



Big DFT  
Challenges

Introduction

Why wavelets

$O(N)$   
BigDFT

Why  $O(N)$

Fragment  
Identification

Purity Indicator

Example

Outlook

Seminar

CENTRE FOR PREDICTIVE MODELLING, WARWICK

*Opportunities from Accurate and Efficient  
Density Functional Theory Calculations for Large  
Systems*

Luigi Genovese

L\_Sim – CEA Grenoble

October 30, 2017



Big DFT  
Challenges

Introduction

Why wavelets

$O(N)$   
BigDFT

Why  $O(N)$

Fragment  
Identification

Purity Indicator

Example

Outlook

## 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**

**A code used in production since 2008**

A new formalism: Opens the path towards new opportunities



Big DFT  
Challenges

Introduction

Why wavelets

$O(N)$   
BigDFT

Why  $O(N)$

Fragment  
Identification

Purity Indicator

Example

Outlook

## “Traditional” BigDFT code

We can reach systems containing up to **a few hundred electrons** thanks to wavelet properties and efficient **parallelization**: (MPI + OpenMP + GPU)

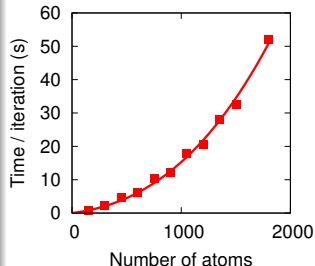
## Varying the number of atoms $N$

DFT operations **scale** differently:

- $O(N \log N)$ : Poisson solver
- $O(N^2)$ : convolutions
- $O(N^3)$ : linear algebra

and have different **prefactors**:

- $C_{O(N^3)} \ll C_{O(N^2)} \ll C_{O(N \log N)}$



For bigger systems the  $O(N^3)$  will dominate

☞ (first) motivation for a **new approach**

# Local orbitals and linear scaling

## KS orbitals

Linear combinations of **support functions**  $\phi_\alpha(\mathbf{r})$ :

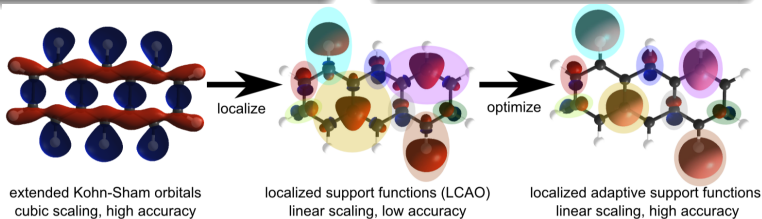
$$\Psi_i(\mathbf{r}) = \sum_{\alpha} c_i^{\alpha} \phi_{\alpha}(\mathbf{r})$$

- localized around atoms
- expanded in wavelets
- **optimized in-situ**

## Density Matrix

Defined via the **kernel**  $K^{\alpha\beta}$  in the  $\phi_\alpha(\mathbf{r})$  basis:

$$\begin{aligned} \rho(\mathbf{r}, \mathbf{r}') &= \sum_i f_i \Psi_i(\mathbf{r}) \Psi_i(\mathbf{r}') \\ &= \sum_{\alpha, \beta} \phi_{\alpha}(\mathbf{r}) K^{\alpha\beta} \phi_{\beta}(\mathbf{r}') \end{aligned}$$



**Localized form for the Density and the Hamiltonian**

Localization  $\rightarrow$  sparse matrices  $\rightarrow O(N)$



Big DFT  
Challenges

Introduction  
Why wavelets

$O(N)$   
BigDFT

Why  $O(N)$

Fragment  
Identification  
Purity Indicator  
Example

Outlook

# Comparison with the cubic version



Big DFT Challenges

Introduction  
Why wavelets

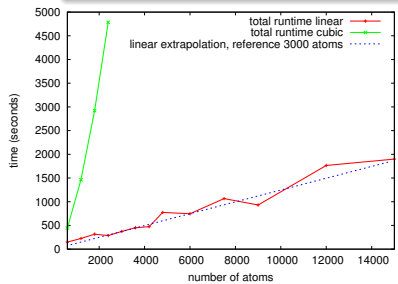
$O(N)$   
BigDFT

Why  $O(N)$

Fragment  
Identification  
Purity Indicator  
Example

Outlook

## Energy and forces with accuracy of a systematic approach



- 20 min for 18 000 atoms
- CPU Time and memory  $\propto$  number of atoms
- Precise DFT computing for thousands atoms **at Institute-Scale** ( $10^2 - 10^3$  CPU cores)

Different levels of precision cutoff radii  
Without fine-tuning converges to **absolute** energy differences of the order of 10 meV/atom, and almost **exact** forces.

High flexibility, like the cubic code

- Charged systems, various BC (free, surfaces, periodic)
- System sizes: 100 - 30K atoms  $\rightsquigarrow$  100 k Basis functions

# Features of the basis set

## Ideal properties for DFT at many thousand atoms scale

- Accurate results with good localization (high sparsity)
- Low No. of degrees of freedom
- ☞ (very!) Low condition number (quasi-orthogonality)
- ☞ Small Spectral Width (PSP)



Big DFT  
Challenges

Introduction

Why wavelets

$O(N)$   
BigDFT

Why  $O(N)$

Fragment  
Identification

Purity Indicator

Example

Outlook

		S		H	
system	(#atoms)	sparsity	$\kappa$	sparsity	SW (AU)
DNA	(15613)	99.57%	2.29	98.46%	1.81
bulk pentacene	(6876)	98.96%	2.26	97.11%	1.55
perovskite	(768)	90.34%	2.15	76.47%	1.73
Si nanowire	(706)	93.24%	2.16	81.61%	1.53
H <sub>2</sub> O droplet	(1800)	96.71%	1.57	90.06%	1.41

# Reference paper for $O(N)$ BigDFT implementation

Algorithm is robust and reliable on a variety of systems



Big DFT  
Challenges

Introduction

Why wavelets

$O(N)$   
BigDFT

Why  $O(N)$

Fragment  
Identification

Purity Indicator

Example

Outlook

**Accurate and efficient linear scaling DFT calculations with universal applicability**  
S. Mohr, L. E. Ratcliff, L. Genovese, D. Caliste, P. Boulanger, S. Goedecker and T. Deutsch  
*Phys. Chem. Chem. Phys.*, 2015, 17, 47, 31360-31370.  
DOI: 10.1039/c5cp00437c

Volume 17 Number 47 22 December 2015 Pages 3135-3208

**PCCP**  
Physical Chemistry Chemical Physics

Themed issue: Real-space numerical grid methods in quantum chemistry

ROYAL SOCIETY OF CHEMISTRY  
11th Edition of the International Union of Pure and Applied Chemistry (IUPAC) nomenclature of inorganic chemistry (2016)

Included in the **Real-space numerical grid methods in quantum chemistry** themed issue of *PCCP*  
Guest-edited by Luca Frediani (The Arctic University of Norway) and Dage Sundholm (University of Helsinki)

# Why Large Scale DFT?

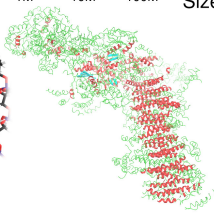
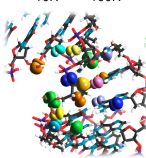
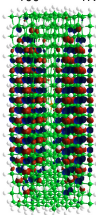
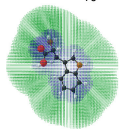
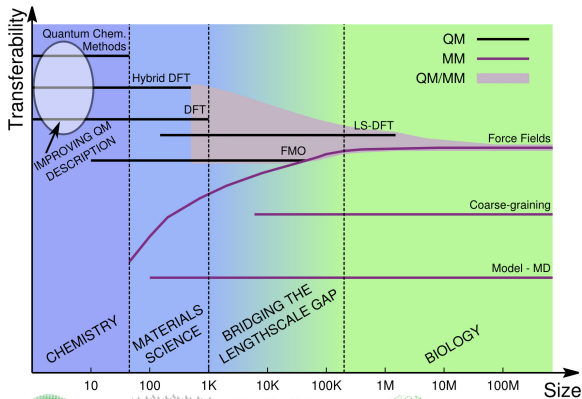
Present-day situation



Big DFT Challenges

Introduction  
Why wavelets  
 $O(N)$   
BigDFT  
Why  $O(N)$

Fragment Identification  
Purity Indicator  
Example  
Outlook







## Challenges in large scale quantum mechanical calculations

Laura E. Ratcliff,<sup>1</sup> Stephan Mohr,<sup>2</sup> Georg Huhs,<sup>2</sup> Thierry Deutsch,<sup>3,4</sup> Michel Masella<sup>5</sup> and Luigi Genovese<sup>3,4\*</sup>

During the past decades, quantum mechanical methods have undergone an amazing transition from pioneering investigations of experts into a wide range of practical applications, made by a vast community of researchers. First principles calculations of systems containing up to a few hundred atoms have become a standard in many branches of science. The sizes of the systems which can be simulated have increased even further during recent years, and quantum-mechanical calculations of systems up to many thousands of atoms are nowadays possible. This opens up new appealing possibilities, in particular for interdisciplinary work, bridging together communities of different needs and sensibilities. In this review we will present the current status of this topic, and will also give an outlook on the vast multitude of applications, challenges, and opportunities stimulated by electronic structure calculations, making this field an important working tool and bringing together researchers of many different domains. © 2016 John Wiley & Sons, Ltd

How to cite this article:

*WIREs Comput Mol Sci* 2016. doi: 10.1002/wcms.1290

New calculation paradigms are emerging



Big DFT  
Challenges

Introduction

Why wavelets

$O(N)$   
BigDFT

Why  $O(N)$

Fragment  
Identification

Purity Indicator

Example

Outlook

# Express localized information

Support Functions describe “exactly” KS Density Matrix

$\phi_\alpha$  can be used to **map** quantum information to localized DoF

Extract reliable localized quantities

☛  $O(N)$  BigDFT provides also **ideal** set up to condense information coming from large-scale (many thousands atoms) QM calculations.

- Partial Density of States  
Each atom can be associated a subset of basis functions
- Hamiltonian and density matrices  
Direct consequence of nearsightedness

BigDFT information can be used for

- Express efficiently and accurately the QM results (DFT)
- Define a *hierarchy* of various levels of theory



Big DFT  
Challenges

Introduction

Why wavelets

$O(N)$   
BigDFT

Why  $O(N)$

Fragment  
Identification

Purity Indicator

Example

Outlook

# Nearsightedness implications in Simulation setups

Localized basis set helps in understanding systems' behaviour

## Locality of the density matrix

- Identify fundamental building blocks of the system (Fragments, residues)
- Understand and model the interaction between them (electrostatic embedding)
- Manipulate fragment quantities to extract excited-state properties (Constrained-DFT)

## Actions needed for such investigations

- Build/Optimize **basis functions** for different systems
- Inspect the **Density Matrix** in this basis
- Manipulate the **Hamiltonian** matrix elements



Big DFT  
Challenges

Introduction

Why wavelets

$O(N)$   
BigDFT

Why  $O(N)$

Fragment  
Identification

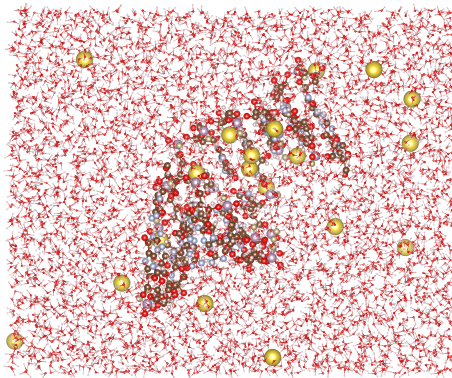
Purity Indicator

Example

Outlook

11 base pairs + Na / H<sub>2</sub>O solution: 15613 atoms

- Complete DFT calculation: 2h15m (800 MPI, 8 OMP)
- Let us perform a *atomic* Mulliken Population analysis
- Different questions arise



Big DFT  
Challenges

Introduction

Why wavelets

$O(N)$

BigDFT

Why  $O(N)$

Fragment  
Identification

Purity Indicator

Example

Outlook

# Arbitrariness of the results



Big DFT  
Challenges

Introduction

Why wavelets

$O(N)$

BigDFT

Why  $O(N)$

Fragment

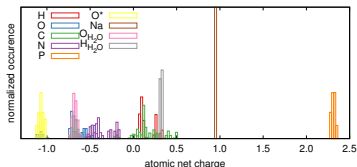
Identification

Purity Indicator

Example

Outlook

- Some atomic charges seem sound (e.g. Na), other look strange (e.g. P)... why?
- If we use more complete basis results change *considerably* (known problem of MPA and LPA)
- However, we *know* that the basis functions represent (very) well the Ground State



## Definition of a meaningful quantity

Atomic charges are (in general) *not* physical observables.  
Is it possible to identify fragment-related  
*pseudo-observables*?      🗑️ “meaningful” fragmentation

# Fragmentation procedure: the **purity indicator**

Suppose that a QM system, identified by  $\hat{F} \equiv |\psi\rangle\langle\psi|$  can be split into  $M$  “perfect” fragments  $\mathfrak{F}$ .

We indicate with  $|\psi^{\mathfrak{F}}\rangle$  the Fragment’s physical state.

☛ There must exist a **Fragment projection** operator  $\hat{W}^{\mathfrak{F}}$  s.t.  $\hat{W}^{\mathfrak{F}}|\psi\rangle = |\psi^{\mathfrak{F}}\rangle$ .

If the fragments are independent, i.e.  $\langle\psi^{\mathfrak{F}}|\psi^{\mathfrak{F}'}\rangle = \delta_{\mathfrak{F}\mathfrak{F}'}$ , the **Fragment density matrix** satisfies

$$\hat{F}^{\mathfrak{F}} \equiv \hat{F}\hat{W}^{\mathfrak{F}} = |\psi^{\mathfrak{F}}\rangle\langle\psi^{\mathfrak{F}}| = \left(\hat{F}^{\mathfrak{F}}\right)^2$$

When expressing this quantities in a localized basis:

$$\hat{F} = \sum_{\alpha\beta} |\phi_{\alpha}\rangle K^{\alpha\beta} \langle\phi_{\beta}|, \quad \hat{W}^{\mathfrak{F}} = \sum_{\alpha\beta} |\phi_{\alpha}\rangle R_{\mathfrak{F}}^{\alpha\beta} \langle\phi_{\beta}|$$

we obtain that a “good fragment” should satisfy

$$\Pi \equiv \frac{1}{Q_{\mathfrak{F}}} \text{tr} \left( (\mathbf{K}\mathbf{S}_{\mathfrak{F}})^2 - \mathbf{K}\mathbf{S}_{\mathfrak{F}} \right) \simeq 0, \quad \mathbf{S}_{\mathfrak{F}} \equiv \mathbf{S}\mathbf{R}_{\mathfrak{F}}\mathbf{S}$$



Big DFT Challenges

Introduction

Why wavelets

$O(N)$

BigDFT

Why  $O(N)$

Fragment

Identification

Purity Indicator

Example

Outlook

# Meaning of the Purity Indicator

If the above conditions are satisfied, the expectation value of any observable  $\hat{O}$  may also be associated to the fragment  $\mathfrak{F}$ :

$$\langle \hat{O} \rangle_{\mathfrak{F}} = \text{tr}(\hat{F}^{\mathfrak{F}} \hat{O}) = \text{tr}(\mathbf{KSR}_{\mathfrak{F}} \mathbf{O})$$

The PI is **not** an observable

- Explicit functional of  $\{|\phi_{\alpha}\rangle\}$  **and** of the *choice* of  $\mathbf{R}_{\mathfrak{F}}$
- $\Pi \simeq 0$  is a **necessary** condition. If it is not satisfied, the basis **and/or** the projection method do not single out  $\mathfrak{F}$  as a meaningful fragment.

☞ Interplay bw the basis set *and* the projection method

## Different Realizations

Pioneering atomic population analysis might be generalized to fragments. For instance:

Mulliken:  $\mathbf{R}_{\mathfrak{F}} = \mathbf{T}_{\mathfrak{F}} \mathbf{S}^{-1}$  Loewdin:  $\mathbf{R}_{\mathfrak{F}} = \mathbf{S}^{-1/2} \mathbf{T}_{\mathfrak{F}} \mathbf{S}^{-1/2}$   
where  $\mathbf{T}_{\mathfrak{F}}$  selects the indices  $\alpha \in \mathfrak{F}$ .



Big DFT  
Challenges

Introduction

Why wavelets

$O(N)$   
BigDFT

Why  $O(N)$

Fragment  
Identification

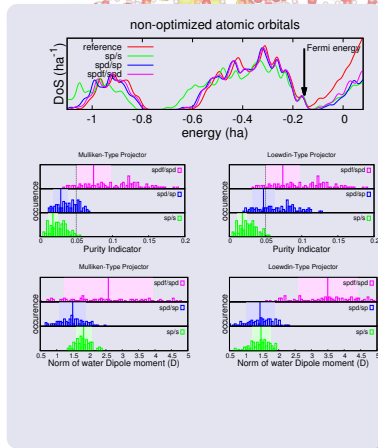
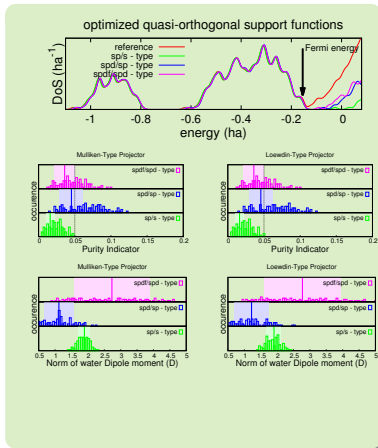
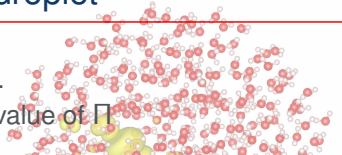
Purity Indicator

Example

Outlook

# Example, 100 Molecule Water “droplet”

Different basis (SF and AO) qualities.  
Correlation between H<sub>2</sub>O dipoles  $\leftrightarrow$  value of  $\Pi$



Big DFT  
Challenges

Introduction

Why wavelets

$O(N)$   
BigDFT

Why  $O(N)$

Fragment  
Identification

Purity Indicator

Example

Outlook



# Advantages of minimal bases

Atomic description might be replaced by molecular one

Molecules are good fragments **for minimal basis**

	sp/s optimized			sp/s atomic orbitals		
	H <sub>2</sub> O	O	H	H <sub>2</sub> O	O	H
Mulliken	0.02(1)	0.16(1)	0.45(0)	0.03(1)	0.16(1)	0.46(1)
Löwdin	0.03(1)	0.16(1)	0.45(0)	0.03(1)	0.17(1)	0.48(0)
quality	✓	✗	✗	✓	✗	✗

Meaningful fragmentation even with “simple” projections

	optimized			atomic orbitals		
	sp/s	spd/sp	spdf/spd	sp/s	spd/sp	spdf/spd
DoS	✓	✓	✓	✗	✓	✓
non-purity	✓	✗	✗	✓	✗	✗
H <sub>2</sub> O dipole	✓	✗	✗	✓/✗	✗	✗



Big DFT  
Challenges

Introduction  
Why wavelets

$O(N)$   
BigDFT  
Why  $O(N)$

Fragment  
Identification  
Purity Indicator  
Example

Outlook

# Coming back to the original case study



Big DFT Challenges

Introduction

Why wavelets

$O(N)$

BigDFT

Why  $O(N)$

Fragment

Identification

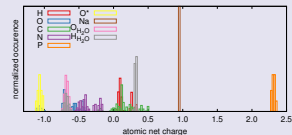
Purity Indicator

Example

Outlook

## Atomic description is useless with this projection (MPA)

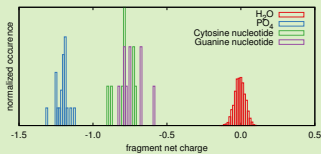
The only meaningful “atomic fragment” is the ionized Na of the solvent



	H	C	N	O	Os	Na	P
$\Pi$	0.48	0.48	0.32	0.15	0.12	0.04	0.34
	✗	✗	✗	✗	✗	✓	✗

## Fragments are - unsurprisingly - DNA residues

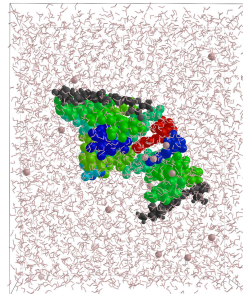
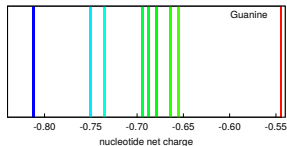
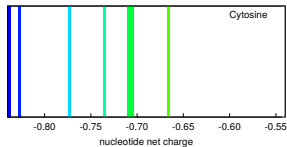
They can be identified **a posteriori** as the relevant quantities



	PO <sub>4</sub>	Cyt	Gua	H <sub>2</sub> O
$\Pi$	0.05	0.01	0.01	0.01
	✓	✓	✓	✓

## Interpreting QM properties of the solute with DFT

- Inspection of the Density Matrix allows to *identify* fragments (DNA residues in this case)
- Electrostatic multipoles on such fragments can be considered as observable quantities
- A general technique which takes benefit from a optimized minimal basis (JCCP 2017, 13, 4079-4088)



Big DFT  
Challenges

Introduction

Why wavelets

$O(N)$

BigDFT

Why  $O(N)$

Fragment

Identification

Purity Indicator

Example

Outlook

# Interplay between fragmentation and embedding

A good fragmentation may also be employed to study the influence of explicit vs. electrostatic solvation



Big DFT Challenges

Introduction

Why wavelets

$O(N)$   
BigDFT

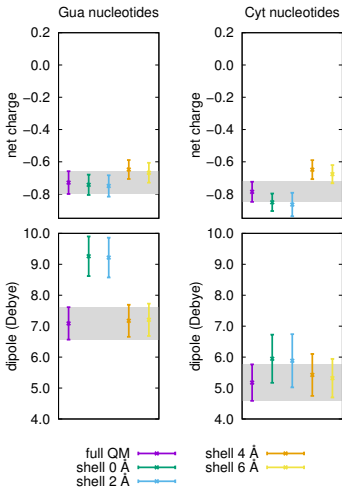
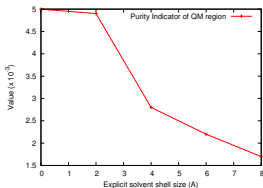
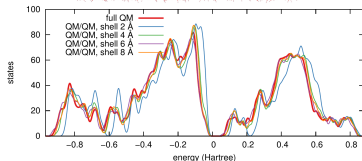
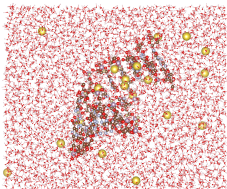
Why  $O(N)$

Fragment Identification

Purity Indicator

Example

Outlook



Reduce overpolarization

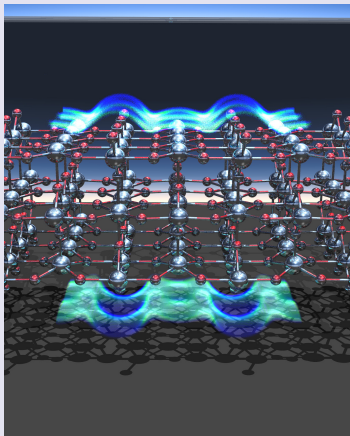
# Different modelling of the environment

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

Allows an efficient and accurate treatment of implicit solvents

The cavity can be

- rigid (PCM-like)
- determined from the Electronic Density (SCCS approach)
- Can treat various BC (here  $\text{TiO}_2$  surface)



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



Big DFT  
Challenges

Introduction

Why wavelets

$O(N)$   
BigDFT

Why  $O(N)$

Fragment  
Identification

Purity Indicator

Example

Outlook

# BigDFT basis set in Fragment approach

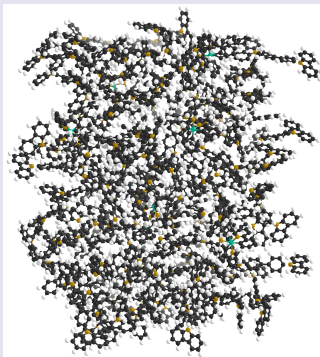
We can **duplicate** the Support Functions for similar portions of large systems → considerably reduces the cost

☛ Enables **manipulation** of optimized basis sets (database)

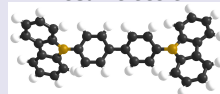
Impact of the environment in OLEDs charge transport

Realistic 'host-guest' morphology:

**6192 atoms, 100 molecules**

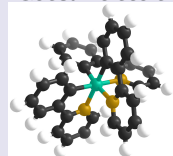


*Host molecule*



4,4'-N,N'-dicarbazole-biphenyl

*Guest molecule*



tris(2-phenylpyridine)iridium



Big DFT  
Challenges

Introduction

Why wavelets

$O(N)$   
BigDFT

Why  $O(N)$

Fragment  
Identification

Purity Indicator

Example

Outlook

# Future scientific directions with $O(N)$ codes



Big DFT  
Challenges

Introduction

Why wavelets

$O(N)$   
BigDFT

Why  $O(N)$

Fragment  
Identification

Purity Indicator

Example

Outlook

## Linear-Scaling DFT calculations based on wavelets

- **Robust** convergence, high **accuracy** and **flexibility** (BC)
- Reduction in degrees of freedom → **large systems** via **moderate sized machines** ( $\sim$  TFlop/s) Lab-scale
- **Optimal** mapping between KS DoF and atoms
- Different level of descriptions (**controlling the precision**)  
QM  $\supset$  Fragments  $\supset$  Atomic charges
- Opens up **new possibilities**

## Challenges and future directions

- Explore interplay environment  $\leftrightarrow$  electronic excitations (CDFT, QM/MM, statistics. . .)
- Provide high quality back end for extraction of atomic multipoles from QM calculations
- Towards a **control** of the level of theory (QM/QM)