Max Webinar

BigDFT

Complexity Reduction: how BigDFT's basis set provides insights on electronic structure calculations of macromolecular systems

William Dawson

RIKEN Center for Computational Science

November 12, 2020

Virtual Room

Motivation

O(N) Codes In Practice

- O(N) codes have been able to compute large systems for many years now.
- And yet, how often do you encounter research being done with DFT involving systems of many tens of thousands of atoms?

The Issues

- Enthalpy challenge: are DFT functionals really more accurate than a well tuned forcefield?
- Entropy challenge: can DFT capture the full set of conformations of a large system?
- The key value added of DFT is not accuracy, but **insight**.

Density Matrix: A Review

Building the Density Matrix

- Let ψ_i be the set of Kohn-Sham orbitals expanded on some set of support functions $\psi_i = \sum_\alpha c_{\alpha i}\phi_\alpha$.
- We call H the Hamiltonian matrix. Then we can compute those orbitals by solving the eigenvalue problem $H\psi_i = \lambda_i\psi_i$ (ignore S for notational simplicity).
- Then we can construct the density matrix by summing over occupied orbitals $K_{\alpha\beta} = \sum_i^{Nocc} c_{\alpha i}c_{i\beta}$.

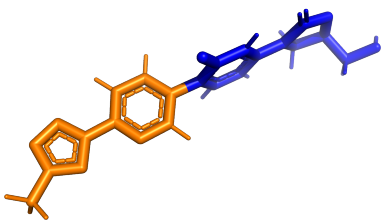
What is the Density Matrix?

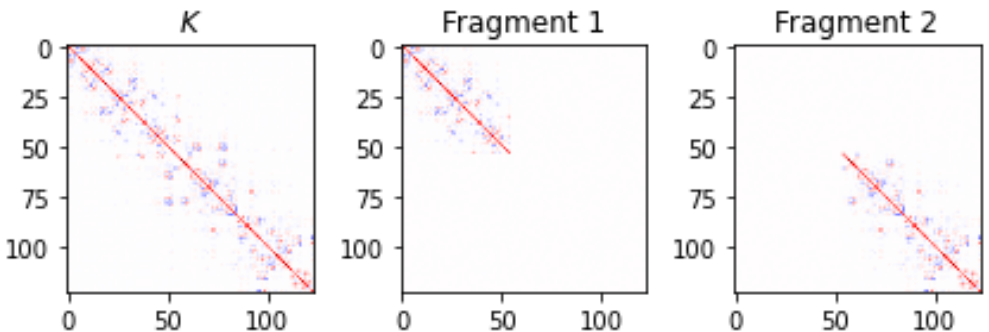
- Mathematically speaking, the density matrix is a projection on to the occupied orbitals.
- This means it fullfills two conditions: $Tr(K) = Nocc$ and it is *Idempotent* $K * K = K$.

Blocking of the Density Matrix

System Fragmentation

- Consider a system of interest, which we have arbitrarily partitioned in to two sets of atoms. We will call these sets *Fragments*.
- Because we are using atom centered support functions, we can partition the Density Matrix into four blocks, guided by the support functions.
- The top left and bottom right block can be considered projections on to each of the two fragments.
- These sub matrices no longer fullfill the idempotency and trace conditions.

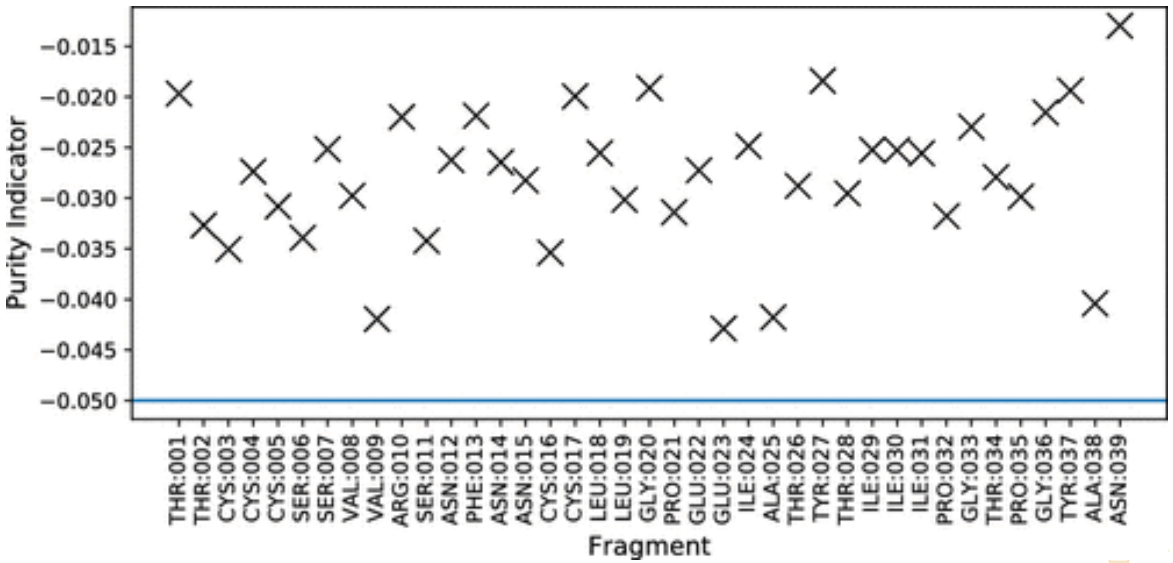




Purity Indicator

Fragment Purity

- We define the *Purity Indicator* as a normalized measure of how much a given projection deviates from the Idempotency condition.
- $\pi_F = \frac{1}{q_F} Tr(K^F K^F - K^F)$, where q is the number of electrons in a given fragment F.
- This error comes from our neglecting of the off-diagonal blocks of the matrix.



Purity Indicator and Bond Order

Block Matrix Analysis

- Using the definition of block matrix multiplication, we will discover terms B_{FE} and B_{EF} which describe the off diagonal contribution.
- We define B as the *Fragment Bond Order*.

$$q_{F+E}\Pi_{F+E} = \frac{q_F\Pi_F}{\mathrm{Tr}\left(K_{FF} \times K_{FF} - K_{FF}\right)} + \frac{B_{FE}}{\mathrm{Tr}\left(K_{FE} \times K_{EF}\right)} + \frac{B_{EF}}{\mathrm{Tr}\left(K_{EF} \times K_{FE}\right)} + \frac{q_E\Pi_E}{\mathrm{Tr}\left(K_{EE} \times K_{EE} - K_{EE}\right)}$$

Complexity Reduction

Two Definitions

- Purity Indicator - a measure of fragment quality. This comes from measuring the error in Idempotency.
- Fragment Bond Order - a measure of fragment interaction. This comes from the off diagonal terms that were neglected.
- Using these two together, we will introduce the concept of *Complexity Reduction* - dividing large, complex systems into chemically meaningful fragments and measuring their interaction.

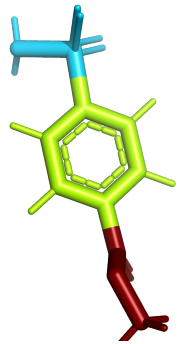
Connection

- Purity Indicator - similar to the classical concept of Atomic Valence.
- Fragment Bond Order - similar to standard atomic bond order.

Automatic Partitioning of Systems

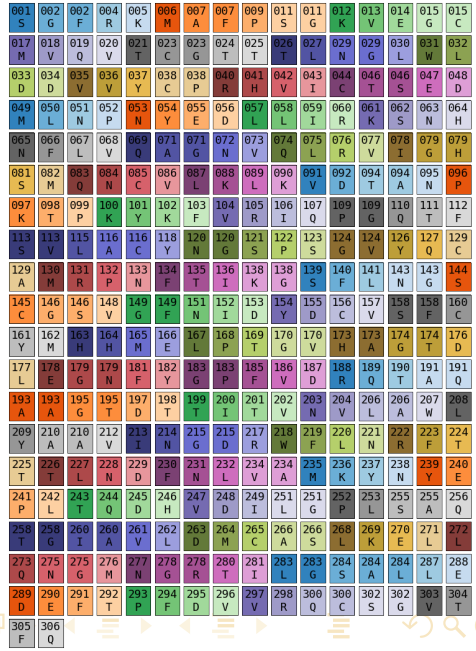
Small Molecules - From Atoms Up

- By combining the purity indicator with optimization algorithms, we can automatically organize a system into fragments.
- We only need to provide a suitable cutoff for terminating the algorithm (0.05 in practice).



Large Biomolecules

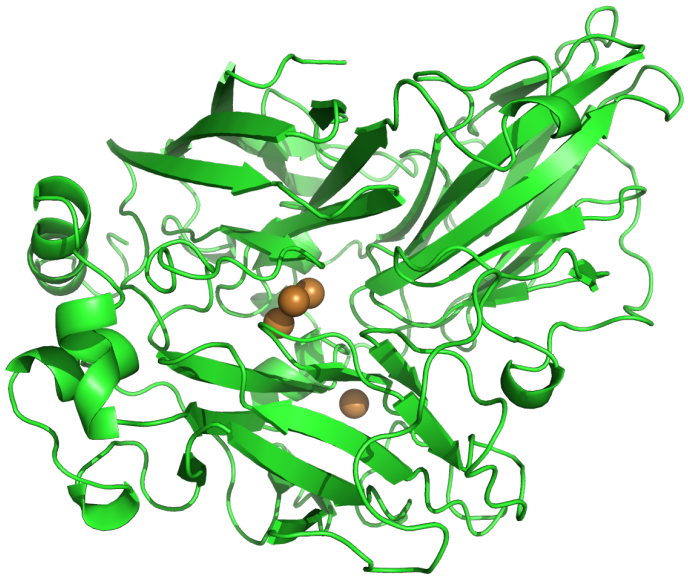
- Proteins are often already divided into fragments based on their Amino Acids.
- Yet not all amino acids are equally good fragments. We can combine them together to build a more coherent picture.

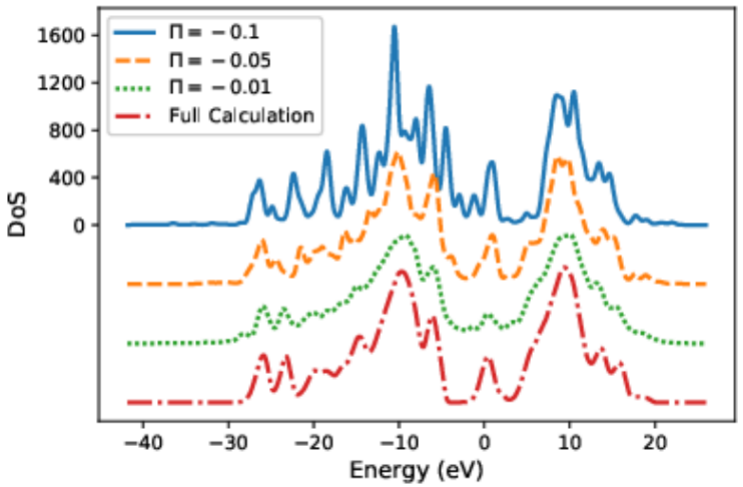


Automatic Partitioning of Systems

Reliable System Partitioning

- When a tight enough threshold is used, we can divide our system into nearly self contained fragments.
- This allows us to reconstruct the properties of a system as a sum of independent calculations performed on each fragment.

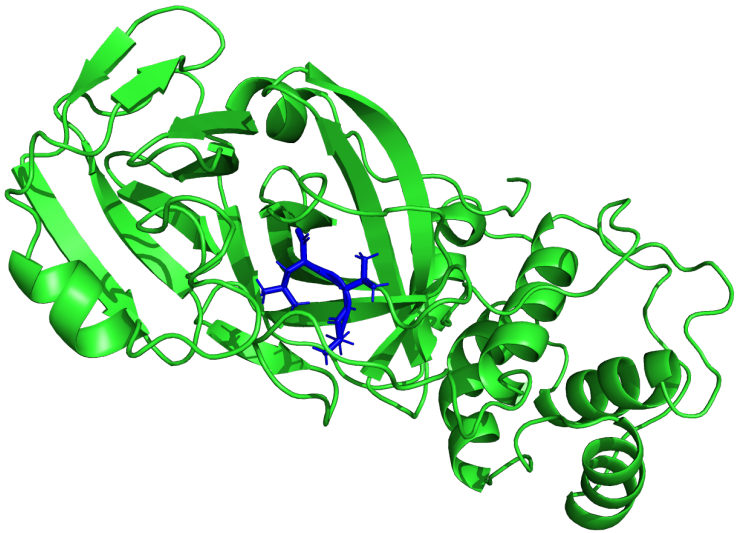


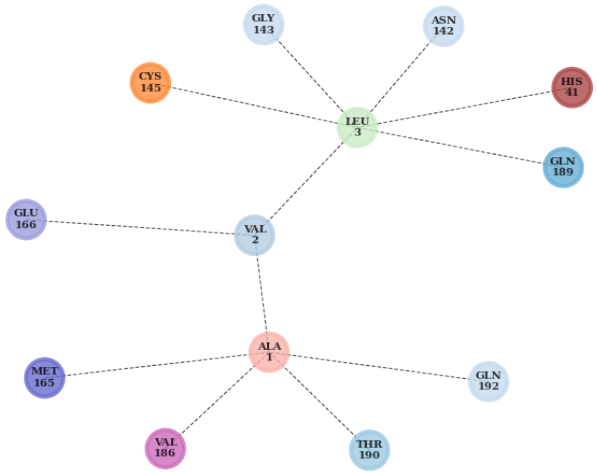


Automatic Graph Views of System

Bond Order Defined Embedding Environments

- We automatically define an embedding environment as the minimal set of fragments such that the sum of the bond order of all remaining fragments with the target is below some threshold.

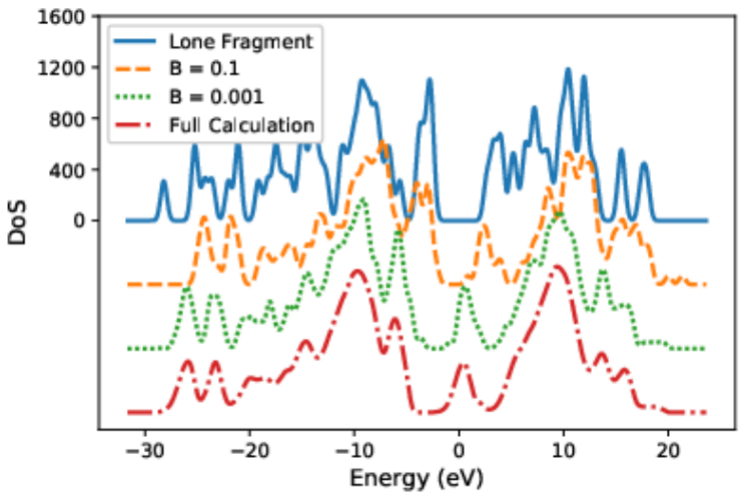


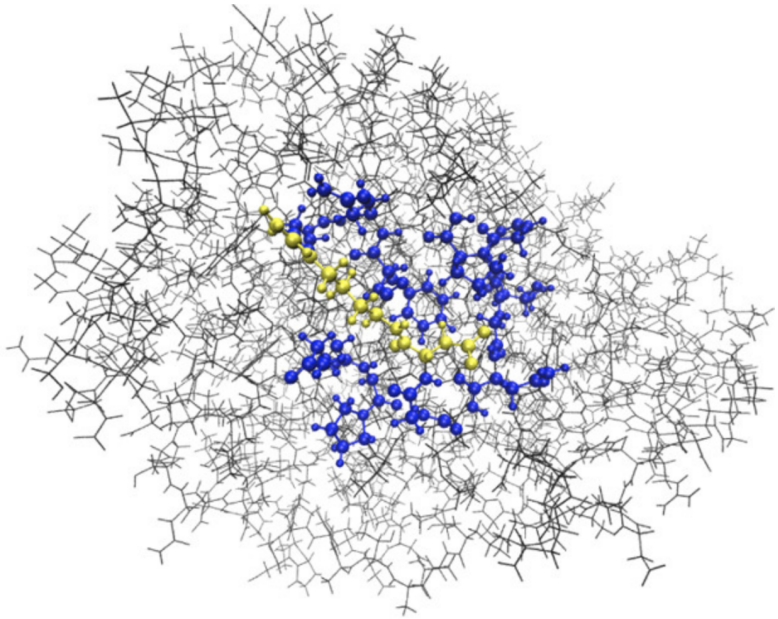


Automatic Embedding of Systems

Reliable System Embedding

- Using this embedding environment, we can construct cluster models of systems.
- The embedded environment serves as a buffer region that enables to compute our desired property of some arbitrary fragment.





Conclusion

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.

Packaging

- The complexity reduction framework presented here is available through our PyBigDFT package.
- Postprocessing can be done even for large systems on a typical workstation.
- If this appears useful to you, we hope to find opportunities to collaborate.