ONETEP Tutorials
7.0.0

Tutorial 12

Tutorial 12: Quantum embedding with (time-dependent) embedded mean-field theory: hydrogenation and excitations of pentacene

◁□ Version: all versions ✎ Joseph Prentice

Introduction

The utility of quantum embedding

Although ONETEP [1] makes first principles calculations of systems containing thousands of atoms feasible, particularly when using semi-local exchange-correlation functionals, it can still be very computationally costly to treat entire systems of this size with higher level theory, such as hybrid functionals, which require the computation of exact exchange. This is particularly relevant for excited state calculations, due to the well-known underestimation of the band gap by semi-local DFT, making excitation spectra performed with semi-local DFT quantitatively, or sometimes qualitatively wrong. However, if the physics/chemistry of interest in the system is known or expected to be localised to a particular subregion -- we call this the `active region` -- with the rest of the system acting as an environment influencing the active region, quantum embedding provides a way to achieve hybrid accuracy with a significantly reduced cost.

This is achieved by treating the active region alone with the higher level of theory, with the rest of the system (the environment) treated at a lower level of theory. In ONETEP at present, this translates to treating the active region with hybrid DFT, and the environment with semi-local DFT. This is done within a single self-consistent calculation, to ensure that the two regions are able to influence one another. As hybrid DFT is only performed on the active region, which is typically small compared to the environment, the computational cost is significantly reduced.

Embedded mean-field theory

There are many different quantum embedding schemes; the scheme used in ONETEP is embedded mean-field theory (EMFT), originally proposed by Fornace et al. [4] This scheme has several advantages, including two particularly relevant to ONETEP: firstly, it partitions the system at a basis-set level, which works well with ONETEP's atom-centred NGWF basis. Secondly, it is a mean-field theory throughout, like DFT, which means that many existing methods based on DFT can be easily modified to accommodate EMFT. This includes linear-response time-dependent DFT (LR-TDDFT), which allows electronic excitations to be computed -- we refer to this as TD-EMFT. For more details, please see the original EMFT paper [4], and the papers implementing ground state EMFT [2] and TD-EMFT [3] in ONETEP.

EMFT is also fully compatible with the implicit solvent methods available within ONETEP, allowing for multi-level modelling of systems.

Pentacene

In this tutorial, we will look at both ground state EMFT and TD-EMFT, with the pentacene molecule as our test case. Ground state EMFT is demonstrated by looking at the terminal hydrogenation energy of pentacene, following Prentice et al. [2], and TD-EMFT is demonstrated by looking at the first excitation energy of pentacene-doped p-terphenyl, following Prentice. [3]

Ground-state EMFT: terminal hydrogenation of pentacene

Non-EMFT baseline calculations

The terminal hydrogenation reaction for pentacene involves two hydrogen atoms becoming bonded to the two carbon atoms at one end of the pentacene molecule.

Before using EMFT, we must first obtain the terminal hydrogenation energy without EMFT, with everything treated with first the PBE, then the B3LYP functionals. The input files required are `Pentacene.dat`, `HydrogenatedPentacene.dat`, and `H2.dat`.

If you look at these files, you will notice that we have several different labels for carbon atoms (C1, C2, etc.) and similar for hydrogen atoms -- these will be used later to vary the size of our active region, by selecting different atoms to be included within it.

To begin, simply run the input files as they are to obtain the ground state energy for the three structures at the PBE level. The hydrogenation energy can then be obtained as:

$$\Delta E_{\text{hyd}} = E_{\text{Hydrogenated Pentacene}} - (E_{\text{Pentacene}} + E_{\text{H2}}).$$

Make sure you save the `.tightbox_ngwfs` file from the `H2.dat` calculation for use later!

Now repeat this for B3LYP. In all three `.dat` files, change `xc_functional` to be `B3LYP`, and add the following keywords/blocks to set up Hartree-Fock exchange (deleting any species labels in the `species_swri...` block that are not present in that particular structure):

```
 1      %block swri
 2         for_hfx 3 12 V 12 12 WE2
 3      %endblock swri
 4
 5      %block species_swri-for_hfx
 6      C
 7      C1
 8      C2
 9      C3
10      H
11      H1
12      H2
13      H3
14      %endblock species_swri-for_hfx
15
16      hfx_use_ri for_hfx
17      hfx_max_l 3
18      hfx_max_q 12
```

For more details on the meaning of these keywords, see the HFx documentation. The most important point for our purposes here is that we can control which atoms are included in the computation of HFx through the `species_swri...` block -- this will become important for our EMFT calculations.

For reasons of stability, it is best to start the B3LYP H2 calculation from the PBE-optimised NGWFs: bring back the PBE-optimised `.tightbox_ngwfs` file, and add `read_tightbox_ngwfs : T` to `H2.dat`. Make sure you still keep a copy of the PBE-optimised `.tightbox_ngwfs` file safe, as we will need it later.

Run these three calculations and compute the hydrogenation energy at the B3LYP level.

EMFT calculations

4 carbon atoms

We can now start to use EMFT to see if we can get close to the B3LYP result without treating the entire molecule with B3LYP. Initially, our active region will just include the 4 C atoms closest to the site of the reaction, and the hydrogen atoms bonded to them. We will only need to do EMFT calculations for `Pentacene.dat` and `HydrogenatedPentacene.dat` -- the hydrogen molecule will always be in the active region, so should always be treated with B3LYP, although there is one subtlety which will be introduced shortly.

To turn on EMFT, change `xc_functional` back to `PBE`, and add the following keywords to `Pentacene.dat` and `HydrogenatedPentacene.dat` (keep the other modifications you already made for HFx for the moment):

```
 1    use_emft           T
 2    use_emft_follow    T
 3    use_emft_lnv_only  T
 4    block_orthogonalise  F
 5    active_xc_functional B3LYP
 6    parallel_scheme    HOUSE
 7
 8    %block species_ngwf_regions
 9    C1 H1
10    C C2 C3 H H2 H3
11    %endblock species_ngwf_regions
```

A brief explanation of each keyword:

- `use_emft`: this turns on EMFT, so that the active region and environment are treated with different levels of theory.

- `use_emft_follow`: this toggles whether a non-EMFT calculation at the lower level of theory (in this case, PBE) is done first, and then uses that as a starting point for the EMFT calculation. For this to happen, the value should be `T`.

- `use_emft_lnv_only`: this toggles whether EMFT is used to optimise both the NGWFs and the density kernel (value is `F`), or just the density kernel (value is `T`). Typically, EMFT is only used to optimise the density kernel (`T`), as NGWF optimisation is poorly behaved under EMFT -- NGWFs can unphysically optimise towards the region described by the level of theory that predicts the lowest energy, and the block orthogonalisation procedure designed to counteract this (discussed

shortly) makes the NGWF optimisation stall. The NGWFs are optimised at the lower level of theory, and then fixed -- the error this introduces is typically <1% of the difference between the high and low levels of theory. For more details, see Prentice et al. [2]

- `block_orthogonalise` : this toggles whether a block orthogonalisation procedure is applied to the NGWFs before using EMFT. This transforms the NGWFs of the environment so that they are orthogonal to the NGWFs of the active region, so the off-diagonal blocks of the overlap matrix are 0. This prevents the emergence of unphysical solutions that can occur in some systems. For this to happen, the value should be `T`.
- `active_xc_functional` : this selects the functional that will be used in the active region, whilst `xc_functional` selects the functional used in the environment.
- `parallel_scheme` : this decides how MPI processes should be split between the regions. `HOUSE` means that the processes are distributed proportionally to the number of atoms within each region; `SENATE` means that the processes are distributed equally between all regions; and `NONE` means that each region will use all the processes in turn. `HOUSE` is strongly preferred.
- `block species_ngwf_regions` : this assigns species to regions, with one region per line. The first line is the active region by default.

The first four keywords should be turned to `T` in the order they are listed in. The first three keywords should almost always be `T` for an EMFT calculation, with `block_orthogonalise` turned on if the calculation proves unstable without it. Here, we leave it off.

We also need to modify the HFx set-up to match the fact we only want HFx in the active region. To do this, simply delete any species in the `species_swri...` block that are *not* in the active region. Remember that each species in the active region should be on its own line in the `species_swri...` block, whereas all the species in the active region should be on the first line of the `species_ngwf_regions` block.

Once these additions/modifications have been made, run the calculations for pentacene and hydrogenated pentacene.

Before we can compute the hydrogenation energy from these results, there is one more subtlety. As we optimised the active region NGWFs at the lower level of theory, but the active region density kernel with the higher level of theory, we need to do the same in our hydrogen molecule for consistency. To do this, bring back the `H2.tightbox_ngwfs` file you saved from the PBE calculation earlier, and re-run your B3LYP `H2.dat` calculation with the following modifications/additions:

```
1    read_tightbox_ngwfs : T
2    maxit_ngwf_cg : 0
```

You can now use these three results to compute the hydrogenation energy with an active region of this size.

Larger active regions

Next, expand the active region to include the 8 carbon atoms closest to the reaction site. To do this, add the C2 and H2 species to the active region (remember to remove them from the environment region!), and modify the HFx set-up to match. Rerun the pentacene and hydrogenated pentacene calculations,

and compute the hydrogenation energy (you don't need to rerun the hydrogen molecule calculation, as you can just reuse the result obtained using PBE-optimised NGWFs).

Finally, expand the active region further to include the 12 carbon atoms closest to the reaction site, by adding C3 and H3 to the active region. Re-calculate the hydrogenation energy.

If you plot the hydrogenation energy vs. the size of the active region, you should see the hydrogenation energy approach the full B3LYP result. This demonstrates the ability of EMFT to obtain high level results at a reduced cost, even when the boundary between regions cuts through covalent bonds.

This also demonstrates the importance of selecting the appropriate active region. In systems made up of weakly bonded parts (e.g. molecular crystals, solvated systems), the appropriate active region will often be obvious -- it will be the molecule or molecules of interest (examples of multiple-molecule active regions could include a nearest-neighbour dimer or a solute along with some nearest-neighbour solvent atoms). In extended covalent or ionic systems, the choice of active region may be more difficult, and should be carefully converged, in a similar way to that shown in this tutorial.

Further investigations

To further investigate the use of EMFT in ONETEP, you could look at the effects of:

- changing the active region further -- perhaps including more C atoms, excluding H atoms, etc.,
- using block orthogonalisation,
- using other functionals for either the high or low level of theory -- semi-local functionals can be used for the higher level, although this is of course not expected to produce a significant advantage,

and many other possibilities.

TD-EMFT: excitations of pentacene-doped p-terphenyl

Non-EMFT benchmark

Here, we will be looking at the S0 to S1 transition in pentacene, which is the lowest excited state observed in TDDFT. This is significantly affected by the environment. In particular, we are interested in pentacene-doped p-terphenyl, which is of interest for room-temperature maser applications, and how the p-terphenyl environment affects the excitation energy.

To give us a reference for isolated pentacene, we first need to perform a high-level TDDFT calculation for pentacene. We will again use B3LYP as our high-level theory. The input file is `Pentacene_isolated.dat` -- as this tutorial assumes you are already familiar with running TDDFT calculations with ONETEP, we will not go into any detail, and this calculation can just be run as it is.

TD-EMFT calculation

We now perform a TD-EMFT calculation for a pentacene molecule surrounded by 6 p-terphenyl molecules, as extracted from the pentacene-doped p-terphenyl molecular crystal. The input file is `Pentacene_in_p-terphenyl.dat`. This can be run as it is, but one point regarding EMFT should be

noted first. The general set-up of the EMFT calculation is precisely the same as for ground state EMFT, with one addition: the `species_tddft_kernel` block. By using this block, we can specify which species we will restrict our excitations to be localised on. Given that in a TD-EMFT calculation we expect the excitations of interest to be localised within the active region, the species contained within the `species_tddft_kernel` block should be a subset of those in the active region. Typically, the two will be identical, i.e., the contents of the `species_tddft_kernel` block should be the same as the first line of the `species_ngwf_regions` block. For more details, see the LR-TDDFT documentation.

Run this calculation -- this may take some time. If you compare this to the results in Prentice [3], you can see that the result is very close to the experimental value of 2.09 eV.

You can plot the resulting excitation as a `.cube` file, and visualise it using e.g. VESTA.

Files for this tutorial

- ⬇ Pentacene.dat
- ⬇ HydrogenatedPentacene.dat
- ⬇ H2.dat
- ⬇ Pentacene_isolated.dat
- ⬇ Pentacene_in_p-terphenyl.dat
- ⬇ C_NCP19_PBE_OTF.usp
- ⬇ H_NCP19_PBE_OTF.usp

References

1. J. C. A. Prentice, J. Aarons, J. C. Womack, A. E. A. Allen, L. Andrinopoulos, L. Anton, R. A. Bell, A. Bhandari, G. A. Bramley, R. J. Charlton, R. J. Clements, D. J. Cole, G. Constantinescu, F. Corsetti, S. M. M. Dubois, K. K. B. Duff, J. M. Escartin, A. Greco, Q. Hill, L. P. Lee, E. Linscott, D. D. O'Regan, M. J. S. Phipps, L. E. Ratcliff, A. Ruiz Serrano, E. W. Tait, G. Teobaldi, V. Vitale, N. Yeung, T. J. Zuehlsdorff, J. Dziedzic, P. D. Haynes, N. D. M. Hine, A. A. Mostofi, M. C. Payne, and C.-K. Skylaris, *The ONETEP linear-scaling density functional theory program*, J. Chem. Phys. 152, 174111 (2020).

2. J. C. A. Prentice, R. J. Charlton, A. A. Mostofi, and P. D. Haynes, *Combining Embedded Mean-Field Theory with Linear-Scaling Density-Functional Theory*, J. Chem. Theory Comput. 16, 354 (2020).

3. J. C. A. Prentice, *Efficiently Computing Excitations of Complex Systems: Linear-Scaling Time-Dependent Embedded Mean-Field Theory in Implicit Solvent*, J. Chem. Theory Comput. 18, 1542 (2020).

4. M. E. Fornace, J. Lee, K. Miyamoto, F. R. Manby, and T. F. Miller, *Embedded Mean-Field Theory*, J. Chem. Theory Comput. 11, 568 (2015).