ab initio Random Structure Search
The Geometry Generator Documentation

Jan Kloppenburg*
Duke University, MEMS Department, Durham 27708, North Carolina, USA
(Dated: May 22, 2014)

I. THE FILE GEN_GEO.PY

The PYTHON3 script called `gen_geo.py` developed in our group[1] is used to generate random geometries within physically sensible ranges.

Input files:

1. The bulk geometry:
   The geometry information of the underlying bulk material must be defined in the FHI-AIMS[2] format. Atomic coordinates are only read from this file to determine the relevant information for the random generation processes. Nothing is modified or written back into this input file by the `gen_geo.py` script. The default name for the input is `bulk_geometry.in` but can be specified to a different name by using `python3 gen_geo.py -i <input_filename> ( -o <output_filename> )`.
2. Geometry input file format:
   Any additional elements or molecules intented to be put onto the bulk material must be specified in files containing the string *ad_geo*. The specification of atomic coordinates must be in the FHI-AIMS format and may not contain lattice vector specifications or fractional atomic coordinates. In addition to the atomic coordinates there are other possible options as:
   (a) `number_of_adds <int>`:
   This keyword specifies how often the geometry specified in the file will be added to the bulk surface. When the `layer_N` keyword is used this number must be the total number of all members of the family to be added.
   (b) `family <str>`:
   This options is used for output purposes only and defines the name of the atoms or molecules printed during the generation processes and serves as an identifier in this manual.
   (c) `layer_N <N_atoms> <z>`:
   This keyword allows to specify up to 10 layers (`layer_1...layer_10`), each with defined thickness `z` in which the specified number of atoms or molecules $\mathtt{N}_{\mathrm{atoms}}$ will be added. The lines
   `layer_1 5 5`
   `layer_2 2 3`
   specify 2 layers, the first one will contain 5 of the family within 5 Å of the bulk surface and the second layer will contain 2 of the family within 3 Å above the first layer. It is of course possible to define multiple families within the same layer so one can vary concentration among different atoms. This is achieved by generating a separate file for each family and specifying the same layer thickness in which the members shall be deposited.
3. Parameter input file format:
   The default name for the input parameters is `parameters` but can be specified by the `--pars <filename>` option in command line. The script `gen_geo.py` has two ways of randomizing the input geometries and other default parameters than can be set to the desired values controlling deposition and orientation of the families defined before.
   (a) `max_wiggle_len <real>`:
   In this routine the coordinates from the family get randomly wiggeled by an adjustable amount given in units of Å. Each coordinate of the family members will be wiggeled by

$$\begin{bmatrix} \mathrm{x} \\ \mathrm{y} \\ z \end{bmatrix} = \begin{bmatrix} \mathrm{x} \\ \mathrm{y} \\ z \end{bmatrix} - w \cdot \left(\begin{bmatrix} \mathrm{rnd}_{\mathrm{x}} \\ \mathrm{rnd}_{\mathrm{y}} \\ \mathrm{rnd}_{\mathrm{z}} \end{bmatrix} - \frac{1}{2} \right) , \tag{1}$$

   where w denotes the specified wiggle length and each coordinate gets it own random number. This functionality in only intented to add a little distortion the the geometry that is added. It is not meant to be used for tearing bonds apart. If you want to add single atoms instead of clusters do not do it by

tearing clusters apart. Usually one generates a tightly relaxed adcluster structure to be put near the surface. To give this structure a small amount of free energy is the only purpose of this wiggle routine. Adjust this parameter carefully and not above maybe maxmal 10% of the shortest bond length of the family.

(b) Family rotation:
This routine rotates the specified family members individually within a random angle range defined by keywords `theta_x, theta_y` and `theta_z`. These angles are the maximal possible angles so that the resulting angles are obtained by

$$\mathrm{rnd}_n = 2 \cdot \big(\texttt{random()} - 0.5\big) \tag{2}$$

$$\begin{bmatrix} \theta_x \\ \theta_y \\ \theta_z \end{bmatrix} = \begin{bmatrix} \mathrm{rnd}_x \cdot \theta_x \\ \mathrm{rnd}_y \cdot \theta_y \\ \mathrm{rnd}_z \cdot \theta_z \end{bmatrix} , \tag{3}$$

while all random numbers in both algorithms (wiggle and rotate) are generated by the `random()` command in the interval $[0, 1)$. Specifying `theta_x = 180` would therefore make a rotation of the full range of 360° possible. Strictly speaking the exact angle of ±180° can not be reached because of the random number being of the interval $\in [0, 1)$.

(c) `min_z_dist <real>`:
This keyword specifies the minimum distance from the bulk surface that will be kept for all added geometries.

(d) `max_z_height <real>`:
This keyword specifies the maximum distance from the surface within that all family members will be added. It is only used when the `layer_N` keyword is not used.

(e) `min_atom_dist <real>`:
This keyword specifies the minimal distance between any atom added to the geometry. Distances specified in input files of the families will not be checked and are unaffected by this keyword. The user should have made sure of sensible geometries before adding them onto a bulk surface.

(f) `maxtries <int>`:
This keyword specifies the maximum number of tries to add a family member to the current geometry. By trying to add too many family members into too little space the random generator will unavoiable find no place to add them into and the run will be aborted after the number of tries specified here.

4. Non-periodic random structure generation:
It is also possible to generate random structures from atoms or molecules in a non-periodic case. For this task there are basically all rules in place as specified above. Some additional rules do apply, though:

(a) The command line option `--non-periodic` must be specified to trigger this functionality.

(b) The file `bulk_geometry.in` now may only contain the `box_dims` specification:
`box_dims <real x> <real y> <real z>`:
Defines the `x`, `y` and `z` dimensions of the box in which the atoms or molecules will be placed at random.

(c) The files containing the string *ad_geo* will be read and treated by the rules specified above. A layerd based approach is not possible in the non-periodic case.

The output file is called `geometry.in` by default and is ready to use with FHI-aims.

* Electronic address: jan.kloppenburg@duke.edu

[1] B. Lange, J. Kloppenburg, and V. Blum, `http://aims.pratt.duke.edu`.
[2] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, "Ab initio molecular simulations with numeric atom-centered orbitals: FHI-aims," Comp. Phys. Commun. **180**, 2175–2196 (2009).