

Simulating Thousands of Atoms using Linear Scaling BigDFT

Laura Ratcliff

School of Chemistry, University of Bristol
UiT The Arctic University of Norway

17.11.2022

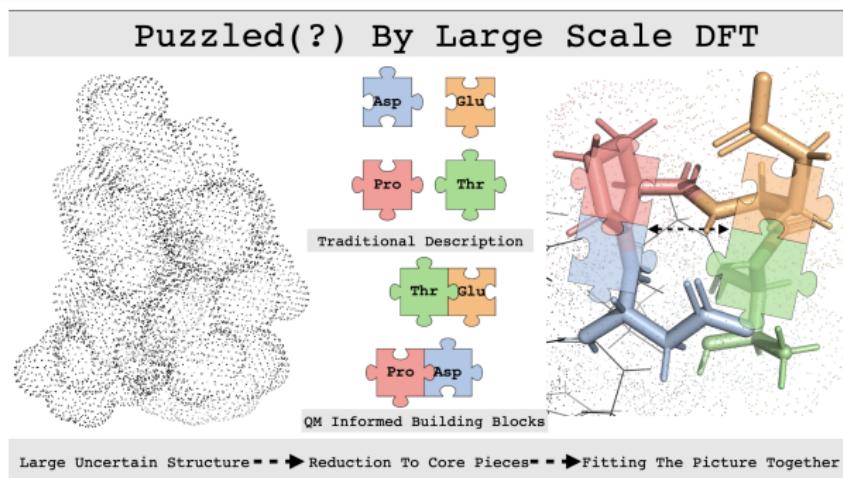


UiT The Arctic University of Norway

Why DFT with 1000s of Atoms?

Why do we Need QM for Large Systems?

- bridge lengthscale gap between QM and MM (validate empirical models, larger QM region etc)
- intrinsically QM quantities, e.g. electronic excitations
- new possibilities for simulating complex materials
- new regime for DFT → new ways of working



Large Scale QM Reviews: LER, Mohr, Huhs, Deutsch, Masella & Genovese WIREs Comput. Mol. Sci. 7, e1290 (2017); Dawson, Degomme, Stella, Nakajima, LER & Genovese WIREs Comput. Mol. Sci. 12, e1574 (2022)

Nearsightedness and Locality

Nearsightedness

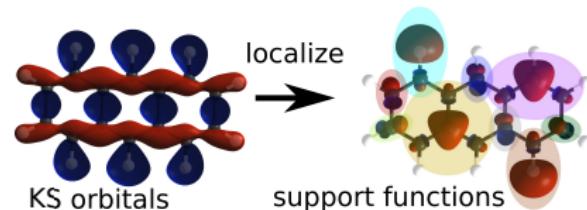
- wavelets + HPC → ~1000 atoms, but **cubic scaling**
- the behaviour of large systems is **short-ranged** (nearsighted)
- the density matrix, $\rho(\mathbf{r}, \mathbf{r}')$, decays exponentially in systems with a gap

→ how can we exploit nearsightedness to treat large systems?

Support Functions (SFs)

- write **extended KS orbitals** in terms of **localised SFs** ($\{\phi_\alpha(\mathbf{r})\}$):

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



Density Matrix Formulation

Avoiding Diagonalisation

- express $\rho(\mathbf{r}, \mathbf{r}')$ in terms of density kernel $K^{\alpha\beta}$ and SFs:

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

- no explicit reference to KS orbitals – avoid diagonalisation
- the density is found via:

$$n(\mathbf{r}) = \rho(\mathbf{r}, \mathbf{r})$$

Matrix Expressions

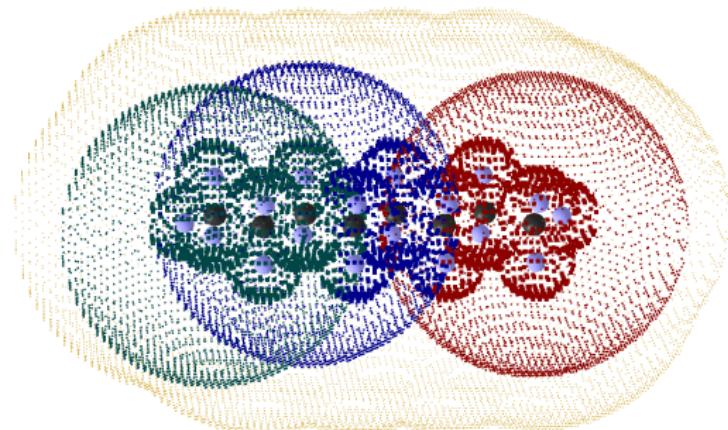
$$H_{\alpha\beta} = \langle\phi_\alpha|\hat{H}|\phi_\beta\rangle; \quad S_{\alpha\beta} = \langle\phi_\alpha|\phi_\beta\rangle$$

$$E = \text{Tr}(\mathbf{KH}); \quad N = \text{Tr}(\mathbf{KS})$$

Support Functions

SF Properties

- strictly **localised** ($\sim 6 - 8 a_0$ radius – systematic convergence)
- (usually) minimal (1 SF per H, 4 per C/N/O...)
- **atom-centred**
- quasi-orthogonal
- numerical functions **expanded in wavelets** → 2 levels of basis
- Γ -point only – real SFs



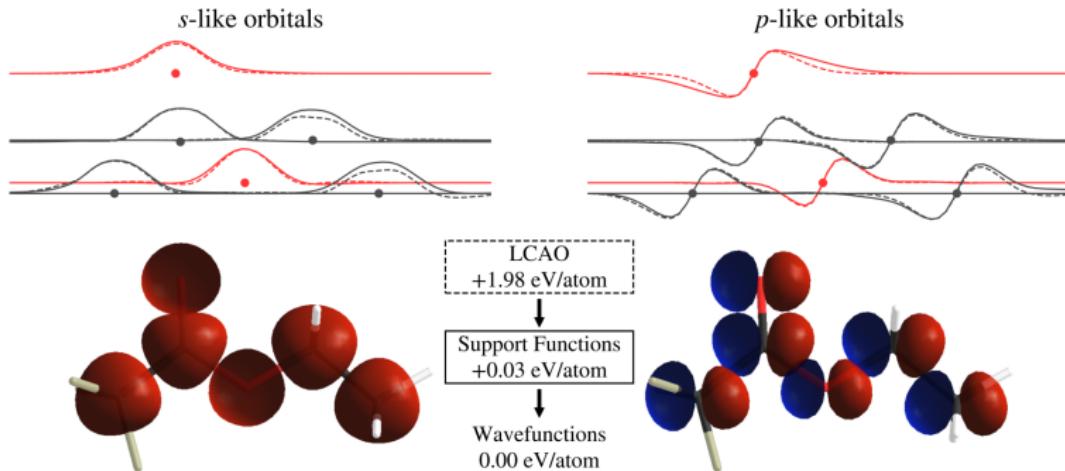
Support Function Optimisation

SF Optimisation

- start with atomic orbital guess for the SFs
- use a gradually decreasing confining potential:

$$\hat{H}_\alpha = \hat{H} + c_\alpha(\mathbf{r} - \mathbf{R}_\alpha)^4$$

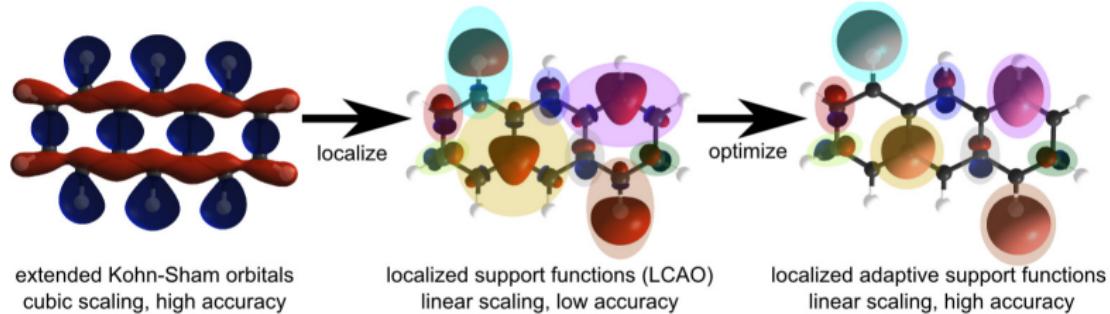
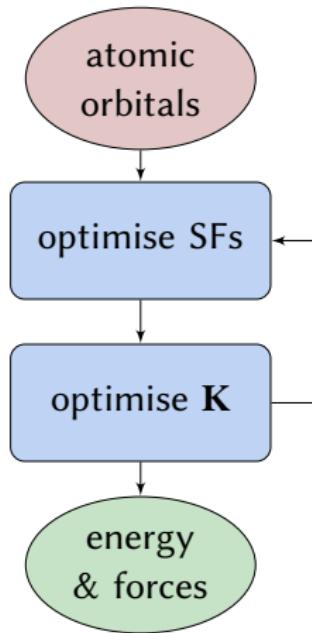
- optimise SFs *in situ* by minimising total energy
→ minimal, localised basis with high accuracy of wavelets



The Algorithm

Calculation Steps

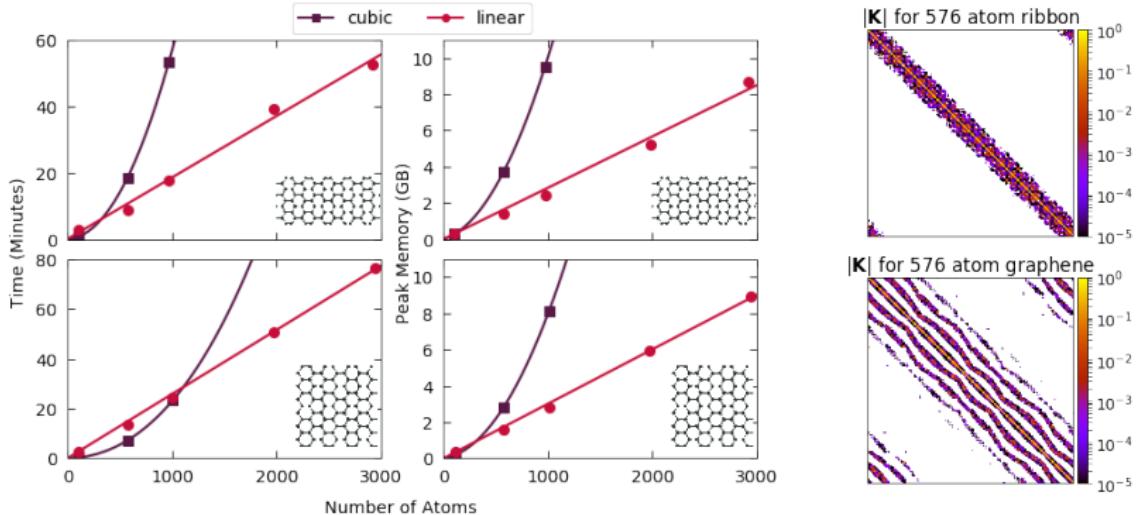
- minimise energy with respect to both SFs and \mathbf{K}
- choice of methods for \mathbf{K} : use FOE for linear scaling
- accurate forces – geometry optimisations, MD
- one-off diagonalisation for KS energies



From Sparsity to Linear Scaling

Sparse Matrices

- strict localization leads to sparse matrices (\mathbf{K} truncation)
- sparsity depends on size, dimensionality, SF radii \rightarrow crossover
- speed also depends on band gap (can treat metals)



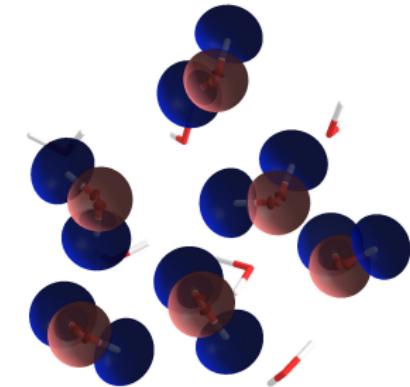
LS-BigDFT References: { Mohr, LER, Boulanger *et al.*, J. Chem. Phys. **140**, 204110 (2014)
Mohr, LER, Genovese *et al.*, Phys. Chem. Chem. Phys. **17**, 31360 (2015)

LS-BigDFT Metals: Mohr, Eixarch, Amsler *et al.*, J. Nucl. Mater. Energy **15**, 64 (2018)

Exploiting Similarity Between Fragments

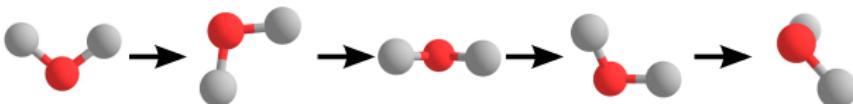
Calculation Bottleneck

- SF optimisation dominates prefactor
- similar chemical environments → similar SFs
- can we reuse SFs?

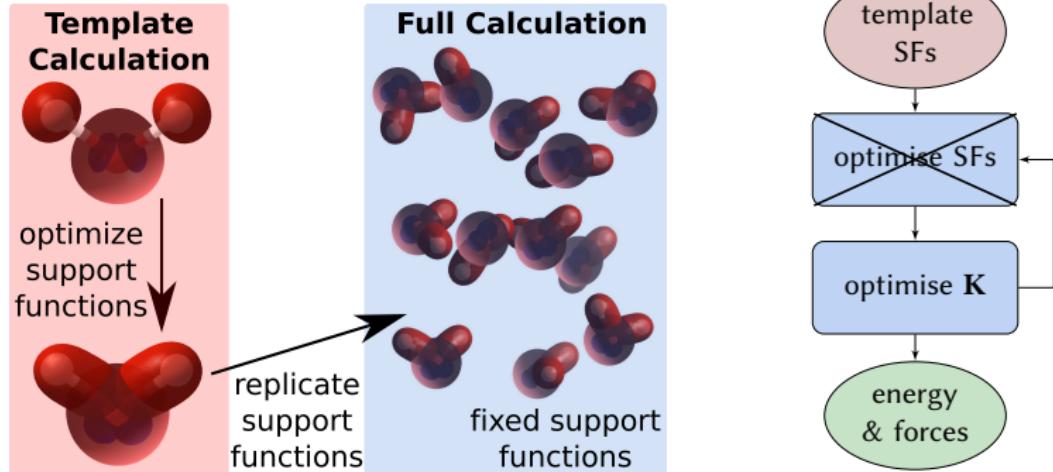


Accounting for Varying Orientations and Positions

- minimise cost function to find rotation from template:
- $$\mathcal{J}(\mathcal{R}) = \frac{1}{N} \sum_{a=1}^N \left\| \mathbf{R}_a^S - \sum_{b=1}^N \mathcal{R}_{ab} \mathbf{R}_a^T \right\|^2$$
- apply accurate and efficient wavelet interpolation scheme



Molecular Fragment Approach



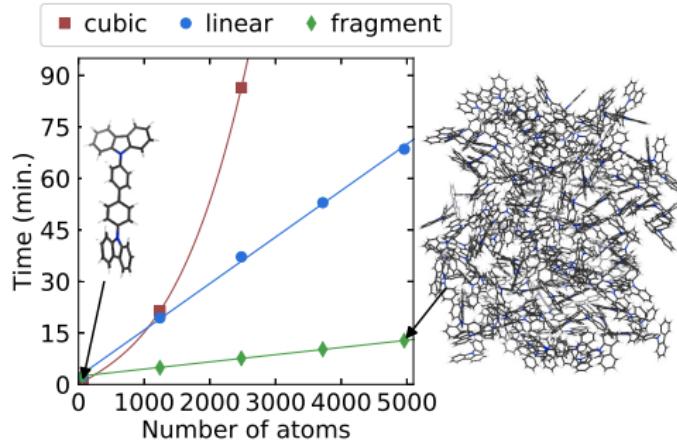
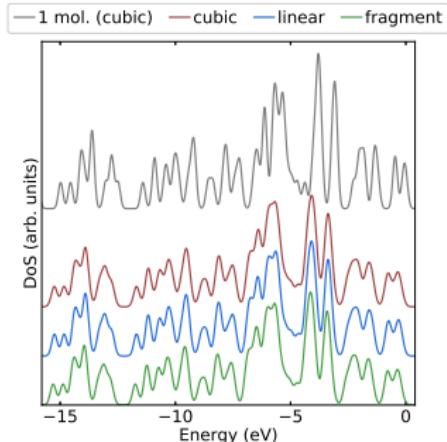
Calculation Steps

- (1) **template calculation:** optimise SFs for isolated fragment
- (2) **reformatting:** replicate and rototranslate template SFs for each fragment instance
- (3) **full calculation:** use fragment SFs as a fixed basis, optimising density kernel only

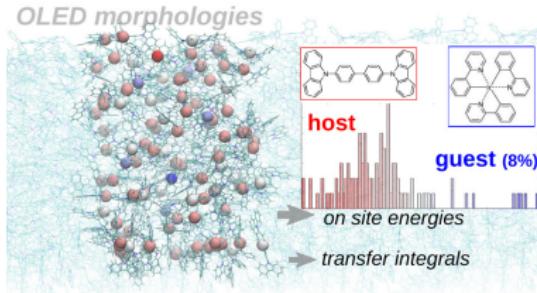
Fragment Approach and OLEDs

Cluster of Rigid CBP Molecules

- ~5000 atom single point calculation (48 nodes on Archer) – fragment approach $\sim 7\times$ cheaper than full linear scaling
- fragment approach reproduces (occupied) DoS
- $E_{\text{frag}} - E_{\text{cubic}} \simeq 30 \text{ meV/atom}$
- → weakly interacting fragments which are not too distorted (\mathcal{J})



OLED Charge Transport Parameters



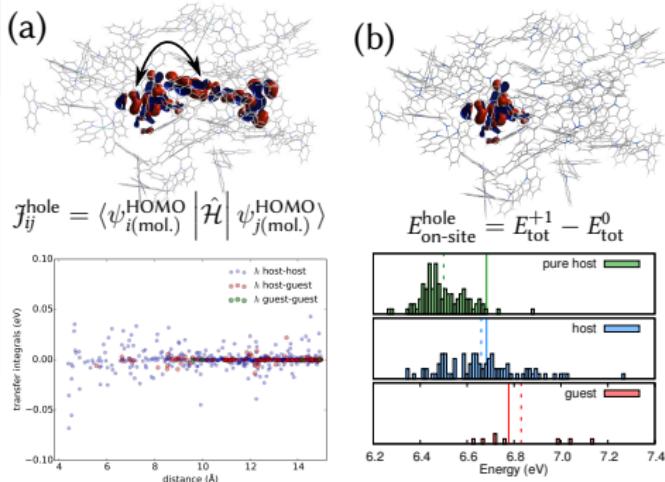
Fragment Setup

- isolated template calculations for each molecule
- calculate transfer integrals (a)
- calculate on-site energies using constrained DFT (b)

Environmental and Statistical Effects

- disorder \rightarrow dispersion in $E_{\text{on-site}}$ and J_{ij}
- environment \rightarrow shift in average $E_{\text{on-site}}$ (- -) vs. isolated molecules (-)

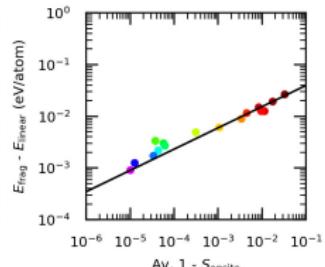
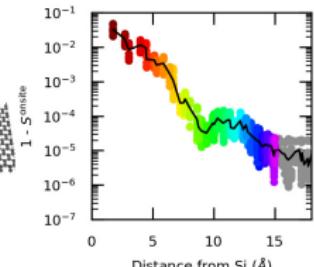
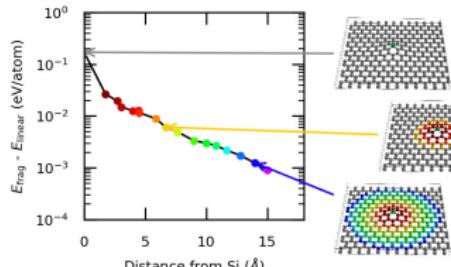
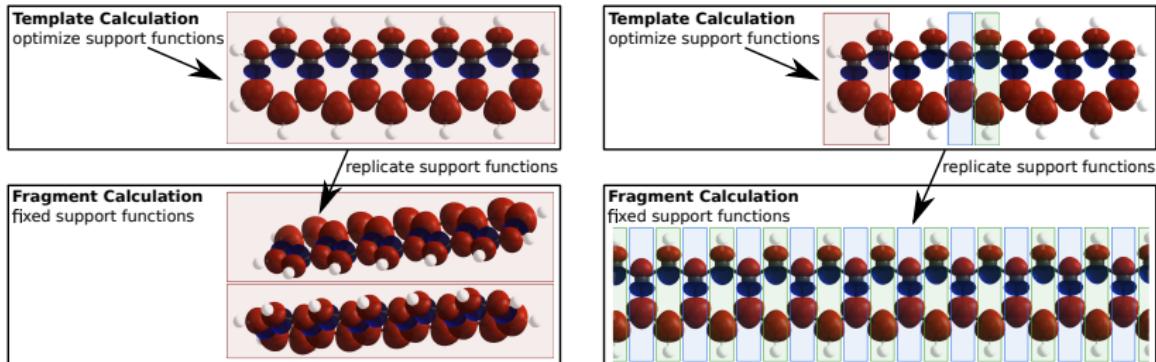
Host-Guest Application: LER, Grisanti, Genovese *et al.*, J. Chem. Theory Comput. **11**, 2077 (2015)



Fragment Approach: Beyond Molecules

Fragments in Extended Systems

- optimise SFs for **embedded** pseudo-fragments
- can define indicators to **predict the accuracy** of a given setup

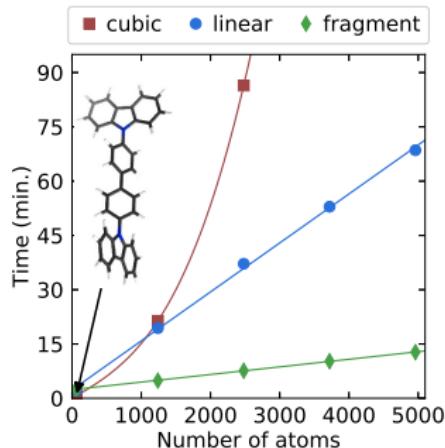
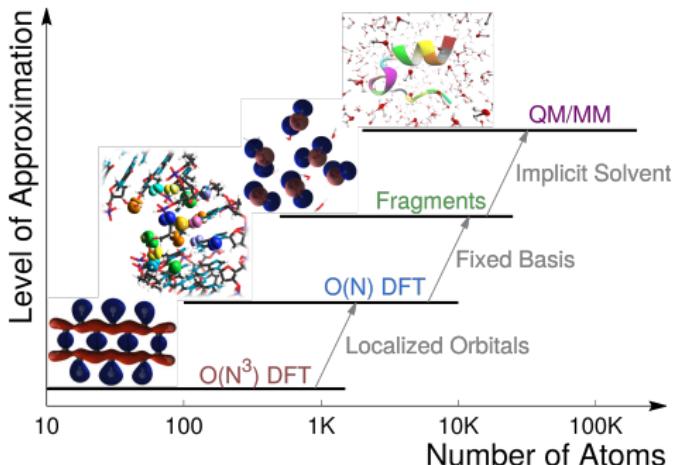


Embedded Pseudo-Fragments: LER and Genovese, J. Phys.: Condens. Matter 31, 285901 (2019)

Across Lengthscales with BigDFT

BigDFT: From Small to Large Systems

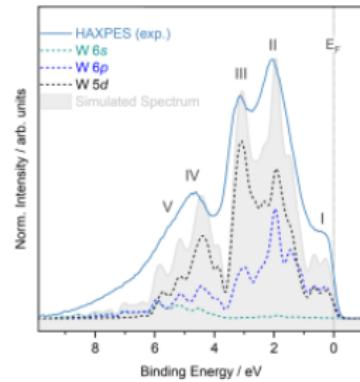
- cubic scaling → restricted to ~ 1000 atoms
- exploit locality *via* optimised SFs → linear scaling
- exploit repetition → fragment approach with lower prefactor
- implicit solvent – combine with cubic or linear scaling approaches
- → reduced cost, but what about analysis?



Exploiting SF Localisation

SFs – Beyond Reducing Cost

- SFs provide a map for **atom-based quantities** e.g. PDOS, atomic charges/multipoles...
- can we use SF locality to extract additional information?



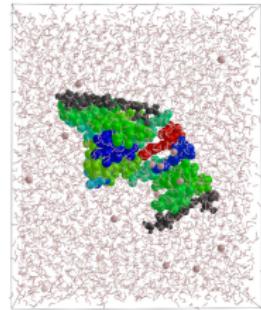
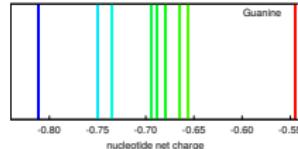
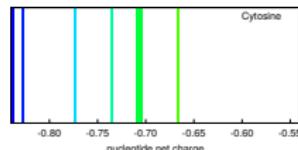
From Atoms to Fragments

- **density matrix** in localised basis → indicators for quantifying fragmentation choice and interactions in large systems
- purity indicator (II) – identify ('pure') **separable fragments**
- fragment bond order – quantify **fragment interactions**
- fragmentation depends on basis – need localised, minimal basis

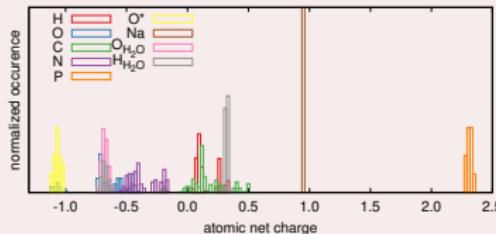
Example I: DNA Charge Analysis

Fragment Analysis

- 11 base pairs in Na/H₂O solution (15,613 atoms)
- full DFT calculation in 2h15m (6400 cores)
- II → unbiased identification of separable fragments

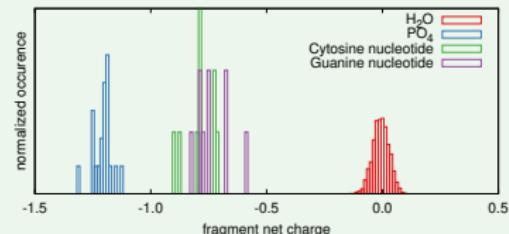


Atoms: Unphysical Charges



	H	C	N	O	O*	Na	P
II	0.48	0.48	0.32	0.15	0.12	0.04	0.34

Fragments: Chemically Sound

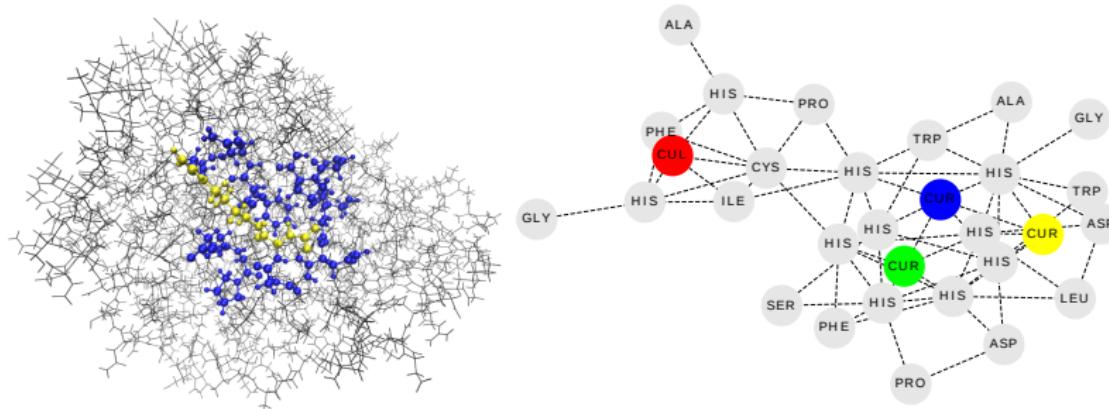


	PO ₄	Cyt	Gua	H ₂ O
II	0.05	0.01	0.01	0.01

Example II: Interactions in Laccase

Complexity Reduction of Laccase Enzyme (~7000 atoms)

- laccase – Cu atoms prevent purely amino acidic decomposition
- use **purity indicator** to identify fragments
- use **bond order** to define embedding region (blue)
- SF-based analysis → **coarse-grained view**
- could **automate** the setup of ‘active’ and ‘embedding’ regions for QM/MM



Example III: Covid Spike Protein

Beyond 50,000 Atoms

- use LS-BigDFT with implicit solvent
- ~52,000 atoms in ~6 hours on 128 nodes (16,384 cores)
- 3D graph view – nodes show residue interactions between the chains (coloured by strength)

