

BigDFT Session

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

The BigDFT project
Wavelets

BigDFT compilation

Poisson Solver
Implicit Solvents

Perspectives
In Practice

Opportunities

Luigi Genovese

L_Sim - CEA Grenoble - France

February 2, 2024

CECAM

Basis sets for DFT calculations



BigDFT

L. Genovese

Plane Waves

ABINIT, CPMD, VASP, ...
Systematic convergence

- ✓ Accuracy increases with the number of basis elements
- ✓ Non-localised, optimal for periodic systems
- ✗ Non adaptive

Gaussians, Slater Orbitals

CP2K, Gaussian, AIMPRO, ...
Real space localized

- ✓ Small number of basis functions for moderate accuracy
- ✓ Well suited for molecules and other open structures
- ✗ Non systematic

FFT

Robust, easy to parallelise

Analytic functions

Kinetic and overlap matrices are calculated analytically



The BigDFT project

Wavelets

BigDFT compilation

Poisson Solver
Implicit Solvents

Perspectives
In Practice

Opportunities

Origin of the BigDFT project



BigDFT

L. Genovese

STREP European project: BigDFT(2005-2008)

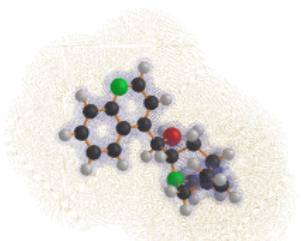
In the beginning: Four partners, 15 people

Now: around 10 active developers, Grenoble, Bristol, Kobe,
Catania, Trieste, Boston

Used in production since fifteen years.

Aim: To develop an ab-initio DFT code
based on **Daubechies Wavelets**, to be
integrated in ABINIT.

BigDFT 1.0 → 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

The BigDFT
project

Wavelets

BigDFT
compilation

Poisson Solver
Implicit Solvents

Perspectives
In Practice

Opportunities

Daubechies wavelets



BigDFT

L. Genovese

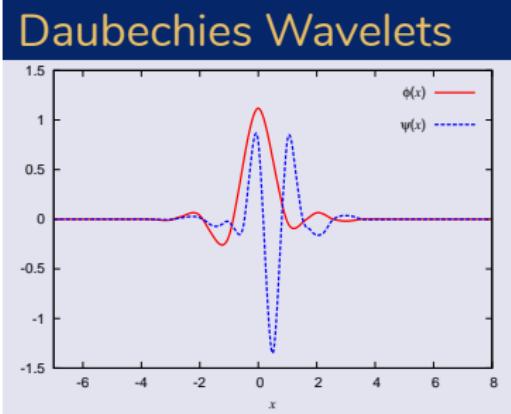
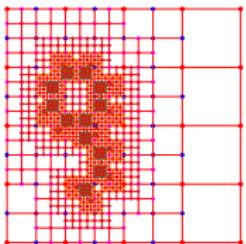
Wavelets

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



A brief description of wavelet theory



BigDFT

L. Genovese

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**.

$$\dots \square \square \dots = \dots \square \dots + \dots \square \dots$$

Wavelets (W)

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

$$\dots \square \dots = \frac{1}{2} \dots \square \dots + \frac{1}{2} \dots \square \dots$$

Increase the resolution without modifying grid space

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

The BigDFT project

Wavelets

BigDFT compilation

Poisson Solver
Implicit Solvents

Perspectives
In Practice

Opportunities

Wavelet properties: adaptivity

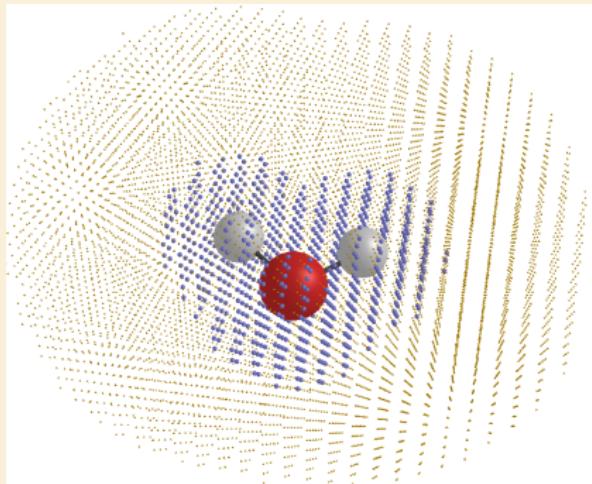


BigDFT

L. Genovese

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**

The BigDFT project

Wavelets

BigDFT compilation

Poisson Solver

Implicit Solvents

Perspectives

In Practice

Opportunities

Tensor product decomposition of the basis

The 3D basis is separable in 1D SF/ W.

$$\phi_{j_x, j_y, j_z}^{e_x, e_y, e_z}(x, y, z) = \phi_{j_x}^{e_x}(x) \phi_{j_y}^{e_y}(y) \phi_{j_z}^{e_z}(z)$$

(j_x, j_y, j_z) are the grid points, $\phi_j^{(0)}$ and $\phi_j^{(1)}$ the SF and the W.

Orthogonality, scaling relation

Daubechies wavelets are **orthogonal** and **multi-resolution**

$$\int dx \phi_k(x) \phi_\ell(x) = \delta_{k\ell} \quad \phi(x) = \frac{1}{\sqrt{2}} \sum_{j=-m}^m h_j \phi(2x - j)$$

Hamiltonian-related quantities are calculated **analytically**

The accuracy is only limited by the basis set ($O(h_{\text{grid}}^{14})$)

Exact evaluation of kinetic energy

Expressed **analytically** by a convolution:

$$f(x) = \sum_{\ell} c_{\ell} \phi_{\ell}(x), \quad \nabla^2 f(x) = \sum_{\ell} \tilde{c}_{\ell} \phi_{\ell}(x),$$

$$\tilde{c}_{\ell} = \sum_j c_j a_{\ell-j}, \quad a_{\ell} \equiv \int \phi_0(x) \partial_x^2 \phi_{\ell}(x),$$

From N^3 to $3N$ calculations for separable objects

We save computational time when performing scalar products with separable functions (e.g. gaussians).

$$\int d\mathbf{r} \psi(\mathbf{r}) e^{-\frac{1}{2} \left(\frac{r}{r_a} \right)^2} = \sum_{ijk} c_{ijk} d_i d_j d_k, \quad d_i = \int \phi_i(x) e^{-\frac{1}{2} \left(\frac{x}{r_a} \right)^2}$$

d_i coefficients can be calculated at **machine precision**.

GTH-HGH: analytic and separable pseudopotentials

Wavelet properties: adaptivity

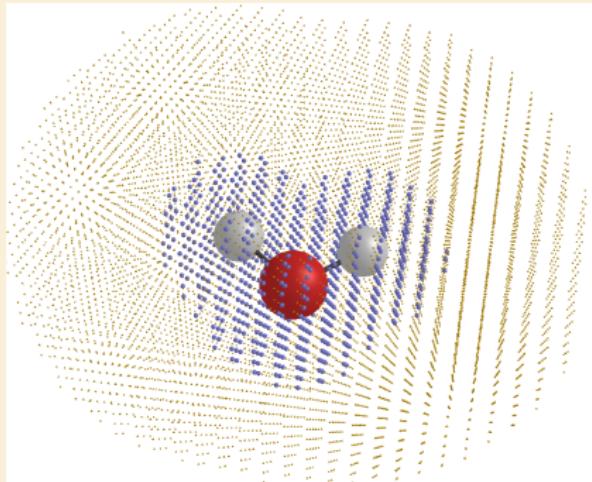


BigDFT

L. Genovese

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**

The BigDFT project

Wavelets

BigDFT compilation

Poisson Solver

Implicit Solvents

Perspectives

In Practice

Opportunities

Systematic basis set



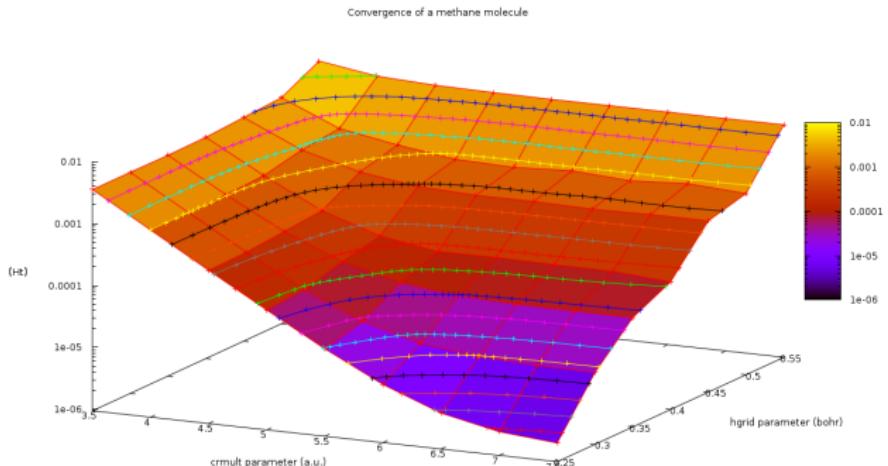
BigDFT

L. Genovese

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`



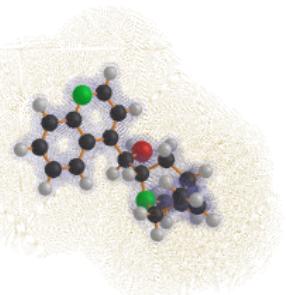
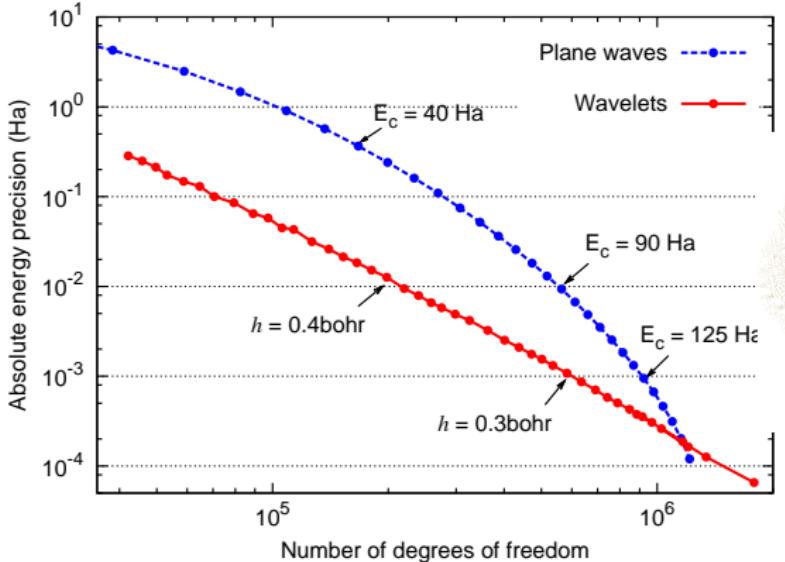
Optimal for inhomogeneous systems



BigDFT

L. Genovese

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

The BigDFT project

Wavelets

BigDFT compilation

Poisson Solver

Implicit Solvents

Perspectives

In Practice

Opportunities

All-electron precision with HGH PSP



BigDFT

L. Genovese

Non-linear core correction

J. Chem. Phys. **138**, 104109 (2013)

- Simple analytic form (a single gaussian as ρ_c)
- Same hardness as HGH
 - ➡ a systematic localized basis is **fundamental**

Precision considerably improved

(energies)

G2-1 test set (Atomization

kcal/mol	MAD	RMSD	MSD	maxAD	minAD
Old HGH	6.85	9.13	-6.76	23.94	0.10
NLCC-HGH	0.51	0.63	0.16	1.50	0.03
PAW Paier	0.46	0.56	-0.43	1.13	0.01
ΔAE (geopt)	0.29	0.70	-0.29	4.21	0.00

AE precision for quantities in different environments

Bond lengths, Pressure (Bulk systems), Dispersion-corrected interaction energies, ...



The BigDFT project

Wavelets

BigDFT compilation

Poisson Solver
Implicit SolventsPerspectives
In Practice

Opportunities

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

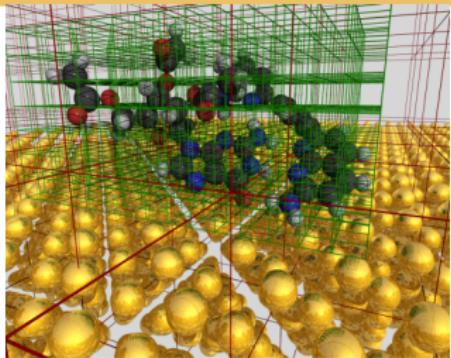
- GPU accelerators since the advent of double-precision GPGPU (2009)
- Various large calculation projects since 10 years
- ✓ A code conceived for supercomputers

A flexible formalism



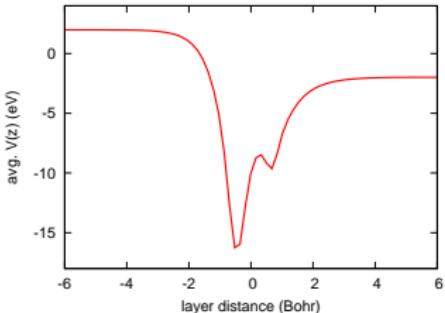
Flexible Boundary Conditions

- 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 (\leftarrow 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



BigDFT

L. Genovese

Non-Exhaustive List of Functionalities

	$O(\mathcal{N}^3)$	$O(\mathcal{N})$	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	✓

The BigDFT
project

Wavelets

BigDFT
compilation

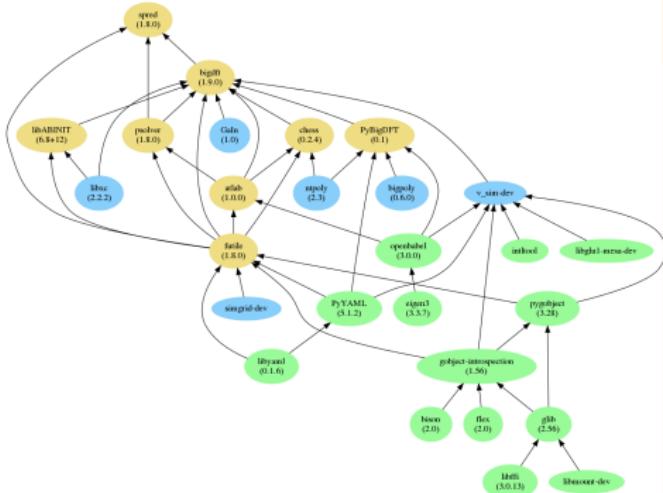
Poisson Solver

Implicit Solvents

Perspectives

In Practice

Opportunities



Modularity first

BigDFT-suite : collection of different independent libraries with own build system.

Third-party libraries (green) and upstream modules (blue)

Dependecies expressed easily in the jhbuild-based bundler.

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

Virtual Machine

- BigDFT is part of MaX flagship codes, is available on the Quantum Mobile virtual machine.
- Great for training/schools

► QM

Package

- Debian package in the making: Easy to install, less optimized.
- Python package for BigDFT run analysis
BigDFT client (futil and PyBigDFT)
- Conda package recipe (release candidate)

In this training you will...

- Get acquainted with BigDFT code API
- Have an overview of how to employ some of the functionalities of this code
- Work to some pre/post processing of the code data/results

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

- Run BigDFT in a production environment (supercomputer)
- Have a realistic idea of the performance of the code
- Have too much time to play with the code runs

👉 bring BigDFT back home!

Interpolating SF Poisson Solver



BigDFT

L. Genovese

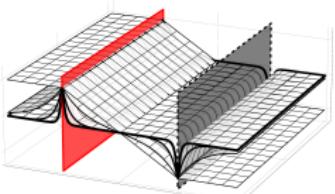
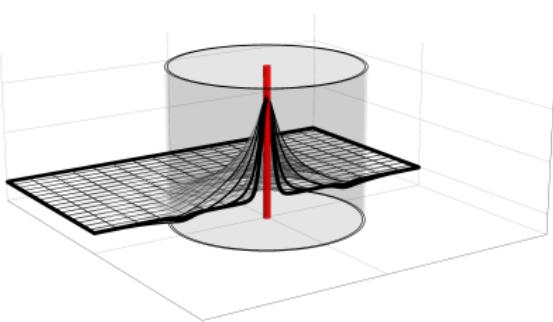
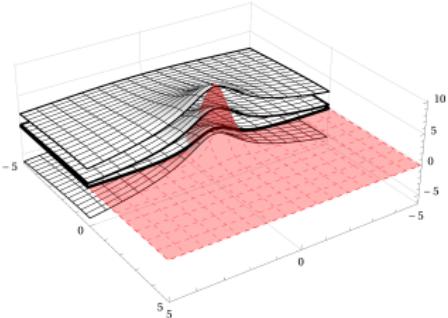
(Screened) Poisson Equation for **any BC** in vacuum

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

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



The BigDFT project

Wavelets

BigDFT compilation

Poisson Solver
Implicit Solvents

Perspectives
In Practice

Opportunities

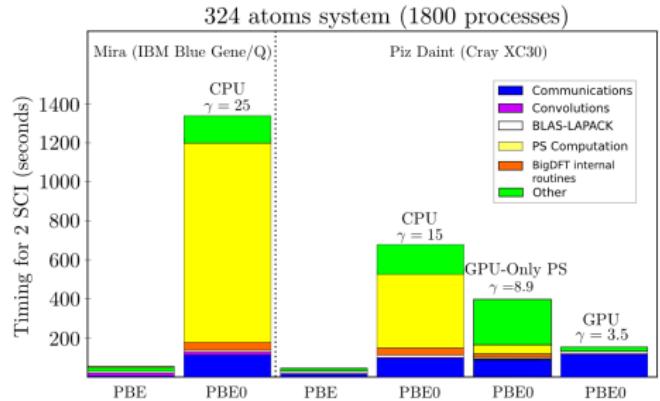
Hybrid Functionals

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



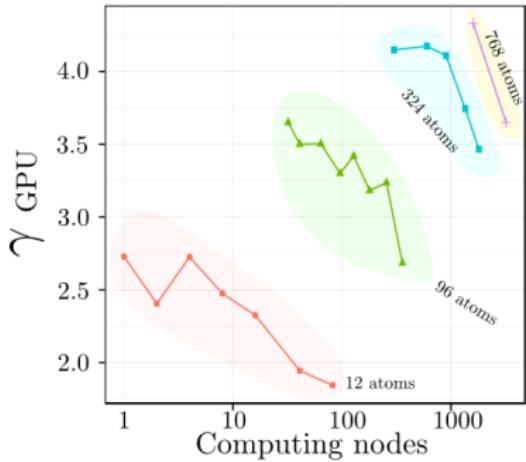
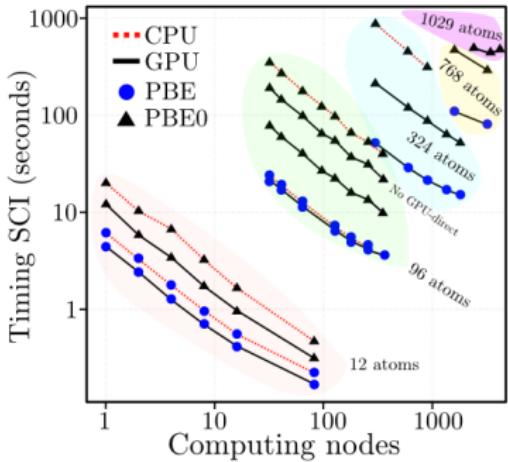
BigDFT

L. Genovese



UO₂ systems:

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



The BigDFT project

Wavelets

BigDFT compilation

Poisson Solver

Implicit Solvents

Perspectives

In Practice

Opportunities

Polarizable Continuum Models



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

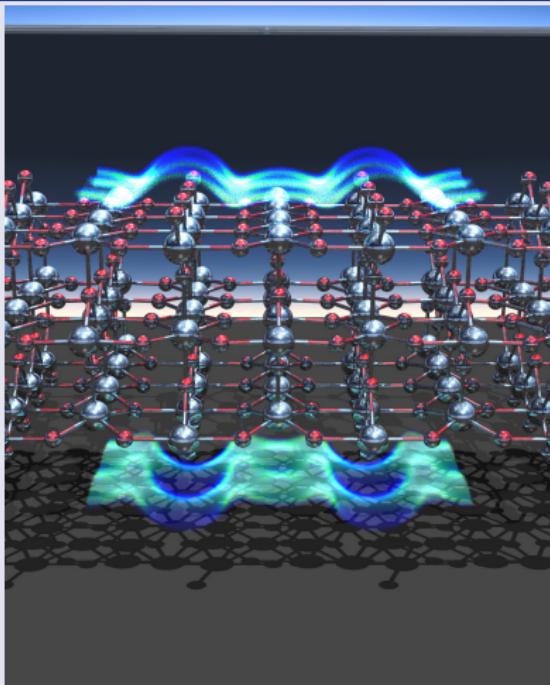
Allows an efficient and accurate treatment of implicit solvents

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)



Can be used in conjunction with $O(N)$ BigDFT

BigDFT

L. Genovese

The BigDFT project

Wavelets

BigDFT compilation

Poisson Solver

Implicit Solvents

Perspectives

In Practice

Opportunities

Performances in full SCF runs

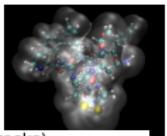
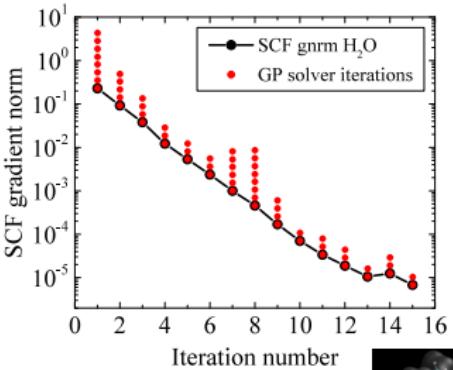


BigDFT

L. Genovese

Blackbox-like usage

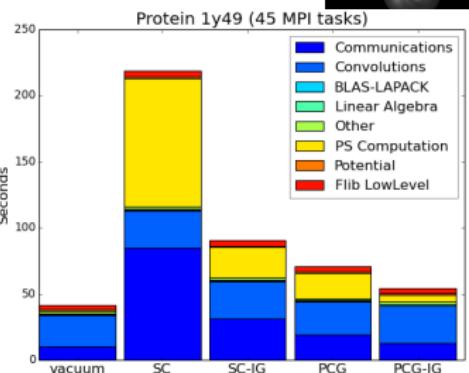
The Generalized PS only needs few iterations of the vacuum poisson solver



Time-to-solution

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

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



The BigDFT project

Wavelets

BigDFT compilation

Poisson Solver

Implicit Solvents

Perspectives

In Practice

Opportunities

Use locality of the basis set

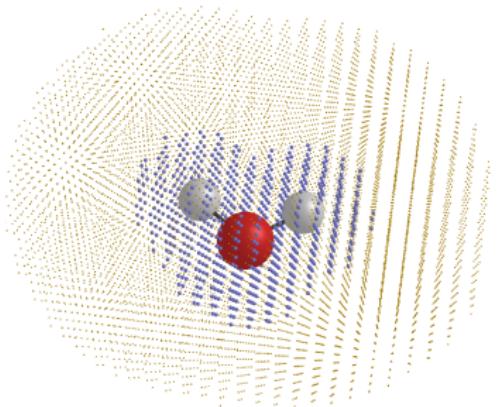


BigDFT

L. Genovese

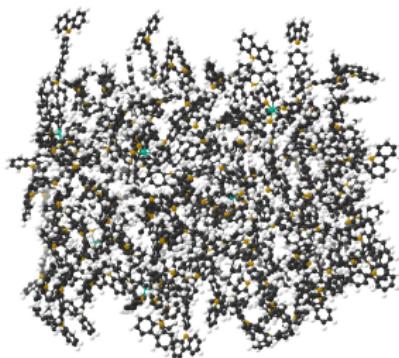
Wavelets

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!

The BigDFT project

Wavelets

BigDFT compilation

Poisson Solver

Implicit Solvents

Perspectives

In Practice

Opportunities

From Cubic Scaling to Multiscale

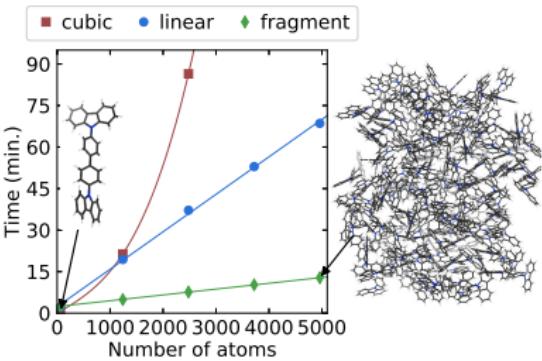
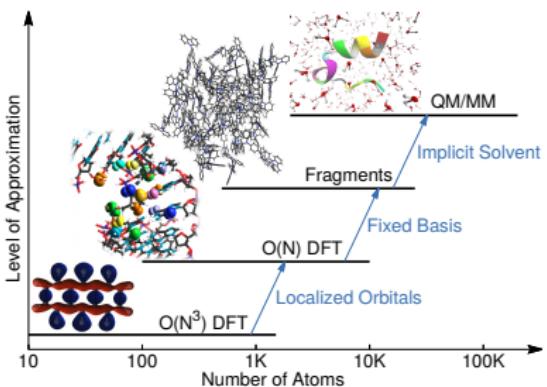


BigDFT

L. Genovese

Across Lengthscales

- extended orbitals → $O(N^3)$
- exploit locality → $O(N)$
- exploit repetition → ↓ cost $O(N)$
- larger systems → increasing complexity
- how to treat **complex systems?**



Features of the approach

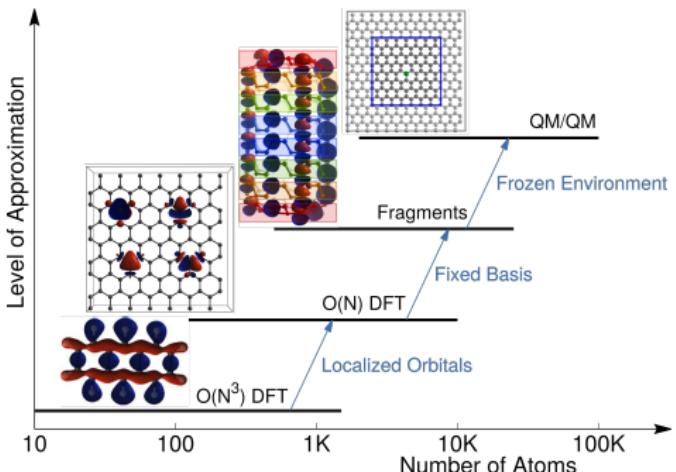


BigDFT

L. Genovese

Across Lengthscales with Wavelets

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



Received: 28 April 2021 | Revised: 31 August 2021 | Accepted: 1 September 2021

DOI: 10.1002/wcms.1574

ADVANCED REVIEW



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

William Dawson¹ | Augustin Degomme² | Martina Stella³ |
Takahito Nakajima¹ | Laura E. Ratcliff³ | Luigi Genovese² 

¹RIKEN Center for Computational Science, Kobe, Japan

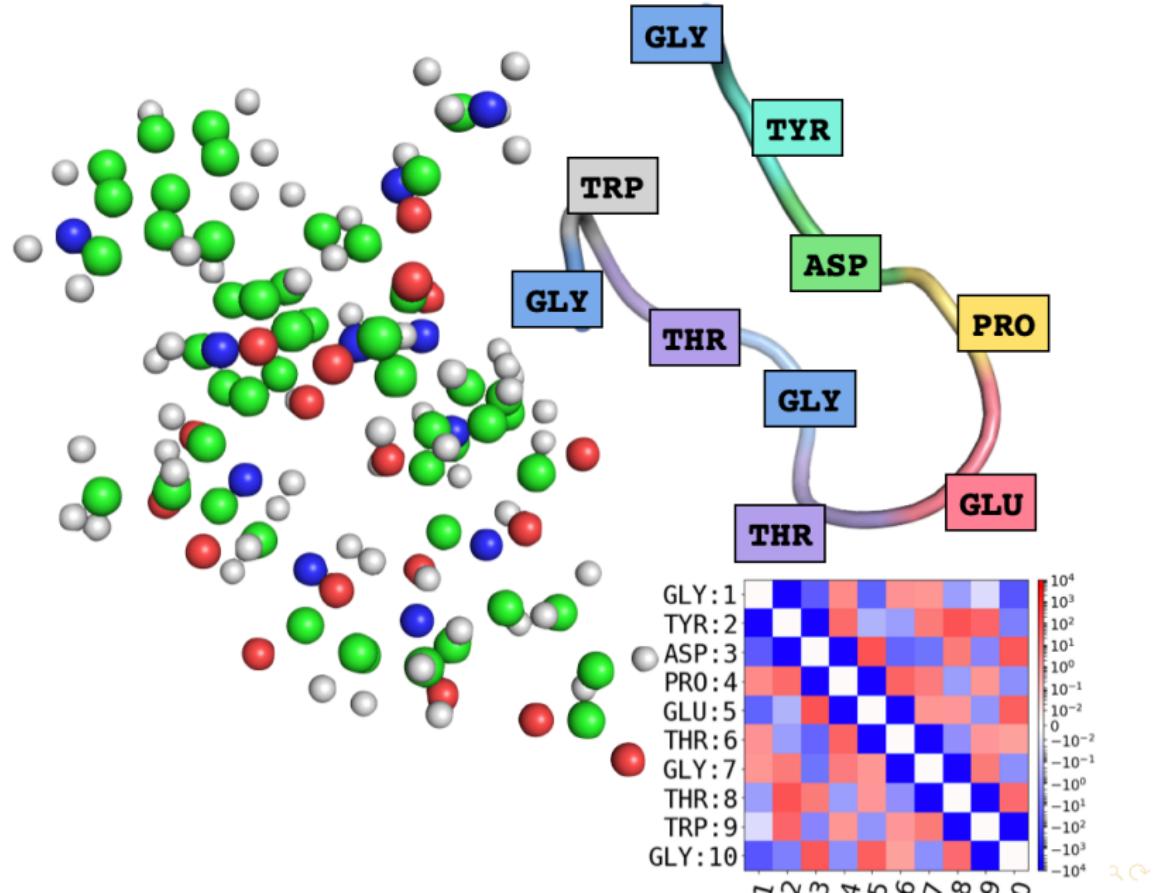
²Université Grenoble Alpes, INAC-MEM, L_Sim, Grenoble, France

³Department 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



BigDFT

L. Genovese

The BigDFT
project

Wavelets

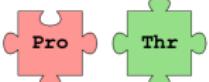
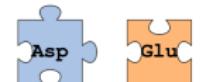
BigDFT
compilation

Poisson Solver
Implicit Solvents

Perspectives
In Practice

Opportunities

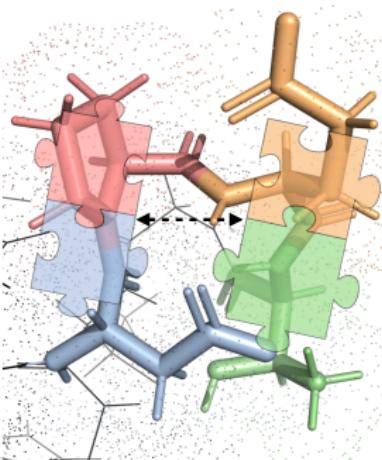
Puzzled(?) By Large Scale DFT



Traditional Description



QM Informed Building Blocks



Large Uncertain Structure → Reduction To Core Pieces → Fitting The Picture Together

Automatic Fragmentation of Systems

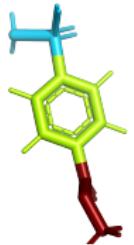


BigDFT

L. Genovese

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.

931	932	T	933	R	934	K	935	H	936	A	937	P	938	S	939	G	940	E	941	C
937	938	I	939	S	940	L	941	V	942	D	943	Y	944	F	945	W	946	M	947	N
948	949	Q	950	N	951	C	952	P	953	Q	954	C	955	P	956	R	957	L	958	Y
959	960	W	961	Q	962	A	963	N	964	N	965	P	966	S	967	N	968	N	969	Q
970	971	Y	972	V	973	I	974	H	975	R	976	T	977	S	978	I	979	H	980	R
981	982	E	983	Y	984	Y	985	Y	986	Y	987	Y	988	Y	989	Y	990	Y	991	Y
992	993	Y	994	Y	995	Y	996	Y	997	Y	998	Y	999	Y	990	Y	991	Y	992	Y
993	994	Y	995	Y	996	Y	997	Y	998	Y	999	Y	990	Y	991	Y	992	Y	993	Y
994	995	Y	996	Y	997	Y	998	Y	999	Y	990	Y	991	Y	992	Y	993	Y	994	Y
995	996	Y	997	Y	998	Y	999	Y	990	Y	991	Y	992	Y	993	Y	994	Y	995	Y
996	997	Y	998	Y	999	Y	990	Y	991	Y	992	Y	993	Y	994	Y	995	Y	996	Y
997	998	Y	999	Y	990	Y	991	Y	992	Y	993	Y	994	Y	995	Y	996	Y	997	Y
998	999	Y	990	Y	991	Y	992	Y	993	Y	994	Y	995	Y	996	Y	997	Y	998	Y
999	990	Y	991	Y	992	Y	993	Y	994	Y	995	Y	996	Y	997	Y	998	Y	999	Y
990	991	Y	992	Y	993	Y	994	Y	995	Y	996	Y	997	Y	998	Y	999	Y	990	Y
991	992	Y	993	Y	994	Y	995	Y	996	Y	997	Y	998	Y	999	Y	990	Y	991	Y
992	993	Y	994	Y	995	Y	996	Y	997	Y	998	Y	999	Y	990	Y	991	Y	992	Y
993	994	Y	995	Y	996	Y	997	Y	998	Y	999	Y	990	Y	991	Y	992	Y	993	Y
994	995	Y	996	Y	997	Y	998	Y	999	Y	990	Y	991	Y	992	Y	993	Y	994	Y
995	996	Y	997	Y	998	Y	999	Y	990	Y	991	Y	992	Y	993	Y	994	Y	995	Y
996	997	Y	998	Y	999	Y	990	Y	991	Y	992	Y	993	Y	994	Y	995	Y	996	Y
997	998	Y	999	Y	990	Y	991	Y	992	Y	993	Y	994	Y	995	Y	996	Y	997	Y
998	999	Y	990	Y	991	Y	992	Y	993	Y	994	Y	995	Y	996	Y	997	Y	998	Y
999	990	Y	991	Y	992	Y	993	Y	994	Y	995	Y	996	Y	997	Y	998	Y	999	Y

The BigDFT project
WaveletsBigDFT compilation
Implicit SolventsPerspectives
In Practice

Opportunities

Lots of Systems of interest in Biology



BigDFT

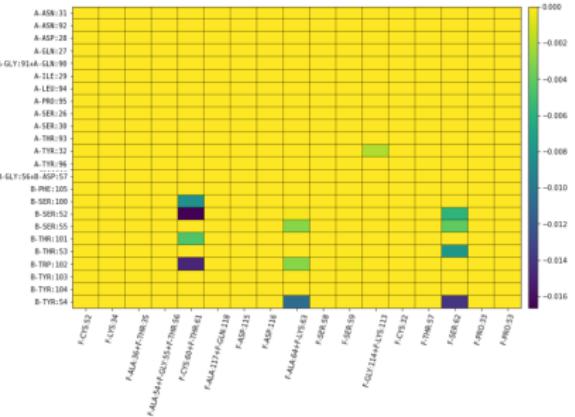
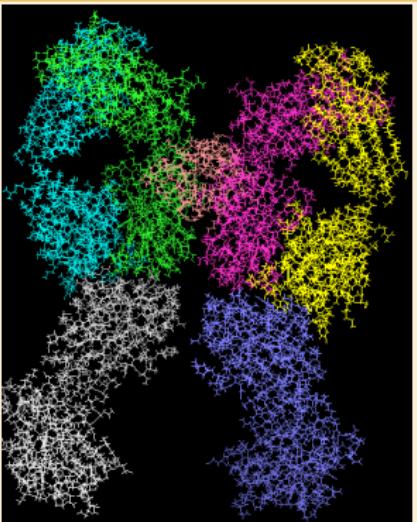
L. Genovese

Large systems are routinely accessible

Linear Scaling code → 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

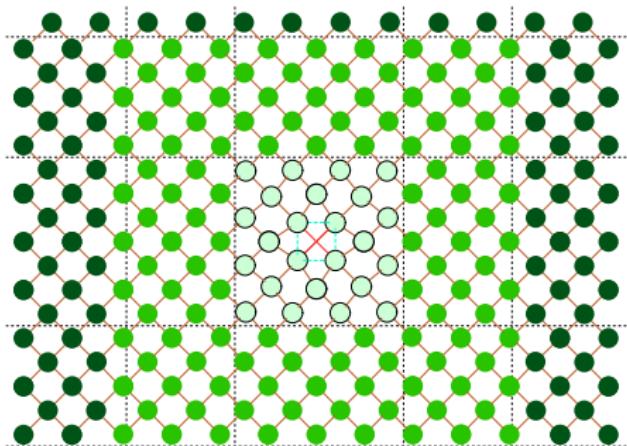


Perspectives (QM/QM)



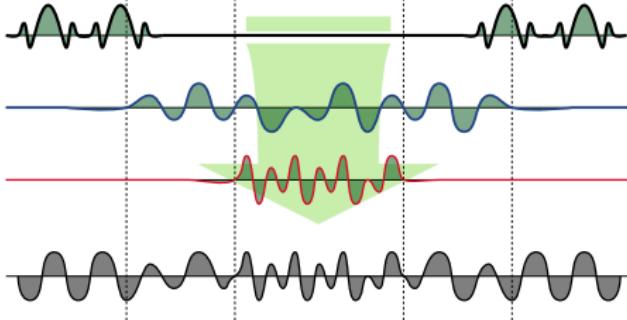
BigDFT

L. Genovese



Region descr.

- active site
- full ab initio
- constrained-DFT approach
- tight-binding like environment



Wavefunc. descr.

- bulk-like frozen wavefunctions
- optimized wavefunc. on a frozen basis set
- optimized wavefunc. and basis set
- complete representation of the active site in bulk environment

The BigDFT
project

Wavelets

BigDFT
compilation

Poisson Solver
Implicit Solvents

Perspectives

In Practice

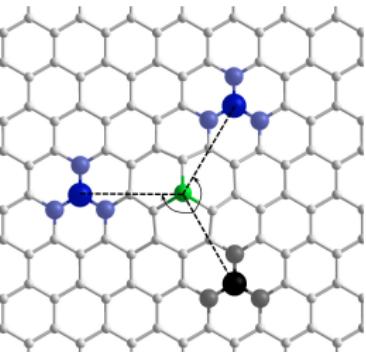
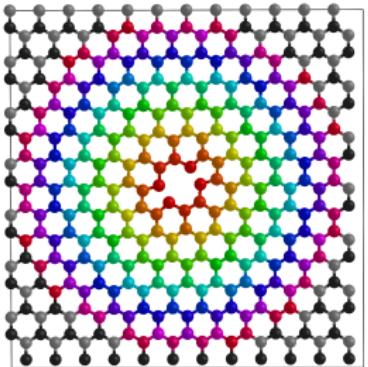
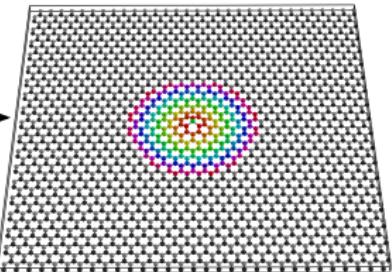
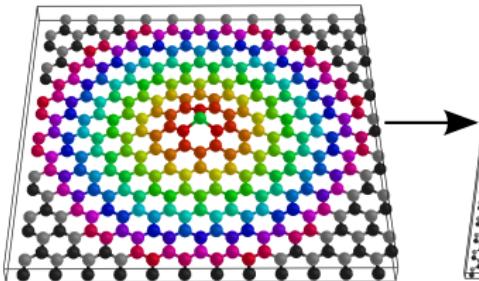
Opportunities

Defective Graphene with Fragments



BigDFT

L. Genovese



close to the defect:
strong perturbation

far from the defect:
bulk-like behaviour

The BigDFT
project

Wavelets

BigDFT
compilation

Poisson Solver
Implicit Solvents

Perspectives

In Practice

Opportunities

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-specialists.
- This combination create interesting opportunities for interdisciplinary collaborations.

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, Commitment ➔ The right guys

Acknowledgments

Bristol L. E. Ratcliff

Trieste M. Stella

Kobe W. Dawson, T. Nakajima

Boston B. Momeni

Grenoble L. Beal, S. Dechamps, D. Caliste, M. Zaccaria,
D. Rolland, LG

ILL V. Cristiglio

SiPearl, Intel A. Degomme, C. Bauinger

Plymouth M. Gisborne