ENCCS Training Workshop

BigDFT Session

Flexibilities of wavelets as a computational basis set for large-scale electronic structure calculations

Luigi Genovese

L_Sim - CEA Grenoble - France

November 17, 2022

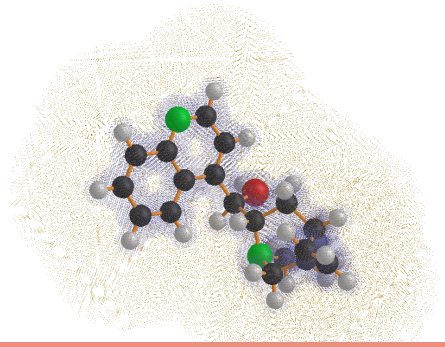
MaX Center of Excellence

Origin of the BigDFT project

STREP European project: BigDFT(2005-2008)

In the beginning: Four partners, 15 people
Now: around 10 active developers, Grenoble, Basel, Bristol, Catania, Trieste, Kobe
Used in production since twelve years.

Aim: To develop an ab-initio DFT code based on Daubechies Wavelets, to be *integrated in ABINIT*.
BigDFT 1.0 $\longrightarrow$ January 2008



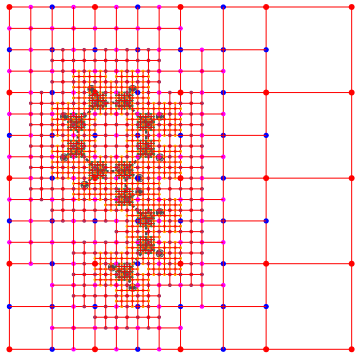
Why have we done this?

- Test the potential advantages of a new formalism
- ☛ A lot of outcomes and interesting results
- Future opportunities and ideas

Daubechies wavelets

Wavelets

A basis with optimal properties for expanding localised information

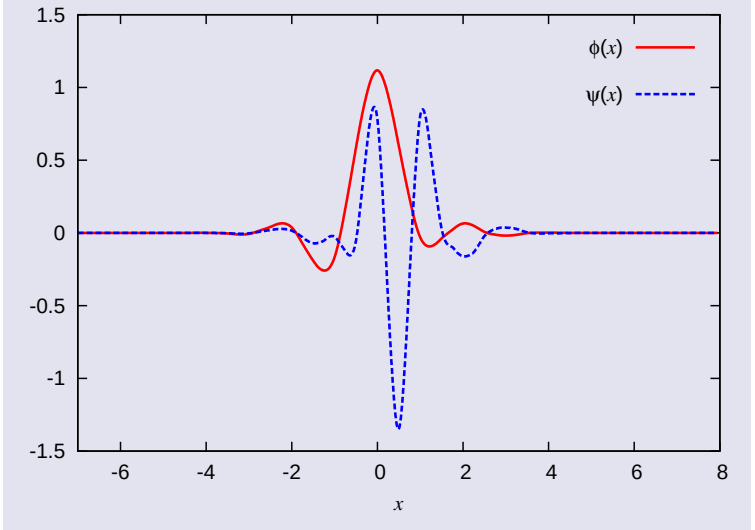


- Localised in real space
- Smooth (localised in Fourier space)
- Orthogonal basis
- Multi-resolution basis
- Adaptive
- Systematic

From early 80's

Applied in several domains
Interesting for DFT

Daubechies Wavelets



A brief description of wavelet theory

A Multi-Resolution real space basis

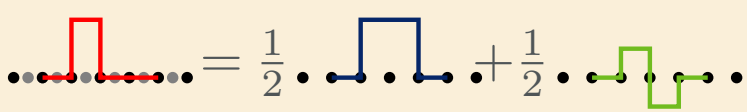
All functions w/ compact support, centered on grid points.
In the wavelet theory we have two kind of basis functions.

Scaling Functions (SF)

The functions of low resolution level are a linear combination of high-resolution functions.

Wavelets (W)

Contain the DoF needed to complete the information lacking due to the coarseness of the resolution.



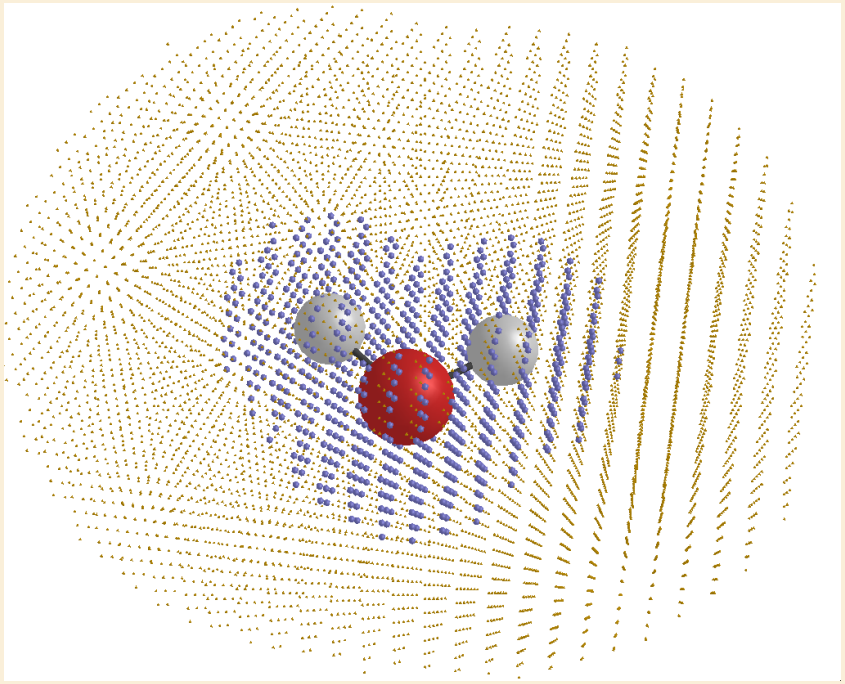
Increase the resolution without modifying grid space

SF + W = Degrees of Freedom of SF of higher resolution

Wavelet properties: adaptivity

Adaptivity

Resolution can be refined following the grid point.



The grid is divided in Low (1 DoF) and High (8 DoF) resolution points.
Points of different resolution belong to the same grid.
Empty regions must not be “filled” with basis functions.

Localization property,real space description

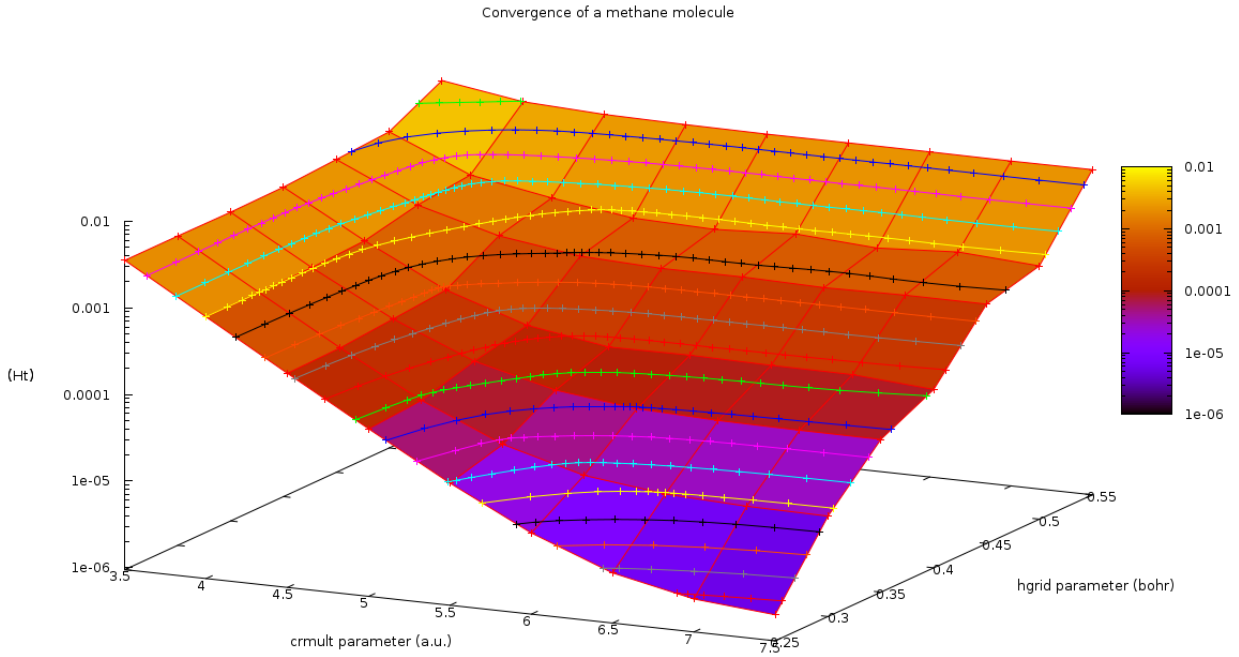
Optimal for big & inhomogeneous systems, highly flexible

Systematic basis set

The absolute accuracy of the calculation is directly proportional to the number of the basis functions

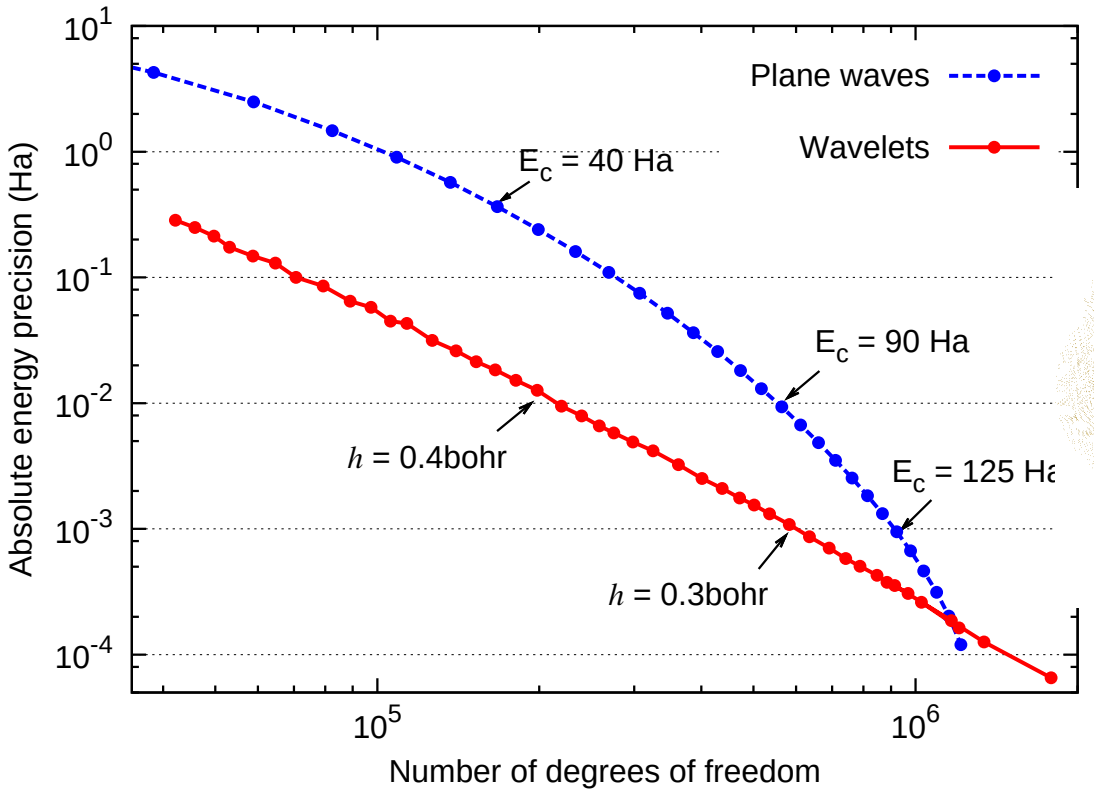
Two parameters for tuning the basis

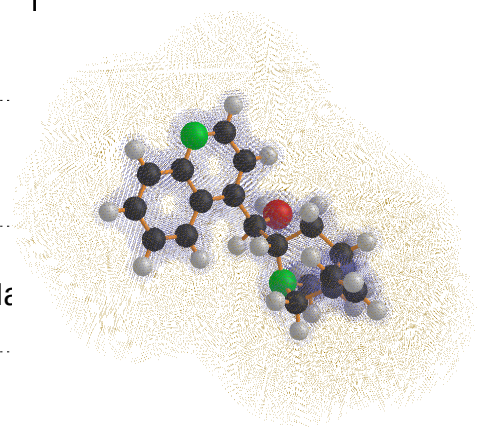
- The grid spacing `hgrid`
- The extension of the Low resolution points `crmult`



Optimal for inhomogeneous systems

Test case: cinchonidine molecule (44 atoms)



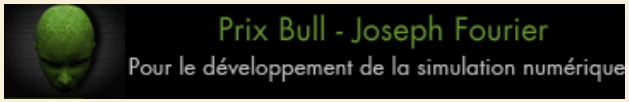


Enables a systematic approach for molecules

Considerably faster than Plane Waves codes.
the above run :10 (5) times faster than ABINIT (CPMD)
Charged systems can be treated *explicitly* with the same time

BigDFT: DFT with wavelets

A DFT code conceived for HPC (`www.bigdft.org`)

- DFT calculations up to many thousands atoms
- 👉 An award-winning HPC code 

- BigDFT has been conceived for massively parallel heterogeneous architectures since more than 10 years (MPI + OpenMP + GPU)

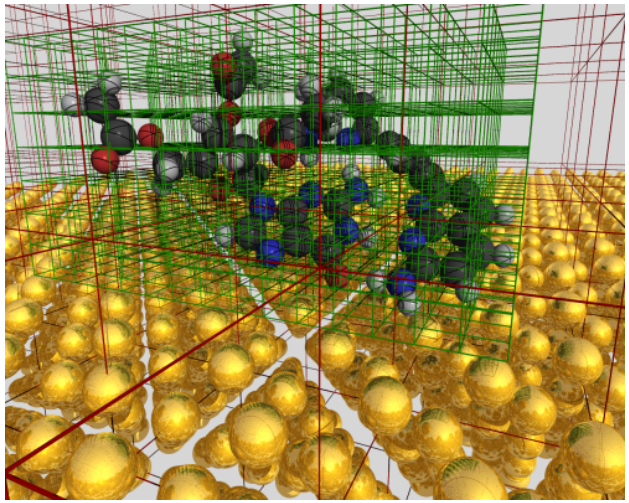
Code able to run routinely on different architectures

- GPU accelerators since the advent of double-precision GPGPU (2009)
- Various large calculation projects since 10 years
- ✔ A code conceived for supercomputers

A flexible formalism

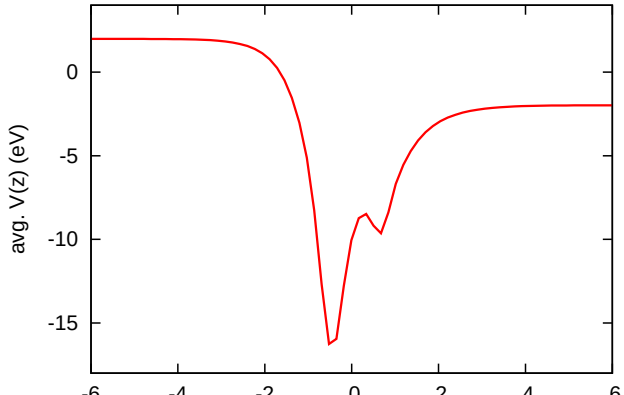
Flexible Boundary Conditions

- Isolated (free) BC
- Wires BC
- Surfaces BC
- Periodic (3D) BC



Systematic approach

Only relevant degrees of freedom are taken into account
Boundary conditions can be implemented explicitly



E.g.: Surfaces BC

2D Periodic + 1D isolated
Optimal to treat dipolar systems without corrections

A code both for Solid-State and Quantum Chemistry

- 3D periodic, Surfaces and Free BC (← Poisson Solver)
- Very high precision (analytic KS operators)
- Usage of analytic HGH pseudopotentials
- AE accuracy, benchmarked in G2-1, S22, DeltaTest

Present functionalities

Traditional functionalities for GS Kohn-Sham DFT (including metals, Hybrid Functionals), LR-TDDFT, empirical VdW
Exhaustive library of Structural Prediction, $O(N)$ calculations

Available Functionalities

Non-Exhaustive List of Functionalities

	$\mathcal{O}(\mathcal{N}^3)$	$\mathcal{O}(\mathcal{N})$	fragment
MPI and OpenMP	✔	✔	✔
GPUs	✔	✘	✘
free/wire/surface boundary conditions	✔	✔	✔
periodic orthorhombic cells	✔	✔	✔
periodic non-orthorhombic cells		in progress	
k-points	✔	✘	✘
forces (geometry optimizations, MD)	✔	✔	✘
metals	✔	✔	✔
hybrid functionals (no k-points)	✔	✘	✘
spin polarization	✔	✔	in progress
explicit charges (free BC only)	✔	✔	✔
external electric field (free/surface BC only)	✔	✔	✔
electrostatic embedding	✔	✔	✔
structure searching	✔	✔	✔
empirical Van der Waals (free BC only)	✔	✔	✔
Raman spectra		in progress	
time-dependent DFT	✔	✘	✘
constrained DFT (no spin or forces)	✘	in progress	✔

Interpolating SF Poisson Solver

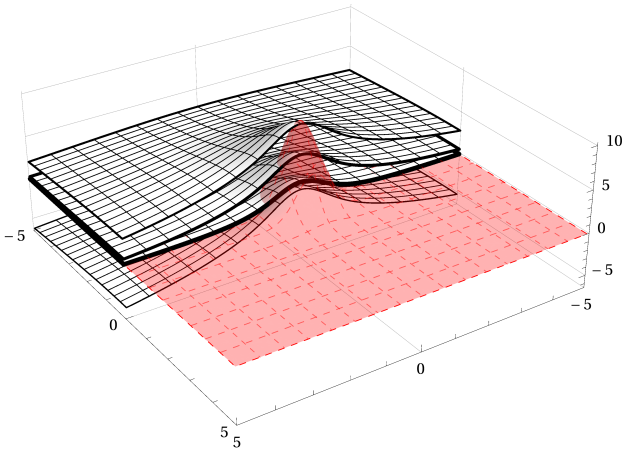
(Screened) Poisson Equation for any BC in vacuum

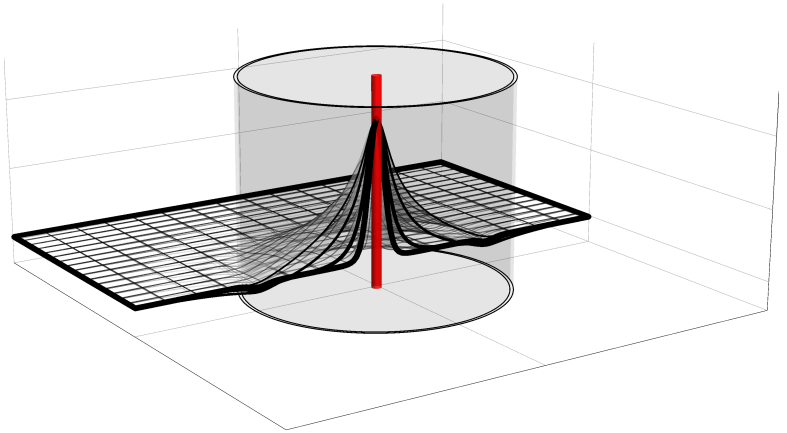
Non-orthorhombic cells (periodic, surface BC):

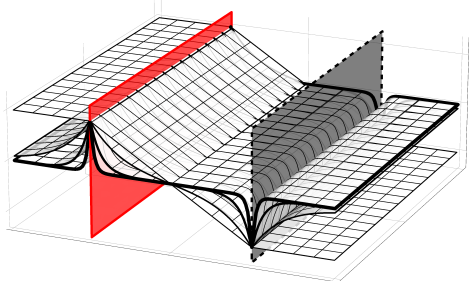
$$(\nabla^2 - \mu_0^2) V(x,y,z) = -4\pi\rho(x,y,z)$$

Machine-precision accuracy J. Chem. Phys. **137**, 13 (2012)

Extended to implicit solvents (JCP 144, 014103 (2016))





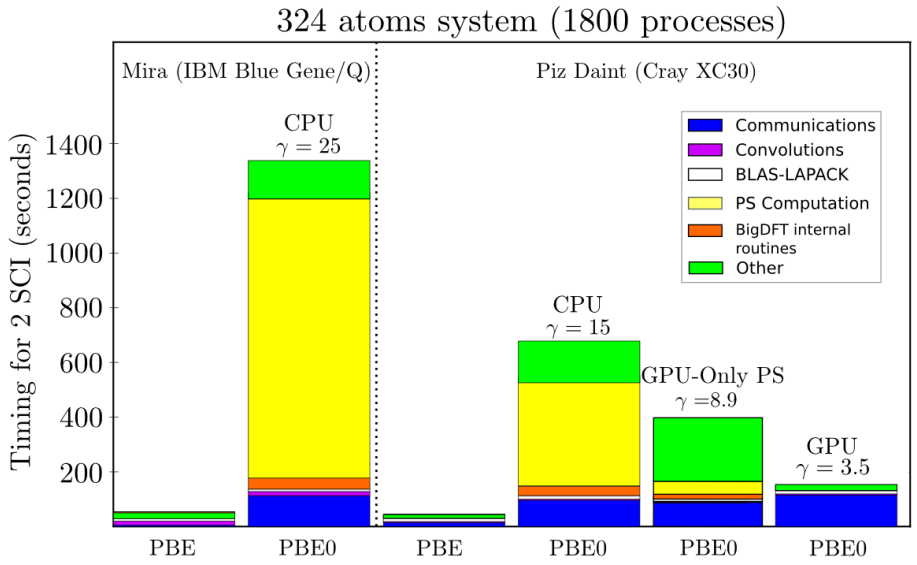


Future developments

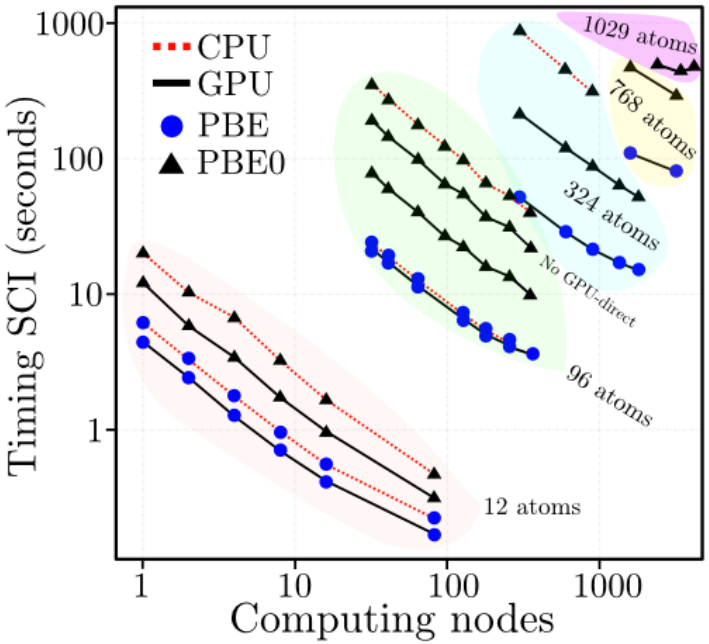
Range-separated

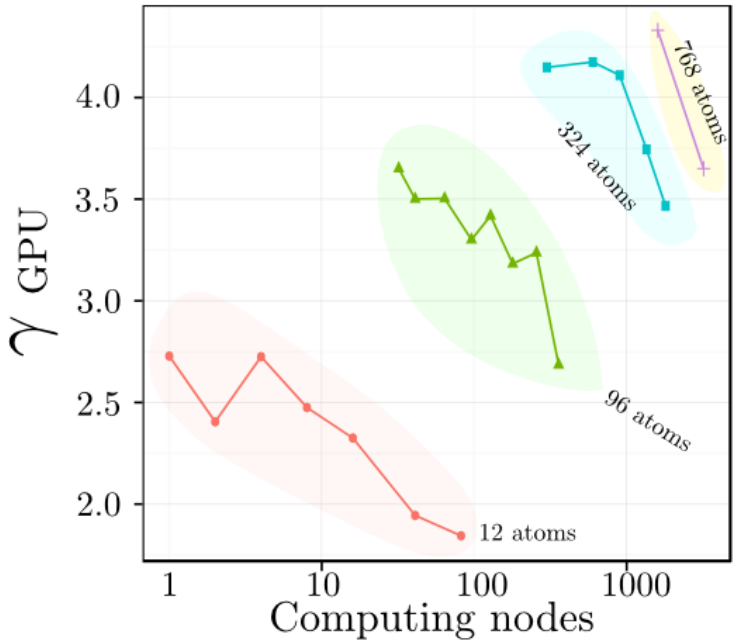
$$\frac{1}{r}\left[\mathrm{erf}\frac{r}{r_0} + \mathrm{erfc}\frac{r}{r_0}\right]$$

Hybrid Functionals (JPCM 30 (9),095901 (2018))



UO_2 systems:	
Atoms	Orbitals
12	200
96	1432
324	5400
768	12800
1029	17150





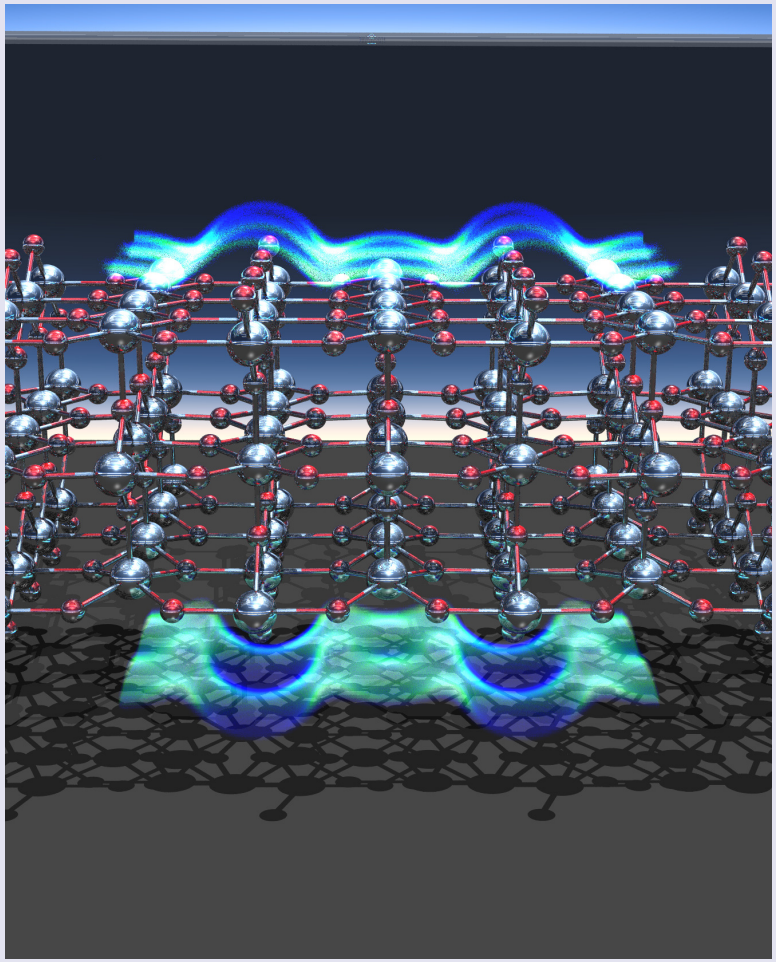
Polarizable Continuum Models

Poisson solver for implicit solvents (2016) JCP 144, 014103

Allows an efficient and accurate treatment of implicit solvents
The dielectric function determine the cavity where the solute is defined.
The cavity can be

- rigid (PCM-like)
- determined from the Electronic Density (SCCS approach)

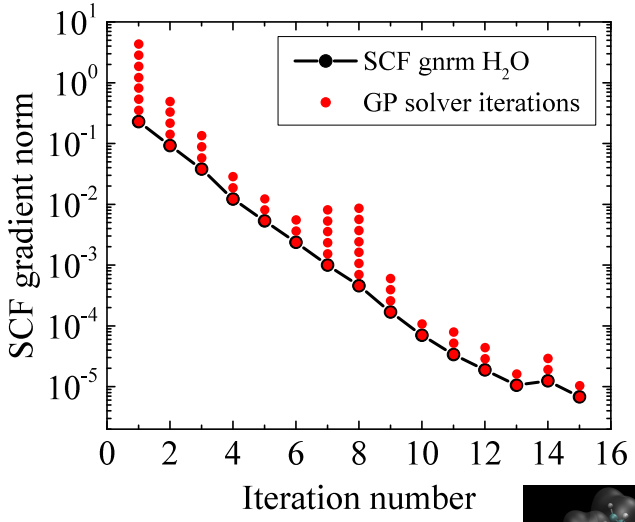
☛ Can treat various BC (here TiO_2 surface)

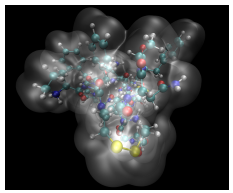


Performances in full SCF runs

Blackbox-like usage

The Generalized PS only needs few iterations of the vacuum poisson solver

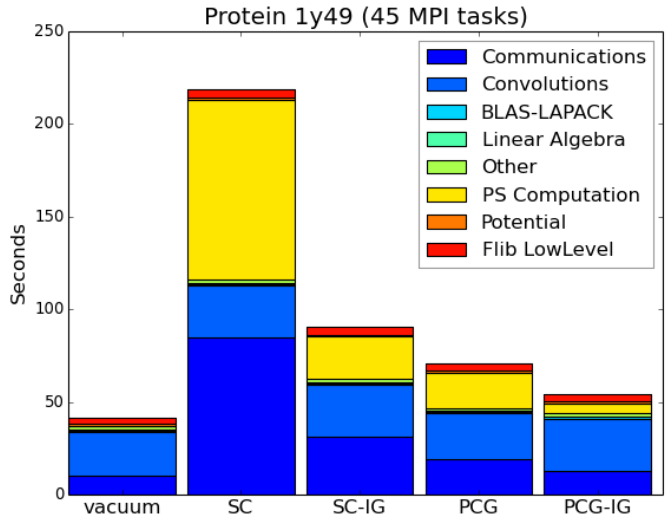




Time-to-solution

Timings for the protein PDB ID: 1y49 (122 atoms) in water

- Full SCF convergence 49 s
- Solvent/vacuum runtime ratio $\alpha = 1.16$



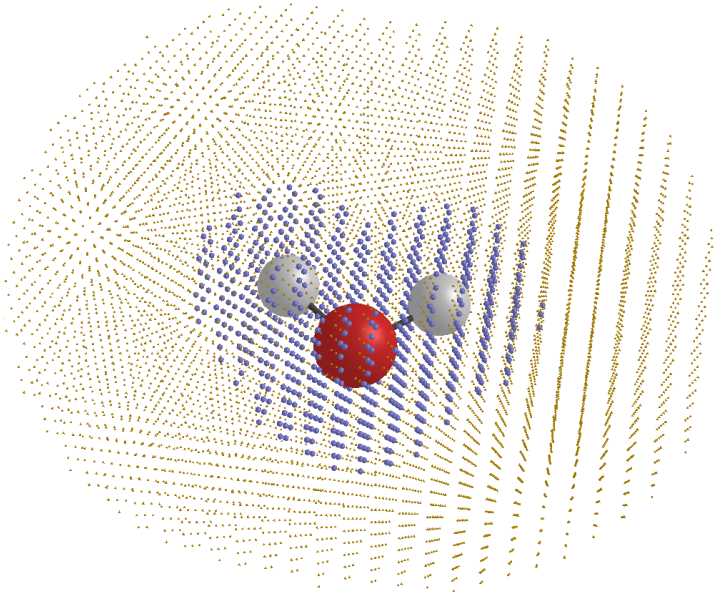
Use locality of the basis set

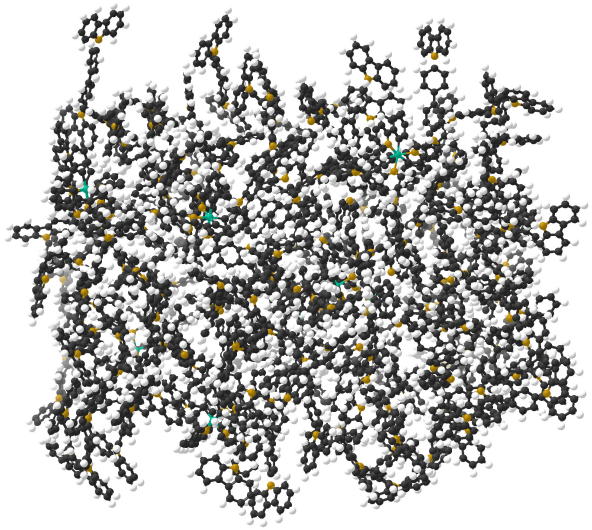
Wavelets

an ideal basis for electronic structure calculations – flexible, systematic and localized

Linear-scaling DFT

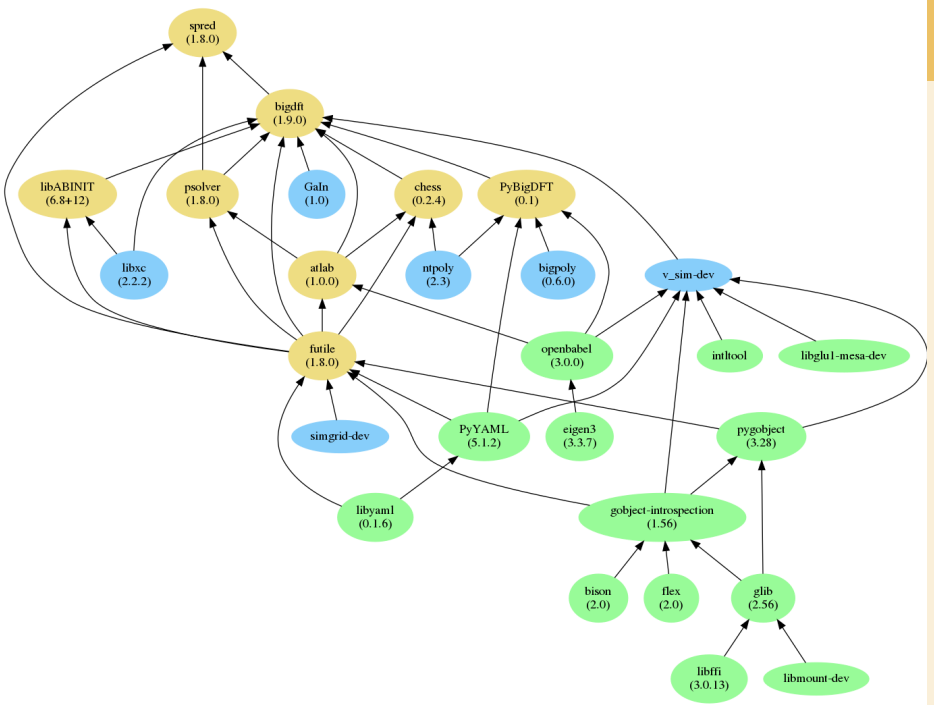
allows us to access very large system sizes via the use of a localized minimal basis set





Combining the two is now possible!

Code release and distribution



Modularity first

BigDFT-suite : collection of different independent libraries with own build system.
Third-party libraries (green) and upstream modules (blue)
☛ Dependencies expressed easily in the jhbuild-based bundler.

- Lots of possible options
- Very versatile
- Python configuration files can be shared, many provided

This training

Google Colab-ready notebooks

In this training you will...

- Have an overview of BigDFT code API
- See how to employ some of the functionalities of this code from a local workstation
- Work to some pre/post processing of the code data/results
- Run some calculations BigDFT in a production environment (supercomputer)

In this training you will **not**...

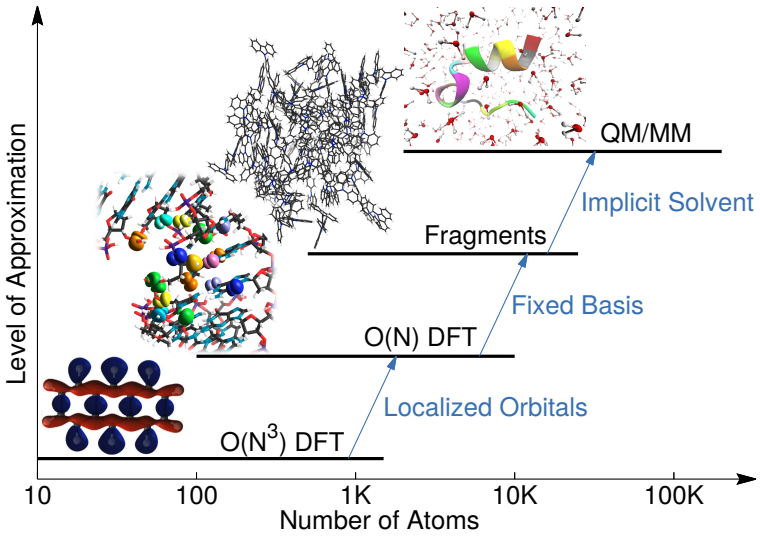
- Perform a throughout overview of the functionalities
- Have lot of time to inspect code performance

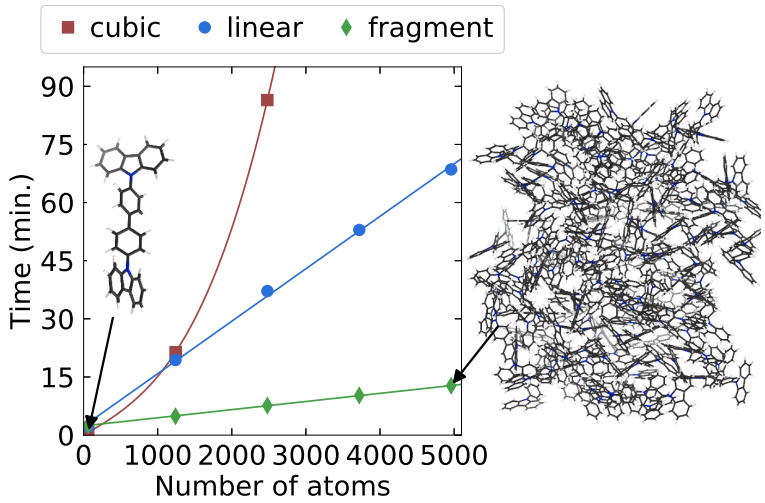
☛ bring BigDFT back home!

From Cubic Scaling to Multiscale

Across Lengthscales

- extended orbitals $\rightarrow O\left(N^3\right)$
- exploit locality $\rightarrow O\left(N\right)$
- exploit repetition $\rightarrow \downarrow$ cost $O\left(N\right)$
- larger systems $\rightarrow$ increasing complexity
- $\rightarrow$ how to treat complex systems?

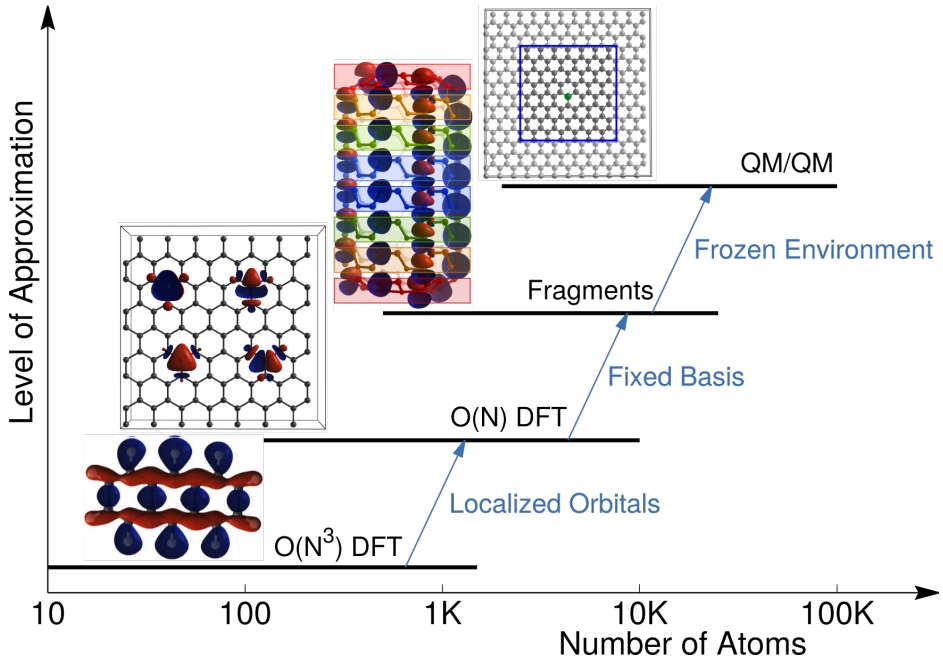




Features of the approach

Across Lengthscales with Wavelets

- three methods in BigDFT with differing levels of approximation fragment $\rightarrow$ linear $\rightarrow$ cubic
- approximations are controllable – can estimate or measure errors



A new mindset is emerging

Received: 28 April 2021 | Revised: 31 August 2021 | Accepted: 1 September 2021
DOI: 10.1002/wcms.1574

ADVANCED REVIEW

WIREs COMPUTATIONAL MOLECULAR SCIENCE WILEY

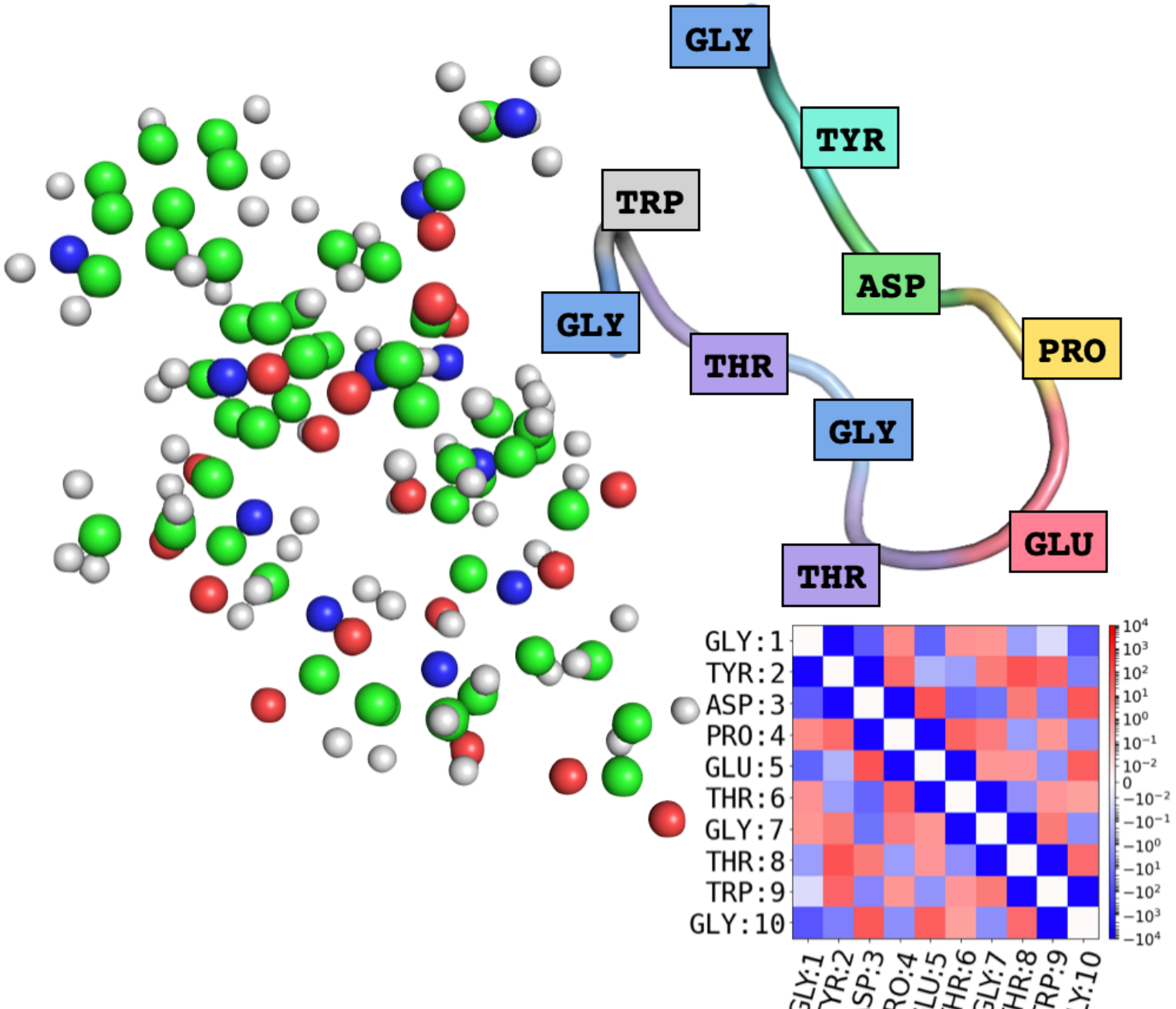
Density functional theory calculations of large systems: Interplay between fragments, observables, and computational complexity

William Dawson[1] | Augustin Degomme[2] | Martina Stella[3] | Takahito Nakajima[1] | Laura E. Ratcliff[3] | Luigi Genovese[2]

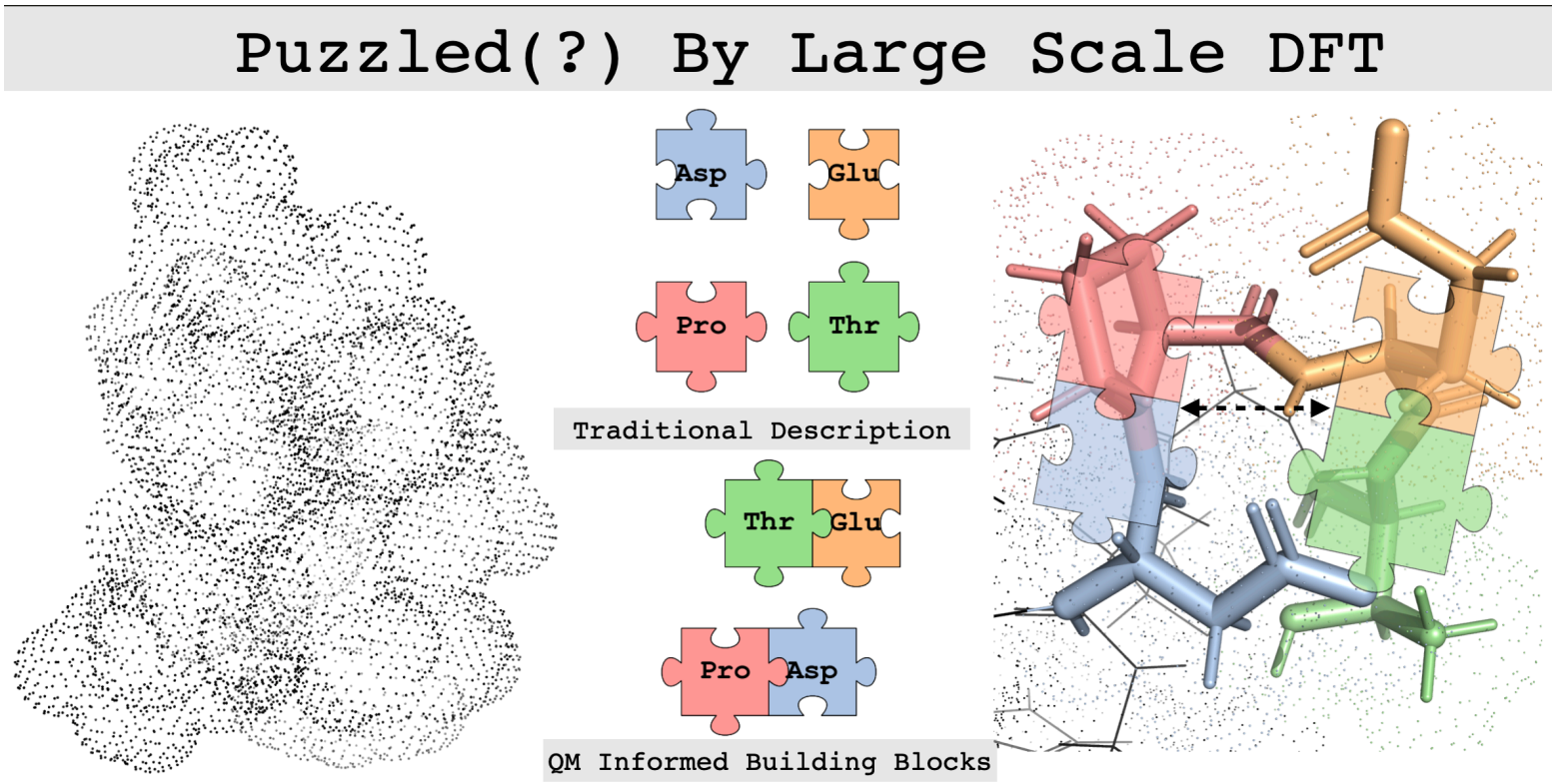
[1]RIKEN Center for Computational Science, Kobe, Japan
[2]Université Grenoble Alpes, INAC-MEM, L_Sim, Grenoble, France
[3]Department of Materials, Imperial College London, London, UK

Abstract
In the past decade, developments of computational technology around density functional theory (DFT) calculations have considerably increased the system sizes which can be practically simulated. The advent of robust high performance computing algorithms which scale linearly with system size has

Example: fragment in peptides



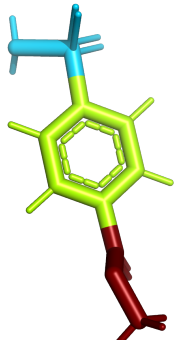
Example 2: the same in protein



Automatic Fragmentation of Systems

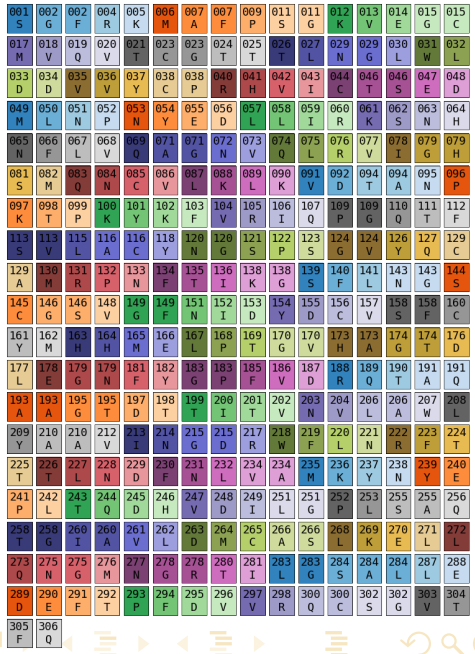
Small Molecules - From Atoms Up

- Automatic: We can re-organize a system into fragments without prior knowledge.
- Robust: Non-expert DFT users can interpret the information coming out of DFT calculations.



Large Biomolecules

- Proteins are often already divided into fragments based on their Amino Acids.
- Yet not all amino acids are equally good fragments. We can combine them together to build a more coherent picture.

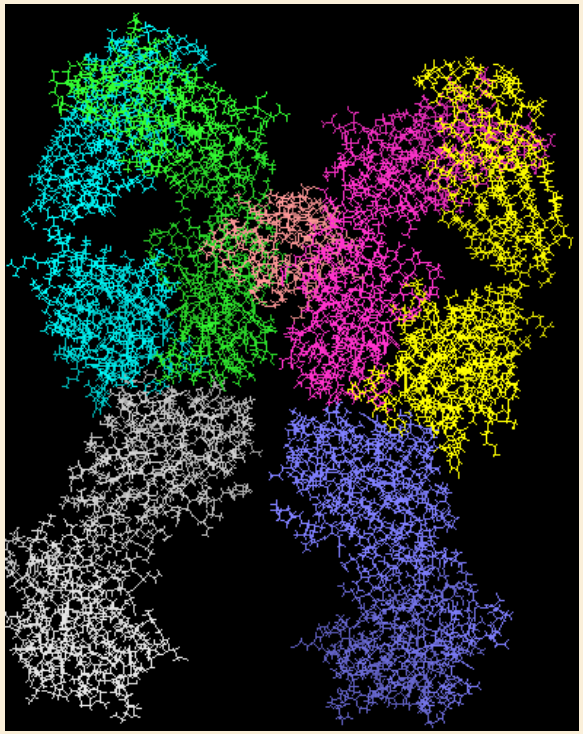


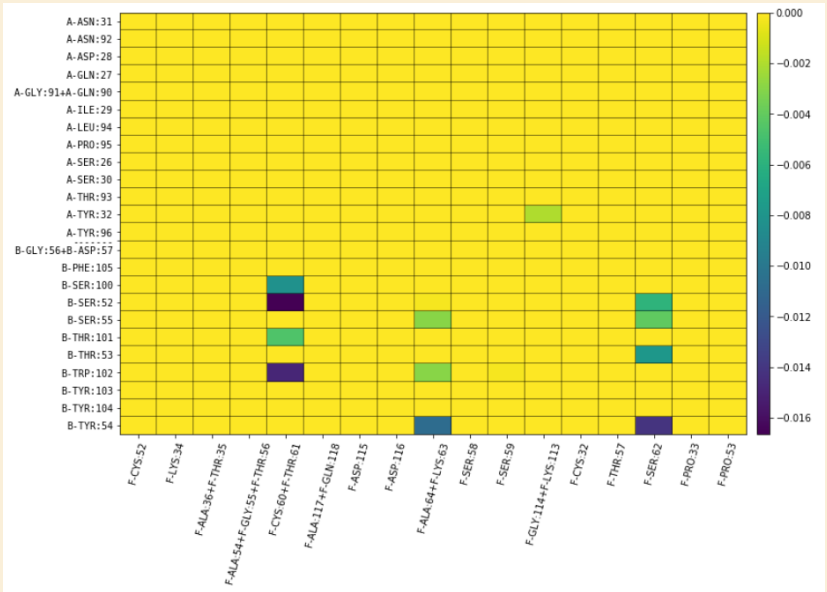
Lots of Systems of interest in Biology

Large systems are routinely accessible

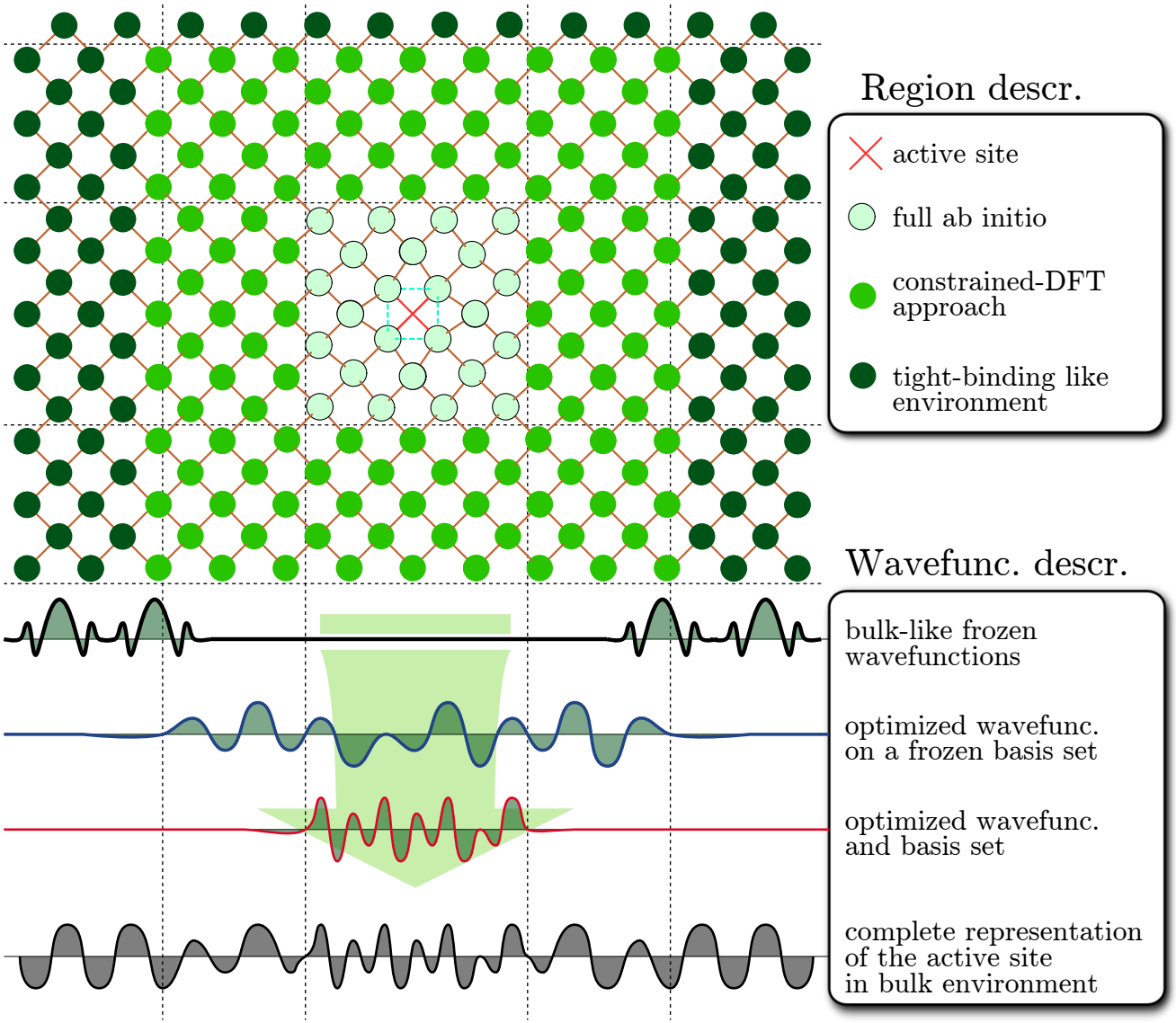
Linear Scaling code $\rightarrow$ CPU time *per atom*
Example: 1400 Residues (One Monoclonal Antibody); 22 thousand atoms: 1.2h of walltime on 32 nodes of IRENE-Rome Machine

Reduce (identify) the interactors in a biological system

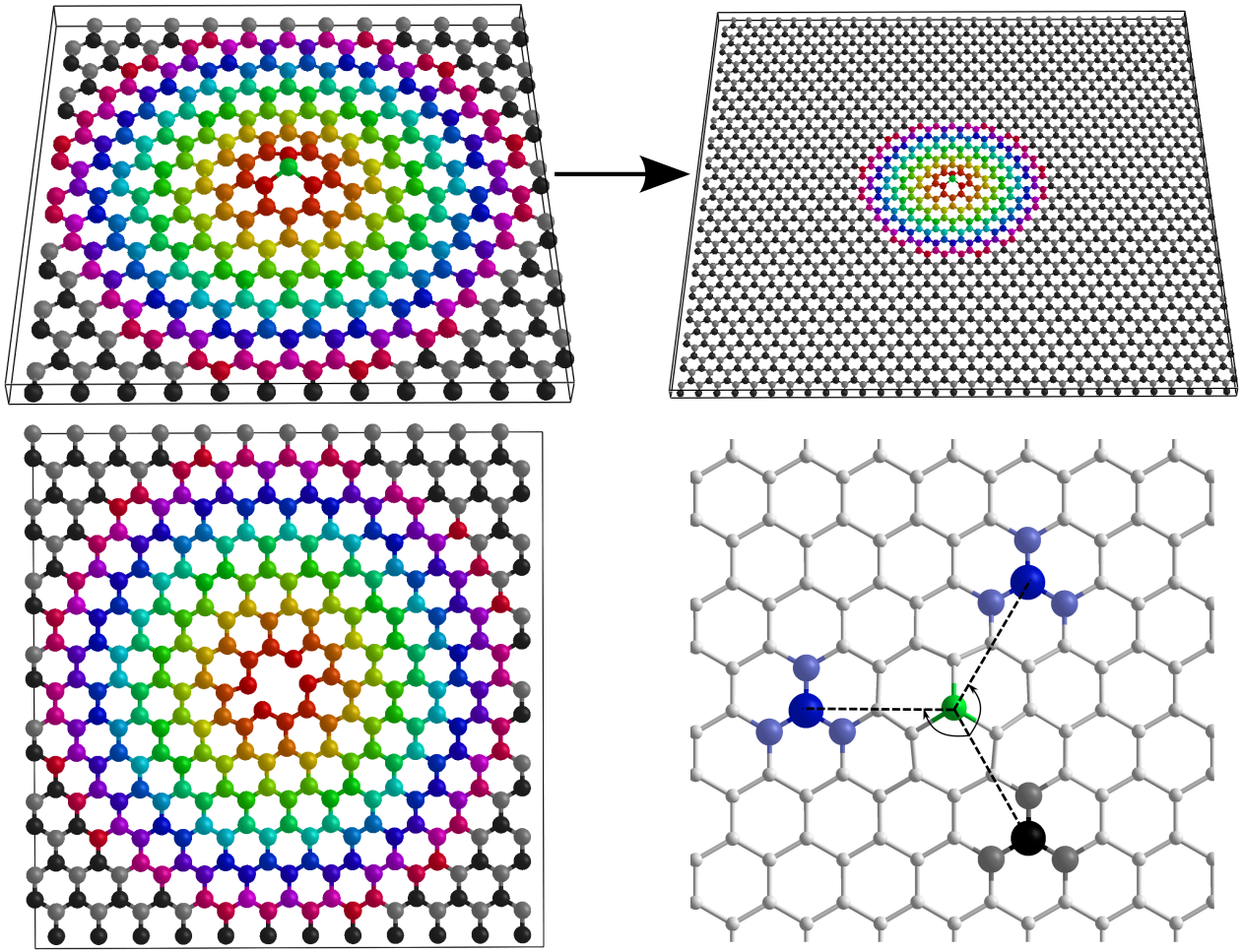




Perspectives (QM/QM)



Defective Graphene with Fragments



close to the defect:
strong perturbation

far from the defect:
bulk-like behaviour

What we can reach with this technique

Quantum Mechanics (DFT) may be needed

- Whenever DFT is necessary to study the electronic structure of the systems, it is important to provide the tools to *interpret* experimental data
- Need of new tools developed especially for the study of biological systems

☛ The BigDFT code provide a new paradigm of analysis

Main ingredients

- PDB files from neutron crystallography, Cryo-TEM, MD simulations, ...
- Remotely accessible (super) computing platform
- A post-processing infrastructure easy-to-use

Quantum-as-a-Service approach

A dedicated Users' platform ?

Collaboration L_Sim (CEA Grenoble) and CS Group (ILL)

PyBigDFT

Pre- and Post-processing of simulations are performed via a `Python` module

HPC

Calculations triggered remotely on a super-computer from a Jupyter notebook (AiiDA framework)

Userclub

Simulation can be processed from a platform next to experimental data (ILL User Club access)

Database

Large databases of biological systems can be created

New insights for (neutron) data analysis

- Dedicated routine for neutron crystallography data interpretation
- Possible improvement from other structural (e.g. SANS) data

In a nutshell

Summary

- DFT shouldn't be employed for large systems just on the hope of accuracy, but instead with the goal of insight.
- Complexity Reduction - We have developed a way to use information from DFT to generate coarse-grained views of a system by defining reliable fragments and measuring their interaction.

From Material Science to other communities

- The complexity reduction framework presented here originate from our expertise on Physics and Material Science.
- Postprocessing can be performed even by non-specialtists.
- This combination create interesting opportunities for interdisciplinar collaborations.

Interdisciplinary considerations

Discussion with biologists

We are not referring to a set of established techniques:

- New objects, definitions, descriptors
- The Physico-Chemical outcome (and only this!) should be highlighted

Difference (I have found) in the approach: example

- For a Physicist the *procedure* is the ground basis for the result
- For Biologists the *result* is the ground basis for the procedure

Interdisciplinarity requires

Rigor, Trust, Vision, Committment ☛ The right guys

Acknowledgments

Bristol L. E. Ratcliff

Trieste M. Stella

Kobe W. Dawson, T. Nakajima

Boston M. Zaccaria, B. Momeni

Grenoble L. Beal, S. Dechamps, D. Caliste, LG

ILL V. Cristiglio

Rome M. Reverberi

Catania G. Fisicaro