids

B (GPa)	exp.	LDA	Δ	GGA	Δ
Si	99	96	-3.0%	83	-16.2%
Ge	77	78	1.3%	61	-20.8%
GaAs	76	74	-2.6%	65	-14.5%
Al	77	84	9.1%	73	-5.2%
Cu	138	192	39.1%	151	9.4%
Ag	102	139	36.3%	85	-16.7%
Ta	193	224	16.1%	197	2.1%
W	310	337	8.7%	307	-1.0%
Pt	283	307	8.5%	246	-13.1%
Au	172	198	15.1%	142	-17.4%

LDA tends to be too “stiff”, GGA to be too “soft”

Lattice parameters of oxides

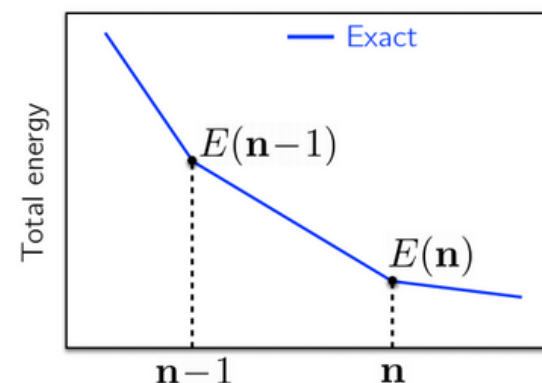
a (Å)	exp.	LDA	Δ	GGA	Δ
MgO	4.21	4.17	-1.0%		
TiO ₂ (a)	4.59	4.55	-0.9%	4.62	0.7%
TiO ₂ (c)	2.96	2.94	-0.5%	2.99	1.0%
Al ₂ O ₃	5.13	5.09	-0.7%	5.19	1.1%
BaTiO ₃	4.00	3.94	-1.5%		
PbTiO ₃	3.90	3.83	-1.7%	3.89	-0.2%
SnO ₂	4.74	4.64	-2.1%		
β -MnO ₂ (a)	4.40	4.35	-1.3%	4.44	0.9%
β -MnO ₂ (c)	2.88	2.81	-2.3%	2.89	0.5%

Same pattern for “difficult” materials like oxides: results not bad, but electronic states often more wrong than structural data

Beyond the reach of LDA and GGA

We do not know the *exact* exchange-correlation (XC) energy and potential functionals, but we know that they must be highly nontrivial objects. In particular:

- the XC potential exactly *cancels the self-interaction* of an electron with itself that is present in the Hartree potential (as in Hartree-Fock: note that the XC potential is *local*: $V_{xc} \equiv V_{xc}(\mathbf{r})$, thus much simpler than nonlocal exchange in Hartree-Fock).
- the XC energy as a function of the number of electrons *has a cusp* when the number of electrons crosses an integer value number. No simple, local function of the charge density (and of its gradient) can fulfil this requirement.

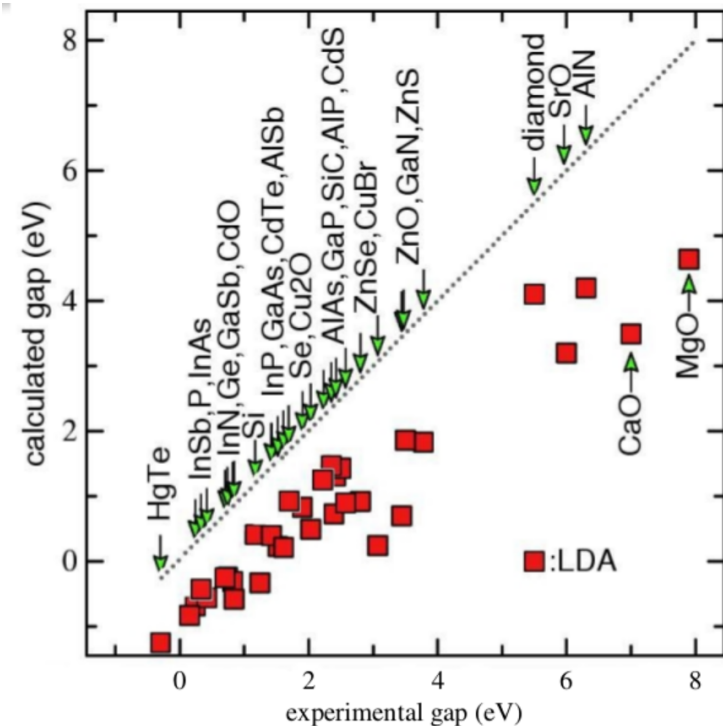


The presence of spurious self-interactions and the lack of discontinuity is at the heart of most problems with approximate DFT. The search for functionals that may solve such problems in an economical way is a very active research field.

Performances of LDA and GGA, summary

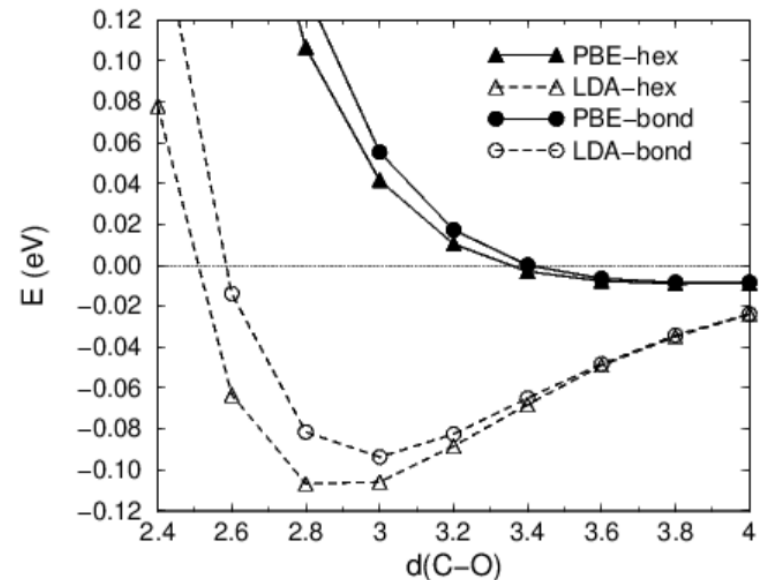
+ **Excellent structural results:** correct prediction of crystal structures, bond lengths, lattice parameters (within 1 ÷ 2%), binding and cohesive energies (GGA 5 to 10%; LDA much worse, strongly overestimates), vibrational properties. Especially good for *sp*-bonded materials but good also for more "difficult" materials, such as transition metal compounds.

– The infamous **band gap problem:** $\epsilon_c - \epsilon_v$ (or HOMO-LUMO in quantum chemistry parlance) wildly underestimates the true band gap: $\Delta = I - A$, where I is the ionization potential ($I = E(n) - E(n-1)$), A the electron affinity ($A = E(n+1) - E(n)$)



Performances of LDA and GGA (2)

- Serious trouble in dealing with **strongly correlated materials**, such as e.g. transition metal oxides: sometimes the occupancy of atomic-like electronic states is badly wrong, sometimes the electronic structure is incorrect in more subtle ways.
- Incorrect treatment of **van der Waals** interactions: LDA works only apparently (due to overbinding), GGA doesn't bind. Deep reason: van der Waals is nonlocal, doesn't depend upon charge overlap. No functional based on local density and gradients can account for it: no charge overlap, no interactions.



(in the picture: energy vs distance for an O₂ molecule over graphene)

“Beyond-GGA” functionals

An incomplete list of advanced functionals, addressing (most or at least part of the) shortcomings of LDA and GGA, especially in strongly correlated materials:

- **Self-Interaction Correction.** is Self-Interaction a problem? let's remove it! The first proposal of a Self-Interaction Correction (SIC) for extended systems, by Perdew and Zunger dates back to 1980. The functional becomes orbital-dependent, leading to some fundamental and numerical problems and unclear performances, so it never really took off. New ideas have since appeared and look more promising.

QUANTUM ESPRESSO implements in the Car-Parrinello code CP a simple form of SIC (D'Avezac et al.) working for systems with a single unpaired spin.

Meta-GGA

In *Meta-GGA* (meta = beyond), the functional has a further dependence upon the *non-interacting kinetic energy density* $\tau(\mathbf{r})$:

$$\tau_{\sigma}(\mathbf{r}) = \frac{1}{2} \sum_i |\nabla \psi_{i,\sigma}(\mathbf{r})|^2, \quad \sigma = +, -$$

The XC energy is written as

$$E_{xc}[n(\mathbf{r})] = \int \epsilon_{mGGA}(n(\mathbf{r}), \nabla n(\mathbf{r}), \tau(\mathbf{r})) n(\mathbf{r}) d\mathbf{r}$$

or, in the spin-polarized version, as

$$E_{xc}[n_+(\mathbf{r}), n_-(\mathbf{r})] = \int \epsilon_{mGGA}(n_+(\mathbf{r}), n_-(\mathbf{r}), \nabla n_+(\mathbf{r}), \nabla n_-(\mathbf{r}), \tau_+(\mathbf{r}), \tau_-(\mathbf{r})) n(\mathbf{r}) d\mathbf{r}.$$

The Kohn-Sham Hamiltonian contains an additional term:

$$\frac{\delta E_{xc}[n]}{\delta \psi_i(\mathbf{r})} = \dots - \frac{1}{2} \left(\nabla \frac{\partial \epsilon_{mGGA}}{\partial \tau(\mathbf{r})} \right) \cdot (\nabla \psi_i(\mathbf{r}))$$

Meta-GGA (2)

Meta-GGA functionals are likely the simplest example of “orbital-dependent” functionals, potentially having more flexibility, better behavior with respect to self-interaction, etc. The formulation of meta-GGA is somewhat more complex, but not much more, than plain GGA, thus making it a potentially cheap extension of GGA.

Meta-GGA consistently improves upon GGA, but it is numerically very unstable (i.e. self-consistently is very slow to converge, or even diverges), and thus not so much used, in spite of its promises. The recently introduced SCAN functional seems to yield excellent results and to be less unstable than early meta-GGA functionals like TPSS. Still, very dense FFT grids are needed in order to deal with the additional term appearing in the Hamiltonian.

QUANTUM ESPRESSO implements a few meta-GGA functionals directly (TPSS, M06L, TB09), others via `libxc` (e.g., SCAN) in ground-state calculations only (i.e. no linear response) and for Norm-Conserving PPs only. Only a few true meta-GGA PPs are currently available.

DFT+U

In DFT+U a Hubbard-like term, accounting for strong Coulomb correlations in systems with highly localized, atomic-like states, is added to the XC functional:

$$E_{DFT+U}[n(\mathbf{r})] = E_{DFT}[n(\mathbf{r})] + E_U[n(\mathbf{r})].$$

In the simplified rotationally invariant method (S. L. Dudarev *et al.*, Phys.Rev. **B**, 57, 1505 (1998)), the Hubbard term has the form:

$$E_U[n(\mathbf{r})] = \frac{U}{2} \sum_{\sigma} \text{Tr}[\mathbf{n}^{\sigma}(1 - \mathbf{n}^{\sigma})],$$

U being a (system-dependent) Coulomb repulsion (typically a few eV) and \mathbf{n}^{σ} is the matrix of orbital occupancies for a set of atomic-like states ϕ_m :

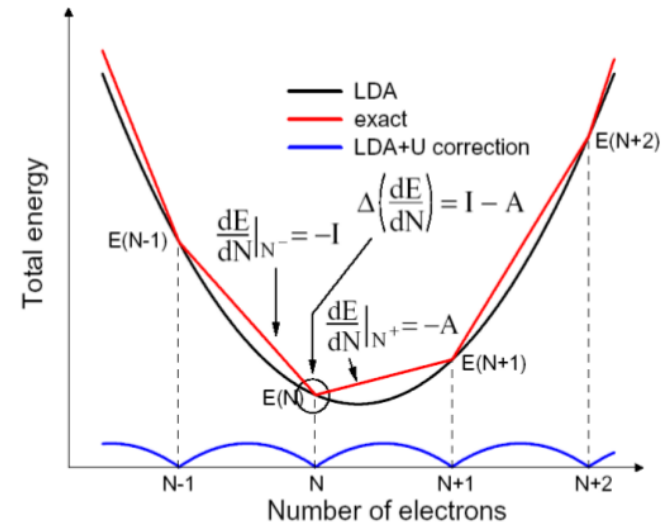
$$n_{mm'}^{\sigma} = \sum_{\sigma} \sum_i f_i^{\sigma} \langle \psi_i^{\sigma} | P_{mm'} | \psi_i^{\sigma} \rangle, \quad P_{mm'} = |\phi_m\rangle \langle \phi_{m'}|$$

The computation of additional U terms in the Kohn-Sham Hamiltonian and in the energy is not a significant overhead. DFT+U calculations are however often slow to converge (or do not converge to the desired ground state)

DFT+U (2)

There are several variants of DFT+U, with more parameters, and several possible choices of atomic-like states (simpler choice: atomic states, not orthogonalized)

DFT+U is an economical solution for a deep problem of DFT: the lack of discontinuity in approximated functionals, due to incomplete self-interactions cancellation, favors fractionary occupancy. In the picture, the behavior of a system connected to a reservoir of electrons, as a function of the occupancy N .



U can be taken as an adjustable parameter, but the above picture suggests that U can be computed from first principles extracting it from the fictitious curvature of the XC functional:

$$U \equiv \frac{\partial^2 E_{DFT}}{\partial n^2}, \quad n = \sum_{\sigma} \text{Tr}[\mathbf{n}^{\sigma}]$$

The procedure is described in detail in M. Cococcioni and S. de Gironcoli, Phys. Rev. B **71**, 035105 (2005). A linear-response based procedure is implemented in code HP.

Hybrid functionals

Hybrid functionals, such as B3LYP or PBE0, containing some amount of *exact* exchange, as in Hartree-Fock theory (spin-restricted case for simplicity):

$$E = T_s + \int n(\mathbf{r})V(\mathbf{r})d\mathbf{r} + E_H \underbrace{-\frac{e^2}{2} \sum_{i,j,\parallel} \int \frac{\psi_i^*(\mathbf{r})\psi_j(\mathbf{r})\psi_j^*(\mathbf{r}')\psi_i(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}'}_{E_x^{HF}}$$

In Hybrid DFT:

$$E = T_s + \int n(\mathbf{r})V(\mathbf{r})d\mathbf{r} + E_H + \overbrace{\alpha_x E_x^{HF} + (1 - \alpha_x) E_x^{DFT} + E_c^{DFT}}^{E_{xc}^{hyb}},$$

with $\alpha_x = 20 \div 30\%$. This is the method of choice in Quantum Chemistry, yielding very accurate results and correcting most GGA errors, at the price of an additional parameter (the amount of exact exchange).

QUANTUM ESPRESSO implements most popular hybrid functionals: B3LYP, PBE0, HSE, gau-PBE, for NC-PPs and (with some restrictions) US-PPs. In a plane-wave basis set, however, hybrid functionals are computationally very heavy.

Exact Exchange Potential with Plane Waves

The calculation of $\hat{V}_x \psi_{\mathbf{k},v}$ proceeds as follows:

- bring bands to real space with FFT:

$$\psi_{\mathbf{k},v}(\mathbf{G}) \xrightarrow{FFT} \psi_{\mathbf{k},v}(\mathbf{r})$$

- for each wave-vector \mathbf{q} and each occupied band v' , build “charge densities” $\rho_{\mathbf{q}}(\mathbf{r})$:

$$\rho_{\mathbf{q}}(\mathbf{r}) = \psi_{\mathbf{k}-\mathbf{q},v'}^*(\mathbf{r}) \psi_{\mathbf{k},v}(\mathbf{r})$$

- bring “charge” to reciprocal space with FFT, solve Poisson equation:

$$\rho_{\mathbf{q}}(\mathbf{r}) \xrightarrow{FFT} \rho_{\mathbf{q}}(\mathbf{G}), \quad V_{\mathbf{q}}(\mathbf{G}) = \frac{4\pi e^2}{|\mathbf{q} + \mathbf{G}|^2} \rho_{\mathbf{q}}(\mathbf{G})$$

- FFT back to real space, multiply by occupied band at $\mathbf{k} - \mathbf{q}$, add to the result:

$$V_{\mathbf{q}}(\mathbf{G}) \xrightarrow{FFT} V_{\mathbf{q}}(\mathbf{r}), \quad (\hat{V}_x \psi_{\mathbf{k},v})(\mathbf{r}) = (\hat{V}_x \psi_{\mathbf{k},v})(\mathbf{r}) + V_{\mathbf{q}}(\mathbf{r}) \psi_{\mathbf{k}-\mathbf{q},v'}(\mathbf{r})$$

In practical calculations, $\mathbf{k}' = \mathbf{k} - \mathbf{q}$ spans the same finite grid of k-points $\{\mathbf{k}\}$. Note that while symmetry can be exploited to compute only $\psi_{\mathbf{k}}$ with \mathbf{k} 's in the Irreducible Brillouin Zone, *all* $\psi_{\mathbf{k}-\mathbf{q}}$ are needed, not just those in the IBZ.

The $\mathbf{q} + \mathbf{G} = 0$ divergence

In periodic systems, the exact exchange contains a divergence when $\mathbf{q} + \mathbf{G} = 0$:

$$E_x^{HF} = -\frac{4\pi}{2\Omega} \times \frac{\Omega}{(2\pi)^3} \int d\mathbf{q} \sum_{\mathbf{G}} \frac{A(\mathbf{q} + \mathbf{G})}{|\mathbf{q} + \mathbf{G}|^2}$$

where

$$A(\mathbf{q} + \mathbf{G}) = \frac{\Omega}{(2\pi)^3} \int d\mathbf{k} |\rho_{\mathbf{k},v}^{\mathbf{k}-\mathbf{q},v'}(\mathbf{q} + \mathbf{G})|^2 \equiv \frac{1}{N_k} \sum_{\mathbf{k}} |\rho_{\mathbf{k},v}^{\mathbf{k}-\mathbf{q},v'}(\mathbf{q} + \mathbf{G})|^2$$

(the finite sum over N_k \mathbf{k} -points is what we actually compute) and

$$\rho_{\mathbf{k},v}^{\mathbf{k}-\mathbf{q},v'}(\mathbf{r}) = \psi_{\mathbf{k}-\mathbf{q},v'}^*(\mathbf{r}) \psi_{\mathbf{k},v}(\mathbf{r}).$$

This divergence is integrable, see: F.Gygi and A.Baldereschi, PRB 34, 4405 (1986)

Beware: this is good for 3D systems, not necessarily for 2D and 1D ones

Practical algorithm

- For any given grid of N_k \mathbf{k} -points, you need to define a subgrid of $N_q \leq N_k$ \mathbf{q} -vectors. In principle it should span all \mathbf{k} -points, in practice it may be smaller.
- Self-consistency is first achieved with closest LDA/GGA functional; then \hat{V}_x is turned on and a double self-consistency loop is started (self-consistency is reached at fixed \hat{V}_x , then \hat{V}_x is updated with new wavefunctions, and so on)
- Scaling: rather catastrophic. $\hat{V}_x\psi$ requires at least $\mathcal{O}(4N_kN_qM^2N \log N)$ floating-point operations for M bands and N points in the FFT grid, vs $\mathcal{O}(2N_kMN \log N)$ for $V_{xc}\psi$ in plain DFT. Tricks for speeding up the calculation:
 - Use a reduced \mathbf{q} -vector grid (when applicable): variables `nqx1`, `nqx2`, `nqx3`
 - Use a reduced cutoff for FFT's done in $\hat{V}_x\psi$: variable `ecutfock`
 - Use many processors, with the “band” parallelization

Even with these tricks, hybrid functional calculations are soooo slow.

Recent advances: ACE

A significant speedup can be obtained with Adaptively Compressed Exchange (ACE: Lin, JCTC, 12, 2242 (2016); Damle, Lin, Ying, JCTC, 11, 1463 (2015)). We project the exchange operator over the set of occupied orbitals:

$$\hat{V}_x^{ACE} = \sum_{i,j} \hat{V}_x |\psi_i\rangle \mathbf{M}_{ij}^{-1} \langle \psi_j | \hat{V}_x$$

where $\mathbf{M}_{ik} = \langle \psi_i | \hat{V}_x | \psi_j \rangle$ is the exchange matrix. The \mathbf{M}^{-1} is factorized via Cholesky decomposition and \hat{V}_x^{ACE} takes the form

$$\hat{V}_x^{ACE} = \sum_i |\xi_i\rangle \langle \xi_i|, \quad (1)$$

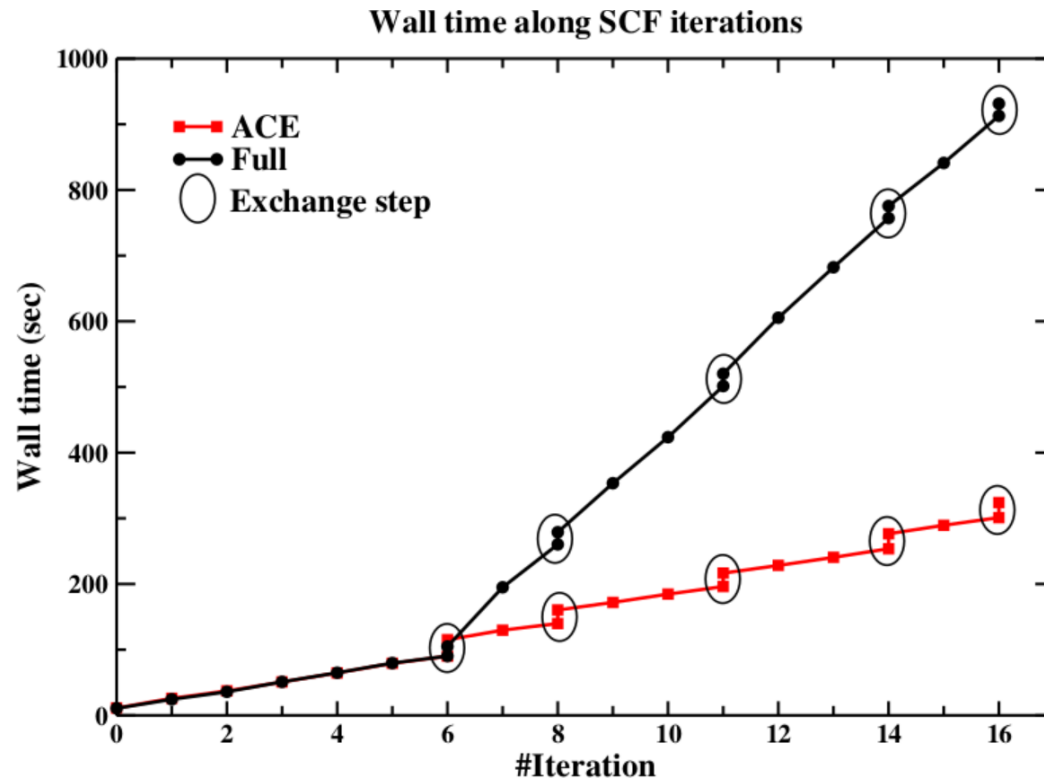
where \mathbf{L}^{-T} is the Cholesky factor of \mathbf{M}^{-1} ,

$$\xi_i = \sum_j \hat{V}_x |\psi_j\rangle \mathbf{L}_{ij}^{-T} \quad (2)$$

\hat{V}_x^{ACE} is used instead of \hat{V}_x in each fixed-exchange self-consistency loop.

Recent advances: ACE (2)

The ACE operator is perfectly equivalent to the true exchange operator on the manifold used for the projection, and the procedure converges to exactly the same results. The computation of ACE is *much* faster than the computation of the true exchange operator. In the picture below, a realistic case.



Recent advances: SCDM localization

The key to achieve even better performances is *orbital localization*. There are various ways to localize orbitals in real space, the most famous being the transformation to *Wannier orbitals*.

The density matrix, $\gamma(\mathbf{r}, \mathbf{r}') = \sum_j \psi_j^*(\mathbf{r}')\psi_j(\mathbf{r})$, is naturally localized, at least in insulators: for large $|\mathbf{r} - \mathbf{r}'|$, $\gamma(\mathbf{r}, \mathbf{r}') \sim \exp(-\alpha|\mathbf{r} - \mathbf{r}'|)$ asymptotically. This means that the functions $w(\mathbf{r}) = \sum_j c_j \psi_j(\mathbf{r})$, with $c_j = \psi_j^*(\mathbf{r}')$, are localized.

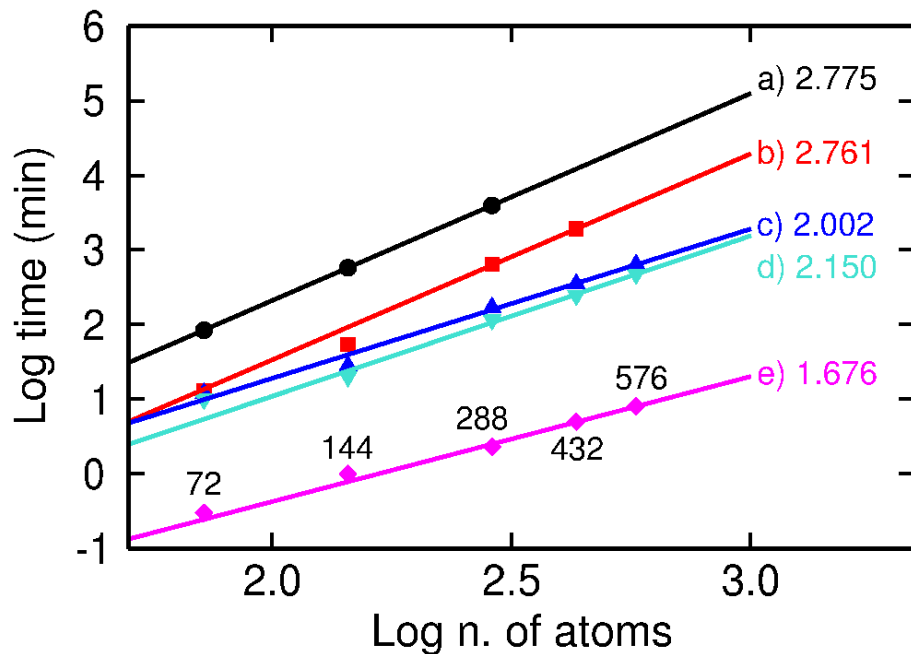
The set of w 's above is redundant: we need a way to select the best manifold among the many possibles, *without* storing the density matrix.

This can be achieved using the *selected columns of the density matrix* (SCDM) algorithm. This is an algebraic procedure that requires a standard QR decomposition (actually, only the pivoting part): Lin, *JCTC*, 12, 2242 (2016); Damle, Lin, Ying, *JCTC*, 11, 1463 (2015).

Once orbitals are localized, exchange integrals between non-overlapping orbitals can be discarded: we just need to choose a criterion and to set a threshold.

Recent advances: SCDM localization (2)

The localization is especially effective in large supercells with many atoms. In the figure below, the wall time for one SCF as a function of the number of atoms of various amorphous silica supercells ($T \propto N^\alpha$, α is the number reported to the right). The criterion for discarding exchange integrals is $S_{ij} = \int d\mathbf{r} |w_i(\mathbf{r})| |w_j(\mathbf{r})| < S_{thr}$

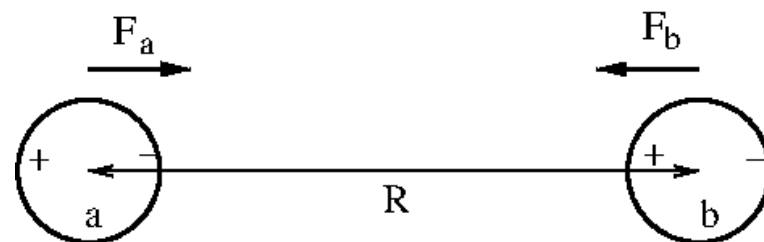


- a) PBE0 hybrid functional, no ACE
- b) PBE0 with ACE, no localization
- c) d) PBE0 with ACE and localization
 $S_{thr} = 0.002$ and $= 0.004$ respectively
- e) PBE

I. Carnimeo, S. Baroni, P. Giannozzi,
Electron. Struct. **1**, 015009 (2019)

Functionals for weak (vdW) interactions

None of the above-mentioned functionals properly accounts for van der Waals (or “dispersive”) forces. No functional based on the local charge/gradients/kinetic energy and/or exact exchange can contain vdW interactions, whose nature is *nonlocal*



Possible solutions include:

- Ignore van der Waals interactions altogether. Sometimes not possible, though
- Add them as a *semi-empirical* correction: very cheap, not very elegant, but often good enough for practical purposes
- As above, with parameters derived from first principles
- Introduce a truly *non-local* functional: more elegant and more computationally demanding, but still feasible

Semi-empirical van der Waals interactions

- **DFT+D** (Grimme): the vdW energy term is added to plain GGA as a semi-empirical correction,

$$E_{DFT+D} = E_{DFT} + E_{vdW}, \quad E_{vdW} = -\frac{s_6}{2} \sum_{i \neq j} \frac{C_6^{ij}}{R_{ij}^6} f_{damp}(R_{ij})$$

where s_6 is a global scaling factor depending upon the specific GGA;

$C_6^{ij} = \sqrt{C_6^{(i)} C_6^{(j)}}$ where $C_6^{(i)}$ are dispersion coefficients for the i -th atom;
a damping function

$$f_{damp}(R) = \frac{1}{1 + e^{-d(R/R_r - 1)}}$$

prevent singularities for $R \rightarrow 0$. Parameters are fitted to experimental or accurate theoretical data. Very cheap and simple, but hardly a first-principle approach.

- **DFT-D3**: a refined model with more parameters and more interaction terms

QUANTUM ESPRESSO implements both Grimme's DFT-D2 and DFT-D3.

Less-empirical van der Waals interactions

- **Tkatchenko-Scheffler**: the C_6 coefficients are computed from first principles, that is, from the charge density. Recent versions of QUANTUM ESPRESSO implement this approach (not yet available for Ultrasoft PP and PAW).
- **XDM** (exchange-dipole model) is another way to obtain interatomic coefficients for the vdW interactions. Available in QUANTUM ESPRESSO, works only for Ultrasoft PP and PAW.

Nonlocal vdW functionals

Nonlocal functionals account for van der Waals (dispersive) forces:

$$E_{nl} = \frac{1}{2} \int n(\mathbf{r}) \Phi(\mathbf{r}, \mathbf{r}') n(\mathbf{r}') d\mathbf{r} d\mathbf{r}', \quad \Phi(\mathbf{r}, \mathbf{r}') \equiv \Phi(n(\mathbf{r}), \nabla n(\mathbf{r}), n(\mathbf{r}'), \nabla n(\mathbf{r}'), |\mathbf{r} - \mathbf{r}'|)$$

Can be computed with a reasonable computational overhead (Soler's technique) yielding generally good results, but often overestimating binding.

QUANTUM ESPRESSO implements several flavors of such functional, notably: vdW-DF (Dion *et al.*, Thonhauser *et al.*), vdW-DF2 (Lee *et al.*), revised VV10 (Sabatini *et al.*, Vydrov and van Voorhis)

For practical usage, the kernel Φ should be computed and stored in a file.

- vdW-DF, vdW-DF2: run auxiliary code `generate_vdW_kernel_table.x`, store file `vdW_kernel_table` in the directory where pseudopotentials (PP) are read from;
- rVV10: run auxiliary code `generate_rVV10_kernel_table.x`, store file `rVV10_kernel_table` in the same directory as above.

In future versions the pre-computation of the kernel table will disappear