comatsci-embed
QM-MM embedding extension to the
computational **mat**erials **sci**ence utility package

Jan M. Knaup, Bremen Center for Compuational Materials Science
Jan.Knaup@bccms.uni-bremen.de

February 5, 2009

Contents

1 Introduction

1.1 Requirements

comatsci-embed is an extension to the comatsci basic package, therefore comatsci-basic must be installed to use comatsci-embed.

comatsci-basic requires the following software/libraries to be present on a system:

- Python version $\geq$ 2.4 (not comatible to python 3)
  - POSIX compatible build environment and header files to compile python c-extensions
- numpy
- pyqt version $\geq$ 3 (only required for the geostats utility)
- ElementTree (included in Python version $\geq$ 2.5)

Note that windows is not an officially supported platform for comatsci.

1.2 Installation

comatsci should be installed via the supplied setup script. Execute `python setup.py install` as root, to install comatsci system-wide. Any other installation tree can be chosen with the `--prefix` option to install, e.g. `--prefix=~` to install into `$HOME/bin` and `$HOME/lib`. The setup script also offers options to generate installers for different operating systems, depending on the platform and installed python version, please refer to the integrated documentation available by calling `python setup.py --help`.

1.3 License

comatsci-embed is provided without any warranty without charge for academic use, hower, you are not entitled to redistribute the package in any way, also you are required to cite comatsci-barrier in any publication of results obtained using comatsci-barrier. Please refer to the attached LICENSE file or appendix A.

2 Utilities

2.1 geostats

comatsci-embed extends the functionality of comatsci-basic. Once comatsci-embed is installed together with comatsci-basic, the geostats utility will offer options for the generation of differently embedded QM zones from a provided two-layer geometry.

In all cases, the geometry must have at least two layers. The QM-MM embedding is performed between two layers specidief in the embedding dialog. If layers with the default names are present, they are preselected in the dialog, otherwise the default layer will be preselected. Make sure to select the correct layers for embedding. The default layer names are “QMZ” for the QM zone and “PCHR” for the external charges zone.

Currently, the following embedding modes are provided (cf. Refs [**?**] for documentation of the different embedding modes):

SLA simple link atoms embedding. QM-MM bonds will be saturated on the QM side by adding hydrogen atoms. The atoms will be added at their eqilibrium distance (computed from the elemental covalent radii) multiplied by the link-atom distance factor. No manipulation of charges is performed.

HCS homogeneous charge subtraction. QM-MM bonds will be saturated on the QM side by adding hydrogen atoms. The atoms will be added at their eqilibrium distance (computed from the elemental covalent radii) multiplied by the link-atom distance factor. The cluster charge will be compensated by subtracting an equal portion from the charge of the QML atom and adding it to the MMH atom. Use with care in inhomogeneous systems!

BCTC bond charge transfer compensation. QM-MM bonds will be saturated on the QM side by adding hydrogen atoms. The atoms will be added at their eqilibrium distance (computed from the elemental covalent radii) multiplied by the link-atom distance factor. The cluster charge will be compensated by subtracting a bond-type specific charge transfer from each QML atom and adding it to its respective QMHA. Bond charge-transfer coefficients can be calculated from the system, entered manually or read from an external file.

2.2 scale-linkdists

This helper script reads an input geometry and writes out a .gen file, after applying a distance scaling factor to the bonds of a block of atoms at the end of th einput geometry. Command line:

```
scale-linkdists [input filename] [output filename] [# of non-link atoms] [distance
factor]
```

none of the positional arguments is optional. The number of non-link atoms specifies the length of the block of atoms at the beginning of the geometry which will not be moved.

2.3 chargeanalys-2D

Reads the outputs of a 2-D map of DFTB outputs and, calculated RMS and absolute charge deviations of the QMZ core atoms and outputs data in gnuplot-friendly 2D mapped data files. Command line:

```
chargeanalys-2D [reference geometry]
```

The reference geometry must be a .fdf file containing only the QM zone, with the QML atoms as one block at the end of the geometry and the QML atoms marked as subtybe “H_l”. (This is the output of QM/MM embedding done via comatsci).

input files:

reference geometry chargeanalys-2D expects the reference geometry to contain atomic charges. Hence it must be in .fdf format, as it is the only supportef format for reading that supports this feature.

calculation results DFTB$^+$ calculation results are expected. Currently detailed.out files from DFTB versions 1.0 or 1.1 are supported. The files are expected to be named `detailed-[x]-[y].out`, as usual, gzip compressed files can be privided instead, which is highly recommended for the mapping calculation results. Conventionally, [x] should be the Gaussian blur width and [y] the link atom distance factor, but the meaning of the mapping parameters is of no practical interest to chargeanalys-2D.

output files:

dQ.dat Summed charge deviation of the QM zone atoms (QML excluded) mapped over the embedding parameters.

RMSQ.dat RMS charge deviation of the QM zone atoms (QML excluded) mapped over the embedding parameters.

2.4 dosanalys-2D

Reads the outputs of a 2-D map of DFTB outputs and writes overall band shift, maximum cross-correlation and, if requested, the state sum in a band-shifted energy range to gnuplot friedly 2D scalar map files. Command line:

```
dosanalys-2D {options} [reference spectrum]
```

options:

-h, –help show this help message and exit

-f PF, –peak-function=PF Calculate the DOS as a superposition of this type of PFs at the eigenvalues. default=lorentz

-w PF, –peak-width=PF Width parameter of the peak function (HWHM for lorentz, sigma for gauss). default=0.3

-s SW, –step-width=SW Sample the DOS-s at every SW eV. default=0.1

–min=ESet lower bound of DOS range to E. default: lowest eigenvalue

–max=E Set upper bound of DOS range to E. default: highest eigenvalue

–shifted-range-states=EL,EH calculate shifted DOS integral difference, between EL and EH, shifted by cross-correlation maximum position

input files:

reference spectrum The reference spectrum must be a band.out file from DFTB^{+}.

calculation results DFTB^{+} calculation results in band.out format are expected. The files are expected to be named `band-[x]-[y].out`, as usual, gzip compressed files can be privided instead, which is highly recommended for the mapping calculation results. Conventionally, [x] should be the Gaussian blur width and [y] the link atom distance factor, but the meaning of the mapping parameters is of no practical interest to dosanalys-2D.

output files:

bandshifts.dat total shift of the band structured determined from the peak in cross-correlation between calculation result and reference DOS.

maxcorrelation.dat maximum cross-correlation value for each mapping point.

shiftedrangestated.dat state sum in the shifted energy range. Cf. **–shifted-range-states** option, only written if this option is supplied.

2.5 dosanalys-3D

Reads the outputs of a 2-D map of DFTB outputs and writes all calculated DOS values into an ASCII legacy format vtk file, suitable for vtk and various data visualizers based on that. Command line:

```
dosanalys-3D {options}
```

options:

-h, –help show this help message and exit

-f PF, –peak-function=PF Calculate the DOS as a superposition of this type of PFs at the eigenvalues. default=lorentz

-w PF, –peak-width=PF Width parameter of the peak function (HWHM for lorentz, sigma for gauss). default=0.3

-s SW, –step-width=SW Sample the DOS-s at every SW eV. default=0.1

–min=E Set lower bound of DOS range to E. default: lowest eigenvalue

–max=E Set upper bound of DOS range to E. default: highest eigenvalue

output files:

DOSvalues-3D.vtk Calulated Density Of States in .vtk format. Data is organized as a 3D rectilinear grid of scalars.

3 Embedding workflow

The QM/MM embedding scheme described here only regards the embedding of the QM zone and the termination of the dangling bonds resulting from the partitioning of the extended system into QM and MM zones. Mechanical embedding, coupling on the MM side and simulations based on the QM/MM scheme are not the subject of comatsci-embed or this manual.

3.1 Constructing an embedded cluster

The QM/MM embedding relies on the availability of QM reference data, which has to be calculated in advance.

1. Perform a DFTB reference calculation of the target system. Typically this would be a periodic supercell model. In the reference calculation, the atomic coordinates should be optimized.

2. Obtain a geometry file of the target system in .fdf format, containing atomic charges.

   E.g. by converting the output of a DFTB$^+$ calculation using:
   ```
   tofdf geo_end.gen -p detailed.out QM-reference.fdf
   ```

3. (only if BCTC embedding is intended) Obtain bond charge transfer coefficients for the reference system:
   ```
   geostats QM-reference.fdf
   ```
   select: statistics->save BCT coefficients

4. Periodically extend the reference system to the intended size of the MM Zone:
   ```
   tofdf QM-reference.fdf -x a:b:c fullsystem.fdf
   ```
   a,b,c denote the numbers of reference unitcells that the final supercell should contain. Alternatively, the “edit->peridoc expand” dialog of geostats can be used.

5. Add a QM zone layer named “QMZ” to the .fdf file, to which all atoms that should form the QM zone can be moved.
   ```
   tofdf --addlayer=”QMZ” fullsystem.fdf
   ```
   Alternatively, add a layer “PCHR” for all atoms that should act as external charges. The layer names “QMZ” and “PCHR” are crucial, naming is case-sensitive.
   Note the layer index of the newly created layer for later use!

6. Obtain a list of atom numbers which will be moved to the QMZ or PCHR layer.

   in VMD console:
   ```
   qmzone=atomselect <molecule number> “<atom selection string>”
   qmzone get serial
   ```

7. Move QMZ or PCHR layers into the respective layer.

   choose one of:
   ```
   tofdf --tolayer <layer index> --atomlist=”<list of atom serial numbers
   counting from 1>” fullsystem.fdf partitioned.fdf
   tofdf --tolayer <layer index> --atomindices=”<list of atom serial numbers
   counting from 0>” fullsystem.fdf partitioned.fdf
   ```
   as appropriate for your atom list.

8. Load the partitioned geometry into geostats and select the desired embedding method from the edit menu.

9. In the embedding dialog, enter the embedding parameters. If you are constructiong a cluster for parameter fitting/validation, set the link atom distance factor to 1.0. Klick “embed” and copy/paste the embedding results displayed in the dialog for your documentation.

10. Save the embedded geometry to a new .fdf file, using the file->save as menu.

11. Save the QM zone atoms to a .gen file as DFTB input
    ```
    togen -l <QMZ layer index> embedded.fdf input.gen
    ```
    If your extended system is periodic, *manually!* Change the geometry mode of input.gen to “c” for cluster and delete the last 4 lines in the file, which describe the periodic boundary conditions.

12. Obtain a list of the serial numbers of atoms in the QMZ input file (input.gen) to keep fixed during the calculations, this should at least include the QMH and QML atoms. Cf. step 6.

13. Save the external charge distribution as an .xyzq file for DFTB input.
    ```
    toxyzq -l <PCHR layer index> embedded.fdf pchr.xyzq
    ```

14. Add the appropriate external electic field specification to the dftb_in.hsd file for your later calculations. Set the Gaussian blur width as determined during fitting.

The external charges files are independent of the link atom distance factor, however, they depend upon the embedding method and any additional parameters an individual method may use. Make sure, not to re-use inappropriate .xyzq files for newly-embedded clusters!

3.2 Fitting of embedding parameters

To determine the proper Gaussian blur widths of the external charge distributin and the appropriate link-atom distance factor, the self-consitent atomic partial charges and the density of states of the embedded cluster must be mapped with respect to the two embedding parameters and compared to the reference system. Refer to Ref. for an in-depth scientific discussion of the fitting criteria.

The workflow for the parameter mapping is

1. Construct a QM surface cluster as described in section 3.1.

2. Save the QM zone atoms to an .fdf file as reference for the fitting
   ```
   tofdf -l <QMZ layer index> embedded.fdf QMZ-reference.fdf
   ```

3. Save the quasi single particle eigenvalues from the reference calculation

4. Do test calculations of the QM zone mapped over Gaussian blur widths and link atom distance factors. The necessary mapping ranges may vary between materials. In aluminna, GBW ranges from 0.1. . .5 a.u. and LADF ranged between $0.5 \ldots 1.2$ have proved reasonable.
   In the mapping calculations, set the maximum number of SCC iterations to a high value (e.g. 1000) to avoid convergence problems. Keep all atomic positions fixed at the coordinates from the reference calculation.

5. Map the atomic charge and DOS reproduction parameters
   ```
   chargeanalys-2D QMZ-reference.fdf
   dosanalys-2D reference-band.out --min=<minimum energy> --max=<maximum
   energy> --shifted-range-states=<reference VBM>,<reference CBM>
   ```
   The energy ranges should be chosen to encompass the valence and conduction bands of the reference calculations with a few eV headroom on both ends to allow for band structure shifts. It may be useful to change the peak function width and sampling step-widths for the DOS analysis from their default values, which are rather coarse.
   The DOS analyis is slow, expect runtimes of ~20 minutes to 1 hour for typical mappings.

6. Plot the mapped embedding results and chose a combination of Gaussian blur width and link atom distance factor based on the results. It has proved reasonable to rank the targets in the following manner:

    (a) RMS charge deviation
    (b) summed charge deviation
    (c) band shift
    (d) state sum in shifted energy range

7. Perform further validation of the embedded cluster at the chosen parameters (cf. section 3.3).

Mapping script example

(Note that mapping ranges are shortened for readability.)

```
#!/bin/bash
binary=/usr/locla/bin/dftb+_1.1
# i runs over Gaussian blur widths
for i in 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4
do
# j runs over link atom distance factors
     for j in 0.600 0.625 0.650 0.675 0.700 0.725 0.750 0.775 0.800
     do
           /usr/local/bin/scale_linkdists qmz.gen qmz-moved.gen 100 $j
           echo -n .
           echo "GaussianBlurWidth="$i > gbw.in
           echo -n .
           $binary > gbw-$i-$j.out
           mv detailed.out detailed-$i-$j.out
           mv band.out band-$i-$j.out
           if [ -f results.tag ];
           then
           mv results.tag results-$i-$j.tag
           fi
           echo -n .
           gzip -9 -f *-$i-$j.out *-$i-$j.tag
           echo -n :
     done;
done;
```

3.3 Validating an embedded cluster

Validation is based on qualitative and quantitative checks, how far the embedded cluster deviates from the reference geometry, when atomic positions of the inner atoms (i.e. all but the QML and QMH atoms) are optimized. No bond-braking or formation should be tolerated, the RMS displacement must be as small as possible and within acceptable limits for the intended application.

Further checks of the differences between reaction energies etc. of small reaction between fully QM reference caluculations and the embedded system are advised.

A comatsci-embed license for academic use

Terms and conditions to obtain a copy of comatsci-embed

Preamble

In the following the term "author" refers to Jan M. Knaup. The term "program" shall mean the copyrighted comatsci-qmmm package (in source, object and executable form).

This licence covers the use of comatsci-qmmm for non-commercial purposes as ene extension to the comatsci package, other uses may be arranged individually by discussion with the author.

Mandatory terms and conditions

I request a copy of the comatsci-qmmm package and therefore agree to the following terms and conditions to use the program and to have access to the source code:

1. All title, ownership and rights to the program or to any copies of it remain with the author, regardless of the ownership of the media on which the program is used or installed.
2. I will not provide a copy of the program in any way to anyone for any reason whatsoever. I may make copies of the code but only for backup purposes, or to use it on any of my group's computer resources. I will refer any request for copies of the program to the author. Anyone directly associated with my research group is covered by the licence and may share my copy of the program on all of my group's computer resources as long as they agree to the mandatory terms and conditions.
3. I will not redistribute any part of the program as part of another software without written permission of the author.
4. The program and any documentation is supplied on an as is basis without any warranty, and thus with no additional responsibility or liability. I understand that support for running the program can not be provided in general, except on the basis of a joint project between the code author and the research partner.
5. It is obvious that modifications of the program can lead to problems where the author may not be able to help. It is encouraged that I report any difficulties encountered in the use of standard versions of the program, which I have not modified, to the author.
6. Users are encouraged to report useful modifications to programs to the author. I agree that any reported modifications could be incorporated into the program, potentially with further modifications. I certify that I am entitled to, and bear responsibility for contributing these modifications to the program and that they will be covered by this licence once incorporated. The contributed modifications are non-exclusively for use in the program, so that my rights to use the original modifications I have contributed for other purposes are not affected.
7. In any publication derived from use of the program, I will make references to relevant papers about the program (comatsci-qmmm), the QM/MM coupling method method and any appropriate extensions used. The manual of the program contains the list of the papers to consider for citation.
8. One copy of each and every publication resulting from use of the program will be sent to the contact address of the author for their attention.
9. comatsci-qmmm may rely on external libraries for operation. If any of these libraries are included in the comatsci-qmmm distribution, the repsective terms and conditions for their redistribution libraries are given in the appropriate source directories.