

# Supplementary Information: Combining embedded mean-field theory with linear-scaling density-functional theory

Joseph C. A. Prentice, Robert J. Charlton, Arash A. Mostofi  
and Peter D. Haynes

November 28, 2019

## 1 Pseudopotentials

For the calculations of the pentacene molecule and pentacene in *p*-terphenyl system, the norm-conserving pseudopotentials for carbon and hydrogen distributed with the ONETEP code (`carbon.recpot` and `hydrogen.recpot`) were used.

For the metallocene@CNT calculations, on-the-fly norm-conserving pseudopotentials generated by CASTEP[?] were used. The pseudopotential strings were:

- Co: 3|1.3|32|37|43|30N:40N:31N:32NN(qc=10.5)
- Fe: 3|1.3|32|37|43|30N:40N:31N:32NN(qc=10.5)
- C: 1|1.2|17|20|23|20N:21L(qc=8)
- H: 1|0.8|14|16|19|10N(qc=8)

## 2 Mixed NGWF basis sets

In the standard embedding calculations presented in the main paper, we use a NGWF basis set that is optimised at the lower level of theory, before applying block orthogonalisation and optimising the kernel using EMFT. To further investigate the effect of using NGWFs optimised at different levels of theory, here we present results obtained using a different approach, whose steps are as follows:

1. A full system calculation is performed at both the low and high levels of theory.
2. The resulting sets of NGWFs are partitioned into sets corresponding to the required embedding regions.

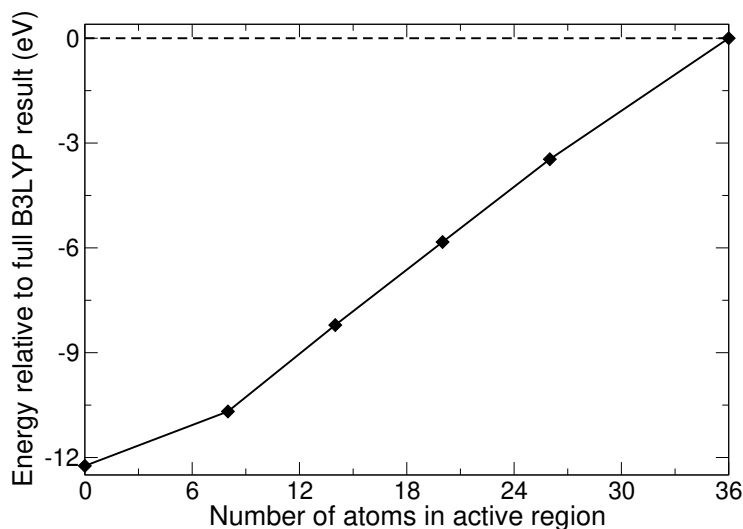


Figure S1: Total energy of a pentacene molecule as a function of the number of atoms treated with B3LYP. NGWFs in the environment are taken from a full system PBE calculation, and NGWFs in the active region are taken from a full system B3LYP calculation. All energies are measured relative to the energy obtained with both the kernel *and* the NGWFs optimised using B3LYP, which is marked with a dotted line.

3. The density kernel of the full system is then optimised with EMFT, using low-level NGWFs for the environment and high-level NGWFs for the active region.

No block orthogonalisation was performed, to ensure that the basis functions used in both regions were directly comparable throughout. The calculations were performed on the pentacene molecule, with the structures and partitionings the same as those shown in Fig. 1 in the main paper. The low level of theory is PBE, and the high level is B3LYP, as in the main paper.

The results are presented in Fig. ?? . It can be seen that there is a smooth convergence from full PBE to full B3LYP. The convergence behaviour is slower than that of the calculations presented in Fig. 2 in the main paper, most likely due to the lack of block orthogonalisation.

## References

- [1] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson, and M. C. Payne. First principles methods using CASTEP. *Z. Kristallogr.*, 220:567, 2005.