Density Functional Theory and Beyond with Numeric Atom-Centered Orbitals

Hands on tutorial:

Accurate scanning of potential energy surface of small metal oxide clusters: An application of *Cascade* Genetic Algorithm

Saswata Bhattacharya

Fritz-Haber-Institut der MPG, Berlin

bhattacharya@fhi-berlin.mpg.de

FHI-aims Developers' and Users' Meeting - August 19-22, 2014

OUTLINE

Motivation: This tutorial aims to solve the problem of accurate scanning of potential energy surface by first principles based global-optimization technique, viz. a massive-parallel cascade genetic algorithm (GA) as available within fhi-aims package.

Task-1: How to find the global minimum of Mg_4O_4 cluster using GA?

In this part of the tutorial the conventional GA will be used to find low energy isomers including the global minimum (GM) of Mg_4O_4 cluster. Here we will introduce the basic input files, keywords needed to run GA and analysis of the output files after GA.

Task-2: Why to switch on cascade?

Here we will introduce the "cascade" feature of the code. The term cascade means a multi-stepped algorithm where successive steps employ higher level of theory and each of the next level takes information obtained at the immediate lower level. In this part of the tutorial we will show if local/semi-local functional yields wrong energetics for a specific system, how a conventional GA scheme fails miserably to address the low energy structures (e.g. GM) of that system. The main objective of this part is to teach how to overcome such functional related issues by introducing "cascade" feature of the code.

Task-1: How to find the global minimum of Mg_4O_4 cluster using GA?

Directory → AIMS-workshop/Mg4O4

Before a GA-run, the working directory should have all the following files & directories:

```
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/Mg4O4$ ls
child.dat  geom1  index.dat      initial_pool         jobs         Math  plot.py   run
control    geom2  indicator.dat  input_settings.dat  master.plx   out   plotting  user_input.in
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/Mg4O4$
```

- The next step is to prepare the input files to run GA.
- Focus on the following important file/directories located in the working directory.

```
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/Mg4O4$ ls
child.dat  geom1  index.dat      initial_pool         jobs         Math  plot.py   run
control    geom2  indicator.dat  input_settings.dat  master.plx   out   plotting  user_input.in
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/Mg4O4$
```

- user_input.in → control file of GA
- initial_pool → directory where to keep random initial structures to start GA
- control → directory where to keep settings and basis sets to prepare control.in to run fhi-aims

Task-1: How to find the global minimum of Mg_4O_4 cluster using GA?

An optimised user_input.in is provided. The first part of it is to provide "USER SPECIFIC SETTINGS", where the user needs to specify some *"values / keywords"* for a specific GA-run.
For meaning of all the *"values / keywords"*, please consult the user manual of GA.

```
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/Mg4O4$ cat user_input.in
USER SPECIFIC SETTINGS
-------------------------------------------------------------------------
General Settings:
surface/cluster:                          cluster
number_of_children                        1
number_of_structures                      100
aims.out:keep/delete                      keep
initial_moment:hund/custom                hund
control.in_species:default/custom         custom
minimum_bond_length                       0.8
existance_minimum                         0.05
-------------------------------------------------------------------------
BIN_DIRECTORY                             "/home/bhattacharya/Codes/bin"
aims_file                                 "aims.052014.mpi.x"
-------------------------------------------------------------------------
Cascade Settings:
cascade_start                             0
cascade_num_of_steps                      0
energy_structure                          none
search_pattern                            0
cascade_window                            0
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/Mg4O4$
```

Task-1: How to find the global minimum of Mg_4O_4 cluster using GA?

- Apart from "USER SPECIFIC SETTINGS" in user_input.in, there are some technical *"keywords"* that are also essential to be inside user_input.in.
- A file named as "input_settings.dat" is provided with all such technical "*keywords*". Paste it after "USER SPECIFIC SETTINGS" in user_input.in.

```
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/Mg4O4$ ls
child.dat  geom1  index.dat      initial_pool        jobs        Math  plot.py   run
control    geom2  indicator.dat  input_settings.dat  master.plx  out   plotting  user_input.in
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/Mg4O4$
```

```
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/Mg4O4$ cat input_settings.dat >> user_input.in
```

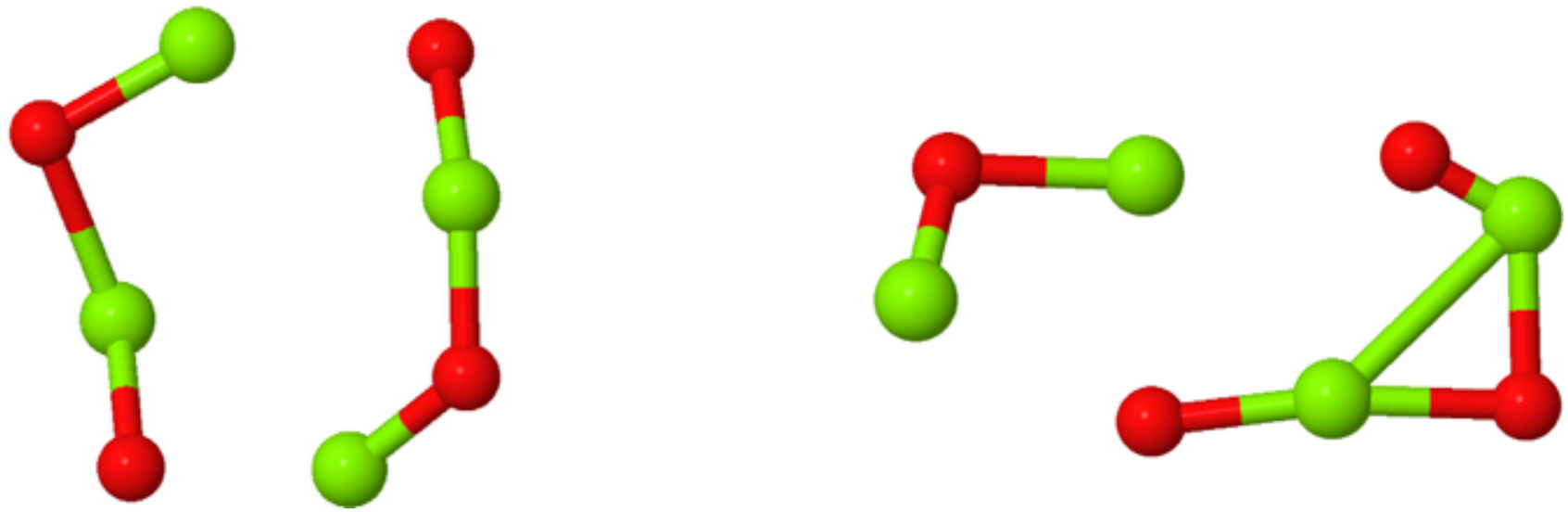
NOTE: Those "*keywords*" as in "input_settings.dat" are optimised after several testing & benchmarking results. Thus it's highly recommended to keep them unchanged, unless the user is really confident enough about such changes.

- user_input.in → control file of GA
- initial_pool → directory where to keep random initial structures to start GA
- control → directory where to keep settings and basis sets to prepare control.in to run fhi-aims

Task-1: How to find the global minimum of Mg_4O_4 cluster using GA?

- initial_pool is the directory where the user needs to keep random structures in geometry.in format. We have already provided two random structures here named as geom1 and geom2.
- For bigger systems, the higher is the number of distinct random structures provided in this directory, the faster will be the GA-convergence.

```
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/Mg4O4$ ls initial_pool/
geom1  geom2
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/Mg4O4$
```



geom1 geom2

- user_input.in → control file of GA
- initial_pool → directory where to keep random initial structures to start GA
- control → directory where to keep settings and basis sets to prepare control.in to run fhi-aims

Task-1: How to find the global minimum of Mg_4O_4 cluster using GA?

- control is the directory where settings and basis sets are provided for generating a control.in to perform local optimisation using fhi-aims.
- settings_main.dat is the settings for control.in and Mg_main and O_main are the basis sets (light settings) of Mg and O atoms respectively.

```
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/Mg4O4$ ls control/
Mg_main  O_main  settings_main.dat
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/Mg4O4$ cat control/settings_main.dat
  xc                 pbe
  spin               none
  relativistic       atomic_zora scalar 1e-9
  occupation_type    gaussian 0.03
  vdw_correction_hirshfeld
  KS_method          scalapack_fast
  mixer              pulay
  ini_linear_mixing 0
  charge_mix_param   0.4
  spin_mix_param     0.6
  sc_accuracy_rho   1E-2
  sc_accuracy_eev   1E-2
  sc_accuracy_etot 1E-1
  sc_iter_limit     100
  empty_states      5
  relax_geometry trm 5E-2
  sc_accuracy_forces 1e-2
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/Mg4O4$
```

- user_input.in → control file of GA
- initial_pool → directory where to keep random initial structures to start GA
- control → directory where to keep settings and basis sets to prepare control.in to run fhi-aims

Task-1: How to find the global minimum of Mg_4O_4 cluster using GA?

- We are all set to run GA. We will submit a job in the local machine as follows:
- ./run/start_GA.plx > out & (Ignore the warnings that are coming from perl).
- Wait for at most two hours to get the global minimum structure of this system.

Analysis of Data:

- Once the start_GA.plx starts running a few important directories viz. "progress", "structures", "structure_temp", etc. will be formed.
- In "progress" directory, there will be a file named as getot.datxxxxxx [xxxxxx is some six digit random number] will give us all the details of hierarchy of different isomers.
- "structures" will store all the new generated structures in "structures/structures_i" directories. This will contain the final geometry.in, aims.out (fhi-aims output file), energy.dat (total energy), etc.
- "structure_temp" will be used for performing temporary local fhi-aims optimisation.

```
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/Mg4O4$ ls
child.dat          geom1      indicator.dat       jobs        out       progress      structures       user_structures
control            geom2      initial_pool        master.plx  plot.py   run           structures_temp
copy_initial_pool  index.dat  input_settings.dat  Math        plotting  stoichiometry  user_input.in
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/Mg4O4$
```

Task-1: How to find the global minimum of Mg_4O_4 cluster using GA?

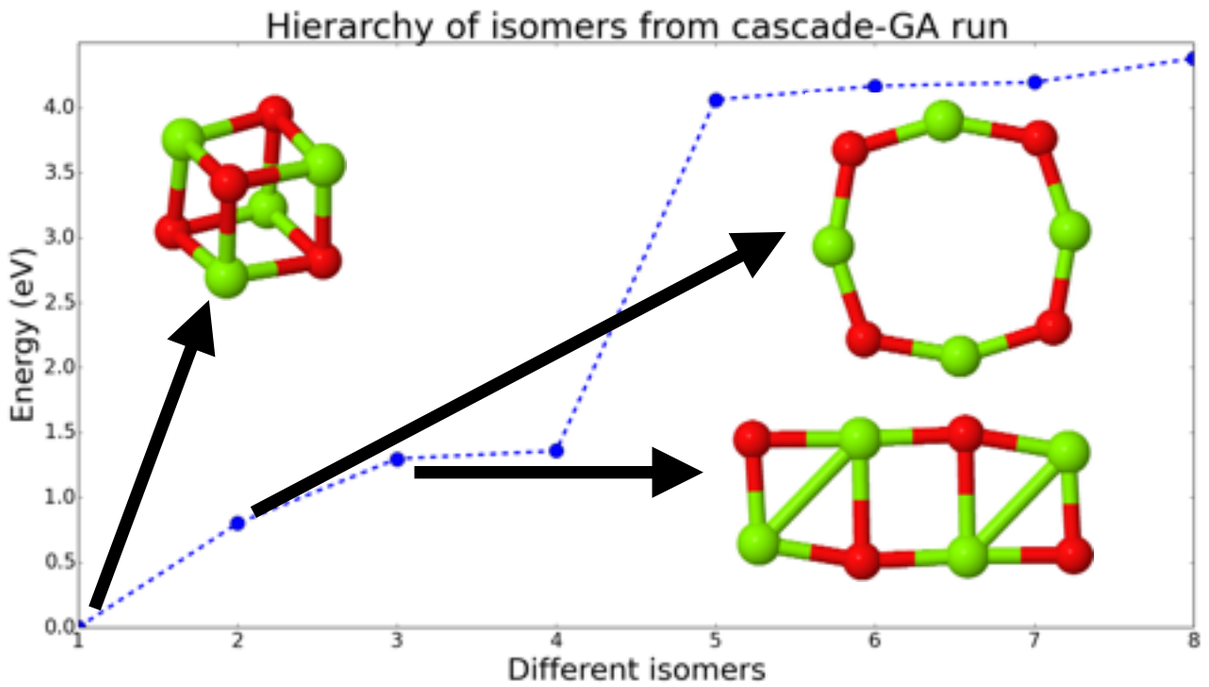
Analysis of Data:

```
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/pre-run-results/Mg4O4$ cat progress/getot.dat189947
#STRUCTURE NAME          ENERGY                                    STOICHIOMETRY

structure_5     -30020.4779808388                                  O_4:Mg_4
structure_2     -30019.6765551307                                  O_4:Mg_4
structure_1     -30019.1803159682                                  Mg_4:O_4
structure_7     -30019.1189970758                                  O_4:Mg_4
structure_6     -30016.4166684199                                  Mg_4:O_4
structure_8     -30016.3118193129                                  O_4:Mg_4
structure_3     -30016.2831888422                                  O_4:Mg_4
structure_4     -30016.0981433278                                  O_4:Mg_4
-------------------------------------------------------------------------
Max_Energy = -30016.0981433278
Min_Energy = -30020.4779808388
Avg_Energy = -30017.9454586145
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/pre-run-results/Mg4O4$ ls structures
structure_1  structure_2  structure_3  structure_4  structure_5  structure_6  structure_7  structure_8
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/pre-run-results/Mg4O4$
```

GM structure

- To plot the hierarchy of isomers keeping GM as 0 run the script ploy.py
- ./plot.py progress/getot.datxxxxxx



Task-2: Why to switch on cascade?

- It's well known the limitation of local/semi-local functional of DFT. There are many systems (e.g. highly oxidised clusters of MgO), where PBE energetics are wrong w. r. t more advanced functionals (e.g. PBE0). Therefore, calculating fitness function using PBE energetics (that we do in conventional GA), becomes very wrong as this gives selection priority to those isomers that are favoured by PBE but totally unfavoured by more advanced functionals. This affects the PES-scanning to get the correct results in determining the GM, and an accurate hierarchy of different isomers.
- In this tutorial, we show one such example and how to circumvent such issues by introducing "cascade" feature of the code. With switching on the "cascade" feature, a local optimisation is done still at the PBE-light level, but that energy is never used to calculate the fitness function. A further step is performed (as requested in the "cascade" keyword as in user_input.in) that takes care of the accurate energetics to be given to calculate the fitness function. For more details we refer to see the manual and our recent publications. [1,2]
- Note: These calculations are expensive and can not be done in the local machines. Thus we recommend to analyse only the pre-run-results using the experience from previous tutorial.

Directory → AIMS-workshop/pre-run-results/Mg2O8

We have three directories to compare:

- GA → performed normal GA but no cascade
- cascade1 → performed cascade GA with PBE-tight
- cascade2 → performed cascade GA with PBE0-tight

[1] "Efficient and accurate *ab initio* schemes for finding thermodynamically stable and metastable atomic structures: Benchmark of cascade genetic algorithms": S. Bhattacharya *et. al.* [to be published 2014]

[2] "Stability and Metastability of Clusters in a Reactive Atmosphere: Theoretical Evidence for Unexpected Stoichiometries of Mg_MO_X": S. Bhattacharya *et. al.* PRL **111**, 135501 (2013)

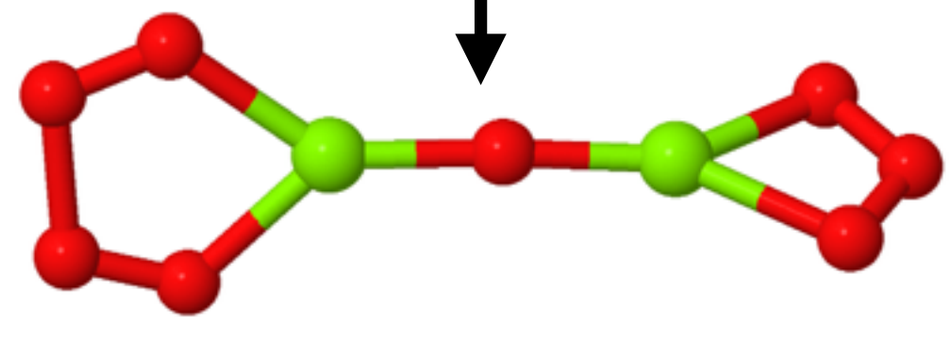
Task-2: Why to switch on cascade?

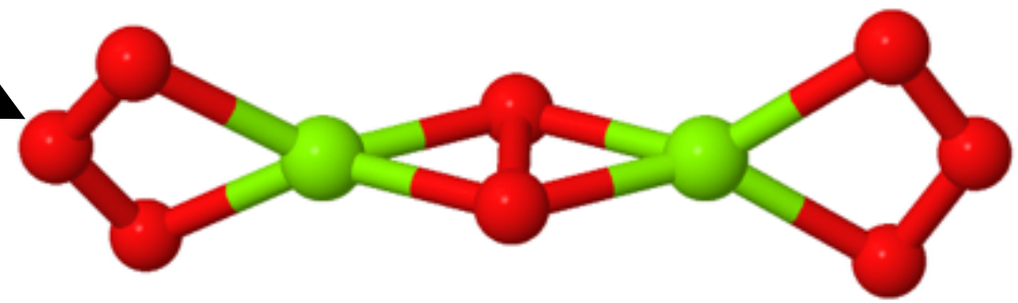
Directory → AIMS-workshop/pre-run-results/Mg2O8/GA

```
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/pre-run-results/Mg2O8/GA$ cat progress/getot.dat864378 | more
#STRUCTURE NAME          ENERGY                                          STOICHIOMETRY

structure_15    -27289.6928732087                                Mg_2:O_8
structure_4     -27289.6871927603                                Mg_2:O_8
structure_22    -27289.637395995                                 Mg_2:O_8
structure_36    -27289.6171771897                                Mg_2:O_8
structure_19    -27289.603605041                                 Mg_2:O_8
structure_5     -27289.5268602086                                Mg_2:O_8
structure_42    -27289.485745731                                 Mg_2:O_8
structure_33    -27289.4448552399                                Mg_2:O_8
structure_11    -27289.4372975139                                Mg_2:O_8
structure_39    -27289.4333105013                                Mg_2:O_8
structure_49    -27289.3580379523                                Mg_2:O_8
structure_32    -27289.3545390086                                Mg_2:O_8
structure_14    -27289.3540606194                                Mg_2:O_8
structure_59    -27289.3524092934                                Mg_2:O_8
structure_28    -27289.3500823892                                Mg_2:O_8
structure_40    -27289.3429699239                                Mg_2:O_8
structure_12    -27289.3416842073                                Mg_2:O_8
structure_25    -27289.3328255649                                Mg_2:O_8
structure_10    -27289.2882737957                                Mg_2:O_8
structure_43    -27289.2691165608                                Mg_2:O_8
structure_18    -27289.2684792269                                Mg_2:O_8
structure_56    -27289.2631147956                                Mg_2:O_8
structure_46    -27289.254886303                                 Mg_2:O_8
structure_61    -27289.2485801047                                Mg_2:O_8
structure_23    -27289.2295444077                                Mg_2:O_8
structure_2     -27289.2069619175                                Mg_2:O_8
structure_47    -27289.1935920688                                Mg_2:O_8
structure_29    -27289.159796125                                 Mg_2:O_8
structure_45    -27289.039755102                                 Mg_2:O_8
structure_50    -27288.954190863                                 Mg_2:O_8
```

GM structure (PBE-light)





Task-2: Why to switch on cascade?

Directory → AIMS-workshop/pre-run-results/Mg2O8/cascade1

✺ Here we dont use the PBE-light energetics to calculate fitness function and we ask the code to perform one additional single point energy calculation at PBE-tight settings (after finishing the local optimisation at the PBE-light level). And the later energy (i.e., PBE-tight) is used to calculate the fitness function.

```
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/pre-run-results/Mg2O8/cascade1$ ls -t progress/getot.dat*
progress/getot.dat189964  progress/getot.dat825111
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/pre-run-results/Mg2O8/cascade1$ more user_input.in
USER SPECIFIC SETTINGS
-----------------------------------------------------------------------
General Settings:
surface/cluster:                        cluster
number_of_children                      2
number_of_structures                    100
aims.out:keep/delete                    keep
initial_moment:hund/custom              hund
control.in_species:default/custom       custom
minimum_bond_length                     0.8
existance_minimum                       0.05
-----------------------------------------------------------------------
BIN_DIRECTORY                           "/u/saswata/fhi-aims/bin"
aims_file                               "aims.020714.scalapack.mpi.x"
-----------------------------------------------------------------------
Cascade Settings:
cascade_start                           1
cascade_num_of_steps                    1
energy_structure                        tight
search_pattern                          0
cascade_window                          0
```

This calculation is more expensive than the previous example without cascade. Thus we have employed parallel run by introducing one more replica.

Cascade settings are switched on.
Note: Some extra files for settings & basis sets named as "settings_tight.dat", "Mg_tight", "O_tight" are also provided in the "control" directory.

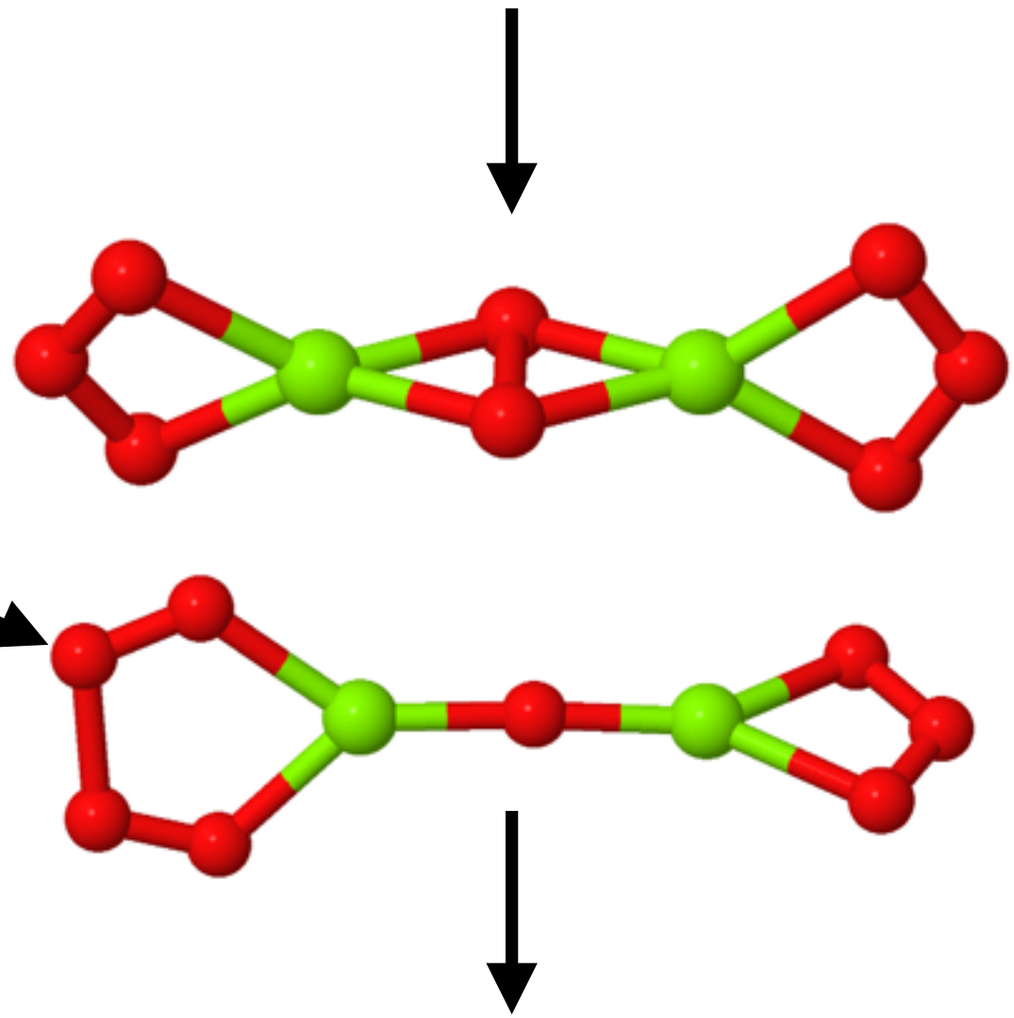
Task-2: Why to switch on cascade?

Directory → AIMS-workshop/pre-run-results/Mg2O8/cascade1

```
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/pre-run-results/Mg2O8/cascade1$ cat progress/getot.dat189964 | more
#STRUCTURE NAME          ENERGY                                                STOICHIOMETRY

structure_4     -27290.7872608886                                    Mg_2:O_8
structure_3     -27290.7719859084                                    Mg_2:O_8
structure_86    -27290.701895717                                     Mg_2:O_8
structure_12    -27290.6790161373                                    Mg_2:O_8
structure_20    -27290.6781484207                                    Mg_2:O_8
structure_43    -27290.6411244382                                    Mg_2:O_8
structure_30    -27290.5617763104                                    Mg_2:O_8
structure_58    -27290.5520315366                                    Mg_2:O_8
structure_28    -27290.5367426399                                    Mg_2:O_8
structure_23    -27290.4544108814                                    Mg_2:O_8
structure_90    -27290.4503146455                                    Mg_2:O_8
structure_9     -27290.4500264413                                    Mg_2:O_8
structure_21    -27290.4487119794                                    Mg_2:O_8
structure_27    -27290.433796813                                     Mg_2:O_8
structure_76    -27290.4304135242                                    Mg_2:O_8
structure_38    -27290.4286252536                                    Mg_2:O_8
structure_52    -27290.420805745                                     Mg_2:O_8
structure_62    -27290.4182489515                                    Mg_2:O_8
structure_34    -27290.4029591525                                    Mg_2:O_8
structure_14    -27290.4006834899                                    Mg_2:O_8
structure_11    -27290.3741692208                                    Mg_2:O_8
structure_7     -27290.368975194                                     Mg_2:O_8
structure_73    -27290.3593021062                                    Mg_2:O_8
structure_16    -27290.3431920166                                    Mg_2:O_8
structure_51    -27290.3327083886                                    Mg_2:O_8
structure_31    -27290.3295145264                                    Mg_2:O_8
structure_67    -27290.3177372789                                    Mg_2:O_8
structure_15    -27290.2985217248                                    Mg_2:O_8
structure_70    -27290.2886310843                                    Mg_2:O_8
structure_19    -27290.2828546073                                    Mg_2:O_8
structure_36    -27290.2592258211                                    Mg_2:O_8
structure_59    -27290.2498177306                                    Mg_2:O_8
structure_79    -27290.2476679924                                    Mg_2:O_8
```

← GM structure (PBE-tight)



Hierarchy is reversed after introducing Cascade at PBE-tight settings!

What happens if we introduce hybrid functional (e.g PBE0-tight) instead of PBE-tight at the cascade level ?

Task-2: Why to switch on cascade?

Directory → AIMS-workshop/pre-run-results/Mg2O8/cascade2

✺ Here we will introduce hybrid functional's energetics (e.g PBE0-tight) to calculate the fitness function. Therefore, in the cascade level, we have to ask single point total energy calculations using PBE0-tight settings. This is the most expensive run out of these three examples. But since we know PBE0-tight energetics are in very good agreement with higher level DFT beyond methods (e.g. rPT2@PBE), this is expected to be the most accurate scanning of this PES.

```
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/pre-run-results/Mg2O8/cascade2$ ls -t progress/getot.dat*
progress/getot.dat445698  progress/getot.dat919573  progress/getot.dat556122  progress/getot.dat203307
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/pre-run-results/Mg2O8/cascade2$ more user_input.in
USER SPECIFIC SETTINGS
-----------------------------------------------------------------------
General Settings:
surface/cluster:                        cluster
number_of_children                      4
number_of_structures                    100
aims.out:keep/delete                    keep
initial_moment:hund/custom              hund
control.in_species:default/custom       custom
minimum_bond_length                     0.8
existance_minimum                       0.05
-----------------------------------------------------------------------
BIN_DIRECTORY                           "/u/saswata/fhi-aims/bin"
aims_file                               "aims.020714.scalapack.mpi.x"
-----------------------------------------------------------------------
Cascade Settings:
cascade_start                           1
cascade_num_of_steps                    1
energy_structure                        PBE0
search_pattern                          0
cascade_window                          0
```

This calculation is more expensive than the previous example of cascade with PBE-tight. Thus we have employed parallel run by introducing two more replicas.

Cascade settings are switched on.
Note: Here the keyword is used as "PBE0". Thus "settings_PBE0", "Mg_PBE0" & "O_PBE0" are provided in the control directory.

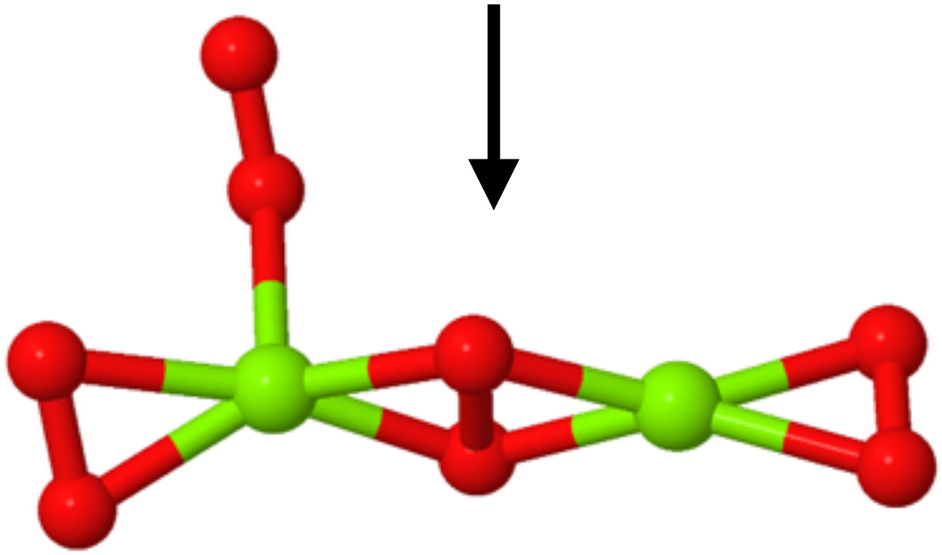
Task-2: Why to switch on cascade?

Directory → AIMS-workshop/pre-run-results/Mg2O8/cascade2

```
bhattacharya@theobook81:~/Codes/GA/AIMS-workshop/pre-run-results/Mg2O8/cascade2$ cat progress/getot.dat445698 | more
#STRUCTURE NAME           ENERGY                                                STOICHIOMETRY

structure_63     -27289.328608731                                  Mg_2:O_8
structure_3      -27289.30853679                          Mg_2:O_8
structure_4      -27289.2823122641                                 Mg_2:O_8
structure_57     -27289.256600264                                  Mg_2:O_8
structure_99     -27289.2289164582                                 Mg_2:O_8
structure_14     -27289.1935686891                                 Mg_2:O_8
structure_100    -27289.1484384294                                 Mg_2:O_8
structure_39     -27289.1375961549                                 Mg_2:O_8
structure_40     -27289.1319882991                                 Mg_2:O_8
structure_50     -27289.1232258254                                 Mg_2:O_8
structure_102    -27289.1216921172                                 Mg_2:O_8
structure_65     -27289.1171926929                                 Mg_2:O_8
structure_91     -27289.1085283347                                 Mg_2:O_8
structure_62     -27289.1067789682                                 Mg_2:O_8
structure_9      -27289.0944667451                                 Mg_2:O_8
structure_80     -27289.0773383517                                 Mg_2:O_8
structure_88     -27289.0549053147                                 Mg_2:O_8
structure_93     -27289.0462879908                                 Mg_2:O_8
structure_21     -27289.0294757398                                 Mg_2:O_8
structure_77     -27289.0285119502                                 Mg_2:O_8
structure_96     -27289.0229372898                                 Mg_2:O_8
structure_92     -27289.0077980291                                 Mg_2:O_8
structure_19     -27288.9969184036                                 Mg_2:O_8
structure_68     -27288.9586228338                                 Mg_2:O_8
structure_94     -27288.9409678739                                 Mg_2:O_8
structure_64     -27288.9386256536                                 Mg_2:O_8
structure_38     -27288.9259921522                                 Mg_2:O_8
structure_97     -27288.9171551086                                 Mg_2:O_8
structure_24     -27288.9026145908                                 Mg_2:O_8
structure_20     -27288.8996315037                                 Mg_2:O_8
structure_23     -27288.8859786706                                 Mg_2:O_8
structure_15     -27288.8039014115                                 Mg_2:O_8
structure_31     -27288.7776579972                                 Mg_2:O_8
structure_12     -27288.7295621244                                 Mg_2:O_8
```

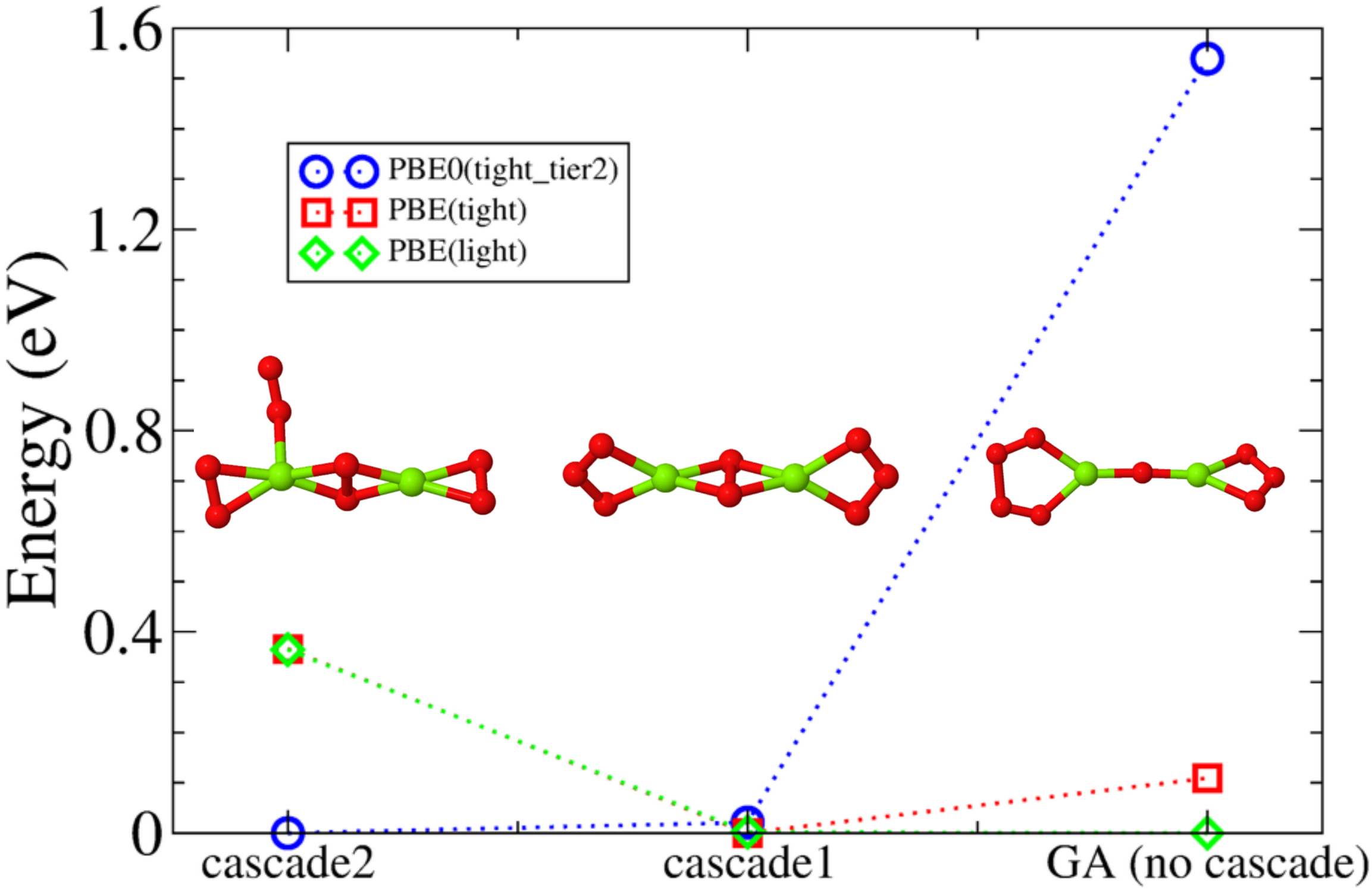
← GM structure (PBE0-tight)



GM is totally changed w. r. t PBE (light/tight).

This structure was 18th and 15th in the energy hierarchy of the previous two GA runs (i.e. the one with and the one without cascade respectively).

Task-2: Why to switch on cascade?



GM from different cascade GA run