comatsci
computational **mat**erials **sci**ence utility package

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

November 7, 2008

Contents

1 Introduction

1.1 Requirements

comatsci 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 is provided without any warranty as open software under the terms of the Non-commercial Open Software License v3.0. Please refer to the attached LICENSE file or appendix A.

2 Utilities

2.1 geostats

Geostats.py is a Qt gui utility to extract some statistical information and to edit the geometry.

The command line is: `geostats.py [options] [<geometryfile>]`

Possible options are:

`-h, --help` show this help message and exit

`-c TYPE, --convert=TYPE` convert input file to format TYPE and exit

gen dftb generic file format

xyz standard .xyz format

fdf SIESTA fdf Geometry specification

pdb protein database file format, suitable for some visualization purposes

fmg flexible molecular geometry xml format

Currently geostats.py can output a bond list and some basic atomic coordination statistics. The statistics output can be saved to “geostats.htm”. Radial distribution functions and atomic charge histograms (per element) can be computed. These Data can be saved as ascii-files suitable for gnuplot, as well as directly plotted and saved in various graphics file formats using gnuplot, if the gnuplot python extension and gnuplot itself are available.

Additonally geostats.py provides an “edit atoms” dialog in which atom properties can be edited in a table view, atoms can be added or deleted and layers can be added or deleted. This is mainly intended as a convenient way to assign atoms different layers.

A “periodic expand” dialog allows to periodically expand supercells.

2.2 geoconv

Geoconv.py serves to quickly convert geometries between different formats known to comatsci. It is can called as geoconv.py or via symbolic links named to[format], where [format] is the specified output format (cf. -f option). geoconv only works on single geometries but not on paths. Known formats are described in section 3.

Command lines are of the format

`<commandname> [options] <inputfilename> [output filename]`

The input file type is determined by the file name extension. If the output file name is ommitted, the program writes a file of the same name as the input file, with the extension replaced appropriately.

Possible command line options are:

`-h, --help` show this help message and exit

-f F, –format=F Write output geometry in format F. Default=gen. Choose from : ['fdf', 'xyz', 'xyzq', 'gen', 'tm', 'fmg', 'pdb']

`-x XTND, --extend=XTND` periodically extend the input geometry to include a:b:c original supercells in the a:b:c lattice directions, default = 1:1:1 (only the original supercell)

`-p F, --population=F` Read dftb Mulliken population from file F. No default.

`-l L, --layer=L` Write Only Atoms from Layer L into output geometry. default: write whole geometry

`-e E, --element=E` Write only atoms of element E into output geometry. default: write all elements

`--tolayer=L` Move atoms defined by –atomindices or –atomlist to layer L. L must be defined in the geometry, either –atomindices or –atomlist must be given. Cannot be combined with -L.

`--translate=V` Translate atoms defined by –atomindices or –atomlist by vector V. V must be in the form x:y:z. Either –atomindices or –atomlist must be given.

`--atomindices=IDX` Declare, which single atoms are to be modified, argument is a whitespace delimited list of integer atom index numbers. Counts from 0. Mutually exclusive with –atomlist.

`--atomlist=LST` Declare, which single atoms are to be modified, argument is a whitespace delimited list of integer atom serial numbers. Counts from 1. Mutually exlusive with –atomindices.

–addlayer=L Add layer with name L to geometry.

2.3 to[format]

Aliases of geoconv, allowing to quickly convert a readable input geometry to the [format] file format. Aliases are:

- togen
- tofmg
- toxyz
- tofdf
- toxyzq
- totm

3 Known geometry file formats

3.1 .xyz (read/write)

The widely used xmol format for carthesian corrdinates of atoms in molecules. The first line starts with an integer giving the number of atoms in the file, the second line is a comment line and ignored. Following lines list the element symbol and x,y,z coordinates in Angstrom of one atom each. Additional fields are ignored, to ensure compatibility with programs that store additional information in .xyz files. On multi-frame .xyz files, only the first geometry is read.

example: CO molecule

```
2

C 0.0 0.0 0.0
O 1.2 0.0 0.0
```

3.2 .gen (read/write)

Generic molecular and crystalline geometry file format, originally from DFTB implementations. This file format is flexible in its support of cluster and supercell geometries. Please refer to the dftb documentation for a detailed description of the geometry file format.

example: CO molecule

```
# number_of_atoms mode flag
2 C
# chemical symbols of atom types
O C
# number_of_atom type_of_atom cartesian_coordinates_in_angstroms
1 1 0.0 0.0 0.0
2 2 1.2 0.0 0.0
```

4 more lines follow for periodic structures. These contain the coordinates of the origin and three vectors characterizing the supercell.

- The atom type number refers to the ordering of the chemical symbols in line 2. These together determine the element of each atom, other than in dftb.
- The atom numbers (first column of the coordinate lines) are for convenience only but required.
- The flag (first line, second column) is either “C” for clusters/molecules or “S” for supercells.

3.3 .fmg (read/write)

The Flexible Molecular Geometry file format is an xml format intended to store all geometry information that can be processed by comatsci. As an .xml format it is user-editable in principle, however, an xml-aware editor is strongly recommended. Currently fmg provides the following features:

- Supercell and cluster geometries
- Specification of multiple geometries, e.g. for storing reaction paths
- Specification of geometry layers to partition the geometry
- Specification of geometry dimers for use with the dimer method
- Specification of atomic properties:
  - charge
  - layer
  - subtype (e.g. for different force-field atom types within one element)
- Storage of additional trajectory information outside the geometry declaration:
  - Energy
  - Velocities

- Forces
- General trajectory information:
  - Number of geometry iterations so far

C.f. appendix B for the xml document type definition of the flexible molecular geometry format.

3.3.1 example: water

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
   <!DOCTYPE fmg>
   <fmg>
   <geometry>
         <mode>C</mode>
         <lattice orgx=''0.0'' orgy=''0.0'' orgz=''0.0'' lunit=''ang''>
               <latvec_a>1.0 0.0 0.0</latvec_a>
               <latvec_b>0.0 1.0 0.0</latvec_b>
               <latvec_c>0.0 0.0 1.0</latvec_c>
         </lattice>
         <layer>
               <lname>default</lname>
               <li>0</li>
         </layer>
         <atom lunit="ang">
               <x>0.98</x> <y>0.0</y> <z>0.0</z>
               <el>1</el>
               <st>H</st>
               <chr>0.0</chr>
               <li>0</li>
         </atom>
         <atom lunit="ang">
               <x>-0.49</x> <y>0.85</y> <z>0.0</z>
               <el>1</el>
               <st>H</st>
               <chr>0.0</chr>
         </atom>
         <atom lunit="ang">
               <x>0.0</x> <y>0.0</y> <z>0.0</z>
               <el>8</el>
               <st>O</st>
               <chr>0.0</chr>
         </atom>
   </geometry>
   </fmg>
```

The following rules apply:

- the `<fmg>` document tag is mandatory
- multiple `<geometry>` tags are allowed. The geoconv and geostats tools ignore all but the first geometry.
- all attributes are optional
- possible value for lunit are:
  - `ang`: Angstrom, default value

  - au: Bohr radii

- <mode> is optional, possible values are:
  - C — cluster, default value
  - S — supercell

- <lattice> is optional
  The org[xyz] attributes specify the supercell origin (as in the .gen format) and are mostly ignored

- <lattice> must contain exactly one of each of the following elements:
  - <latvec_a>: first lattice vector, 3 coordinates
  - <latvec_b>: second lattice vector, 3 coordinates
  - <latvec_c>: third lattice vector, 3 coordinates

- <layer> is optional, multiple layers may be specified. <layer> must contain exactly one <lname> and <li> element

- <atom> must contain exactly one of each of the following elements:
  - <x>: x coordinate
  - <y>: y coordinate
  - <z>: z coordinate
  - <el>: element number

- <atom> may contain one of each of the following elements:
  - : layer index, default 0
  - <st>: subtype, default [element symbol]
  - <chr>: charge, default 0.0. Unit is electrons, negative values signify electron excess
  - <lpop>: l-shell populations, default is empty

- <trjstep> is outside the geometry definition. It stores data applying to one single <geometry>. To store data for a geometry trajectory, an equal number of trjstep elements in the same order should be present. It may contain the following subelements:
  - <nrg>: Image total Energy
  - <velocities>: Image Velocities data block
  - <forces>: Image forces data block

- <trjinfo> is outside the geometry definition. It stores general information applying to the whole trajectory, rather than stepwise data, as in <trjstep>. Only one <trjinfo> element should be present in one <fmg> and only the first element encountered is evaluated. subelements:
  - <stepcount>: Number of geometry iterations performed so far.

- <dimer> must contain exactly one <geometry> element, specifying the dimer midpoint. It must also contain exactly one <DeltaR> element, specifying the dimer translation vector $\vec{R_1} - \vec{R_0}$. Additionally the following subelements are allowed:

- `<NoGradInRot />` Boolean specifier not to use gradient calculation in rotation step. *deprecated!*
- `<curvature>` The curvature in dimer direction
- `<E0 eunit=””>` Calculated energy at dimer midpoint, possible units are eV and H.
- `<E1 eunit=””>` Calculated energy at dimer endpoint 1, units as in E0.
- `<E2 eunit=””>` Extrapolated energy at dimer endpoint 2, units as in E0.
- `<f0>` Calculated force at dimer midpoint in a.u..
- `<f1>` Calculated force at dimer endpoint 1 in a.u..
- `<f2>` Extrapolated force at dimer endpoint 2 in a.u..
- `<fN>` Projected rotational force on dimer in a.u..

3.4 .pdb (write only)

Protein database format. Refer to PDB documentation for details. .fmg layers are written as segments.

3.5 .xyzq (write only)

Carthesian coordinates and atomic charges in a.u.. Suitable as external electric field input in DFTB$^+$.

example: water

```
1.85193164 0.00000000 0.00000000 0.2938
-0.54141424 1.51278662 0.00000000 0.2938
0.05430593 -0.19438394 0.00000000 -0.5877
```

3.6 .tm (write only)

Turbomole `coord:` block. Refer to Turbomole documentation for details.

3.7 .fdf (write only)

Geometry specification in SIESTA fdf input format. Refer to SIESTA documentation for details.

A Non-Profit Open Software License ("Non-Profit OSL") 3.0

Licensed under the Non-Profit Open Software License version 3.0

1) Grant of Copyright License. Licensor grants You a worldwide, royalty-free, non-exclusive, sublicensable license, for the duration of the copyright, to do the following:

a) to reproduce the Original Work in copies, either alone or as part of a collective work;

b) to translate, adapt, alter, transform, modify, or arrange the Original Work, thereby creating derivative works ("Derivative Works") based upon the Original Work;

c) to distribute or communicate copies of the Original Work and Derivative Works to the public, with the proviso that copies of Original Work or Derivative Works that You distribute or communicate shall be licensed under this Non-Profit Open Software License or as provided in section 17(d);

d) to perform the Original Work publicly; and

e) to display the Original Work publicly.

2) Grant of Patent License. Licensor grants You a worldwide, royalty-free, non-exclusive, sublicensable license, under patent claims owned or controlled by the Licensor that are embodied in the Original Work as furnished by the Licensor, for the duration of the patents, to make, use, sell, offer for sale, have made, and import the Original Work and Derivative Works.

3) Grant of Source Code License. The term "Source Code" means the preferred form of the Original Work for making modifications to it and all available documentation describing how to modify the Original Work. Licensor agrees to provide a machine-readable copy of the Source Code of the Original Work along with each copy of the Original Work that Licensor distributes. Licensor reserves the right to satisfy this obligation by placing a machine-readable copy of the Source Code in an information repository reasonably calculated to permit inexpensive and convenient access by You for as long as Licensor continues to distribute the Original Work.

4) Exclusions From License Grant. Neither the names of Licensor, nor the names of any contributors to the Original Work, nor any of their trademarks or service marks, may be used to endorse or promote products derived from this Original Work without express prior permission of the Licensor. Except as expressly stated herein, nothing in this License grants any license to Licensor's trademarks, copyrights, patents, trade secrets or any other intellectual property. No patent license is granted to make, use, sell, offer for sale, have made, or import embodiments of any patent claims other than the licensed claims defined in Section 2. No license is granted to the trademarks of Licensor even if such marks are included in the Original Work. Nothing in this License shall be interpreted to prohibit Licensor from licensing under terms different from this License any Original Work that Licensor otherwise would have a right to license.

5) External Deployment. The term "External Deployment" means the use, distribution, or communication of the Original Work or Derivative Works in any way such that the Original Work or Derivative Works may be used by anyone other than You, whether those works are distributed or communicated to those persons or made available as an application intended for use over a network. As an express condition for the grants of license hereunder, You must treat any External Deployment by You of the Original Work or a Derivative Work as a distribution under section 1(c).

6) Attribution Rights. You must retain, in the Source Code of any Derivative Works that You create, all copyright, patent, or trademark notices from the Source Code of the Original Work, as well as any notices of licensing and any descriptive text identified therein as an "Attribution Notice." You must cause the Source Code for any Derivative Works that You create to carry a prominent Attribution Notice reasonably calculated to inform recipients that You have modified the Original Work.

7) Warranty of Provenance and Disclaimer of Warranty. The Original Work is provided under this License on an "AS IS" BASIS and WITHOUT WARRANTY, either express or implied, including, without limitation, the warranties of non-infringement, merchantability or fitness for a particular purpose. THE ENTIRE RISK AS TO THE QUALITY OF THE ORIGINAL WORK IS WITH YOU. This DISCLAIMER OF WARRANTY constitutes an essential part of this License. No license to the Original Work is granted by this License except under this disclaimer.

8) Limitation of Liability. Under no circumstances and under no legal theory, whether in tort (including negligence), contract, or otherwise, shall the Licensor be liable to anyone for any direct, indirect, special, incidental, or consequential damages of any character arising as a result of this License or the use of the Original Work including, without limitation, damages for loss of goodwill, work stoppage, computer failure or malfunction, or any and all other commercial damages or losses. This limitation of liability shall not apply to the extent applicable law prohibits such limitation.

9) Acceptance and Termination. If, at any time, You expressly assented to this License, that assent indicates your clear and irrevocable acceptance of this License and all of its terms and conditions. If You distribute or communicate copies of the Original Work or a Derivative Work, You must make a reasonable effort under the circumstances to obtain the express assent of recipients to the terms of this License. This License conditions your rights to undertake the activities listed in Section 1, including your right to create Derivative Works based upon the Original Work, and doing so without honoring these terms and conditions is prohibited by copyright law and international treaty. Nothing in this License is intended to affect copyright exceptions and limitations

(including "fair use" or "fair dealing"). This License shall terminate immediately and You may no longer exercise any of the rights granted to You by this License upon your failure to honor the conditions in Section 1(c).

10) Termination for Patent Action. This License shall terminate automatically and You may no longer exercise any of the rights granted to You by this License as of the date You commence an action, including a cross-claim or counterclaim, against Licensor or any licensee alleging that the Original Work infringes a patent. This termination provision shall not apply for an action alleging patent infringement by combinations of the Original Work with other software or hardware.

11) Jurisdiction, Venue and Governing Law. Any action or suit relating to this License may be brought only in the courts of a jurisdiction wherein the Licensor resides or in which Licensor conducts its primary business, and under the laws of that jurisdiction excluding its conflict-of-law provisions. The application of the United Nations Convention on Contracts for the International Sale of Goods is expressly excluded. Any use of the Original Work outside the scope of this License or after its termination shall be subject to the requirements and penalties of copyright or patent law in the appropriate jurisdiction. This section shall survive the termination of this License.

12) Attorneys' Fees. In any action to enforce the terms of this License or seeking damages relating thereto, the prevailing party shall be entitled to recover its costs and expenses, including, without limitation, reasonable attorneys' fees and costs incurred in connection with such action, including any appeal of such action. This section shall survive the termination of this License.

13) Miscellaneous. If any provision of this License is held to be unenforceable, such provision shall be reformed only to the extent necessary to make it enforceable.

14) Definition of "You" in This License. "You" throughout this License, whether in upper or lower case, means an individual or a legal entity exercising rights under, and complying with all of the terms of, this License. For legal entities, "You" includes any entity that controls, is controlled by, or is under common control with you. For purposes of this definition, "control" means (i) the power, direct or indirect, to cause the direction or management of such entity, whether by contract or otherwise, or (ii) ownership of fifty percent (50%) or more of the outstanding shares, or (iii) beneficial ownership of such entity.

15) Right to Use. You may use the Original Work in all ways not otherwise restricted or conditioned by this License or by law, and Licensor promises not to interfere with or be responsible for such uses by You.

16) Modification of This License. This License is Copyright © 2005 Lawrence Rosen. Permission is granted to copy, distribute, or communicate this License without modification. Nothing in this License permits You to modify this License as applied to the Original Work or to Derivative Works. However, You may modify the text of this License and copy, distribute or communicate your modified version (the "Modified License") and apply it to other original works of authorship subject to the following conditions: (i) You may not indicate in any way that your Modified License is the "Open Software License" or "OSL" and you may not use those names in the name of your Modified License; (ii) You must replace the notice specified in the first paragraph above with the notice "Licensed under <insert your license name here>" or with a notice of your own that is not confusingly similar to the notice in this License; and (iii) You may not claim that your original works are open source software unless your Modified License has been approved by Open Source Initiative (OSI) and You comply with its license review and certification process.

17) Non-Profit Amendment. The name of this amended version of the Open Software License ("OSL 3.0") is "Non-Profit Open Software License 3.0". The original OSL 3.0 license has been amended as follows:

(a) Licensor represents and declares that it is a not-for-profit organization that derives no revenue whatsoever from the distribution of the Original Work or Derivative Works thereof, or from support or services relating thereto.

(b) The first sentence of Section 7 ["Warranty of Provenance"] of OSL 3.0 has been stricken. For Original Works licensed under this Non-Profit OSL 3.0, LICENSOR OFFERS NO WARRANTIES WHATSOEVER.

(c) In the first sentence of Section 8 ["Limitation of Liability"] of this Non-Profit OSL 3.0, the list of damages for which LIABILITY IS LIMITED now includes "direct" damages.

(d) The proviso in Section 1(c) of this License now refers to this "Non-Profit Open Software License" rather than the "Open Software License". You may distribute or communicate the Original Work or Derivative Works thereof under this Non-Profit OSL 3.0 license only if You make the representation and declaration in paragraph (a) of this Section 17. Otherwise, You shall distribute or communicate the Original Work or Derivative Works thereof only under the OSL 3.0 license and You shall publish clear licensing notices so stating. Also by way of clarification, this License does not authorize You to distribute or communicate works under this Non-Profit OSL 3.0 if You received them under the original OSL 3.0 license.

(e) Original Works licensed under this license shall reference "Non-Profit OSL 3.0" in licensing notices to distinguish them from works licensed under the original OSL 3.0 license.

B flexible molecular geometry DTD

```
<!ELEMENT mode (#PCDATA)>
<!ELEMENT lattice (latvec_a, latvec_b, latvec_c)>
<!ATTLIST lattice
orgx CDATA "0.0"
orgy CDATA "0.0"
orgz CDATA "0.0"
lunit (ang|au) "ang"
>
<!ELEMENT latvec_a (#PCDATA)>
<!ELEMENT latvec_b (#PCDATA)>
<!ELEMENT latvec_c (#PCDATA)>
<!ELEMENT layer (li, lname)>
<!ELEMENT li (#PCDATA)>
<!ELEMENT lname (#PCDATA)>
<!ELEMENT atom (x, y, z, el, st?, chr?, li?, lpop?)>
<!ATTLIST atom
lunit (ang|au) "ang"
>
<!ELEMENT x (#PCDATA)>
<!ELEMENT y (#PCDATA)>
<!ELEMENT z (#PCDATA)>
<!ELEMENT el (#PCDATA)>
<!ELEMENT st (#PCDATA)>
<!ELEMENT chr (#PCDATA)>
<!ELEMENT nrg (#PCDATA)>
<!ELEMENT lpop (#PCDATA)>
<!ATTLIST nrg
eunit (eV|au) "au"
>
<!ELEMENT velocities (#PCDATA)>
<!ELEMENT forces (#PCDATA)>
<!ELEMENT geometry (mode?, lattice?, layer?, atom+)>
<!ELEMENT trjstep (nrg?,velocities?,forces?)>
<!ELEMENT fmg (geometry+,trjstep*,trjinfo?)>
<!ELEMENT stepcount (#PCDATA)>
<!ELEMENT trjinfo (stepcount?)>
<!ELEMENT DeltaR (#PCDATA)>
```

```
<!ATTLIST DeltaR
lunit (ang|au) "ang"
>
<!ELEMENT NoGradInRot EMPTY>
<!ELEMENT curvature (#PCDATA)>
<!ELEMENT E0 (#PCDATA)>
<!ATTLIST E0
eunit (eV|au|H) "au"
>
<!ELEMENT E1 (#PCDATA)>
<!ATTLIST E1
eunit (eV|au|H) "au"
>
<!ELEMENT E2 (#PCDATA)>
<!ATTLIST E2
eunit (eV|au|H) "au"
>
<!ELEMENT f0 (#PCDATA)>
<!ELEMENT f1 (#PCDATA)>
<!ELEMENT f2 (#PCDATA)>
<!ELEMENT fN (#PCDATA)>
<!ELEMENT Dimer (Geometry,DeltaR,NoGradInRot?,(E0,E1)?,E2?,(f0,f1)?,f2?,fN?,curvature?)>
```