ASCEC-V02 (very short) Instructions

Albeiro Restrepo

Grupo de Química–Física Teórica, Instituto de Química,

Universidad de Antioquia, AA 1226 Medellín, Colombia.

August 21, 2010

Contents

1 Distribution

The program is sent in a single file named ascec.tar suited for Linux operative systems. To unpack it, make a directory to contain the ascec files and copy the ascec.tar file in there, then issue tar xvf ascec.tar. Successful unpacking will produce the following files and directories:

type	name	description	size
file	c_ascec.sh	compiling script	143
file	readme.pdf	this file	53695
directory	papers	relevant papers	4096
directory	p_ascec	source code	4096
directory	examples	worked out examples	4096

The source code, input, and output files were written with Spanish in mind, I translated the information shown in the input and output files, as a consequence, some minor glitches will pop up during execution time, but those are format errors which should not affect the calculations.

2 Capabilities

Please understand that this program is always in a development state. The version I am attaching can perform the annealing on any system containing up to a total of 10,000 atoms and a maximum of 1,000 molecules. All atoms in the periodic can be treated, provided that Gaussian can calculate the energy of the system with the chosen level of theory.

3 Compilation and execution

To compile the program issue sh c_ascec.sh. A successful compilation will produce the ascec-v02 executable file. The compiling script will use gfortran by default, if your machine has a different compiler, change it accordingly (second line of the c_ascec.sh file). If the program is to be available for all users on a given computer, copy the executable to either the /bin or the /usr/bin directories. For execution, issue ascec-v02 < inputfile > outputfile &. For the ASCEC program to work, you will need to have the Gaussian 03 program installed and run under the g03 command.

3.1 Parallel computing

The ASCEC program is not yet written to work in parallel. However, the gaussian program, which is called to calculate the energy of the randomly generated configurations is capable of parallel processing. To activate the parallel option in gaussian,

before compilation (i) un-comment line 20 of the function_E.f file on the p_ascec directory and (ii) give de desired number of processors in the same line.

4 Input

I have provided two typical input files in the examples directory: w5-10.in and w5-a.in. Both files consider the water pentamer with all the molecules initially at the center of a cubic box of 7 Å in length.

4.1 w5-10.in

Running ASCEC on this file will produce 10 random configurations for the water pentamer. This test is very useful to (i) verify that the compilation was successful; (ii) make sure that the random number generator is working properly, and (iii) fix any problems on the input file. The mto_24032010.xyz file was produced after issuing ascec-v02 < w5-10.in > w5-10.out on my computer. The 10 configurations can be visualized using the molekel program. Details of the calculations are on the w5-10.out file.

4.2 w5-a.in

Running ASCEC on this file will produce candidate structures for the water pentamer to be further optimized and refined. The result_29562010.xyz, rless_29562010.out, tvse_29562010.dat files were produced after issuing ascec-v02 < w5-a.in > w5-a.out & on my computer. Details of the calculations are on the w5-a.out file.

4.3 Input dissection

The program is written in fortran, therefore, **the format of the input file must be kept**, this is a critical issue: integers must be differentiated from real numbers and the number of spaces available for each field must be retained. The only difference between the w5-10.in and w5-a.in files is in the first line. I will explain the input file using w5-a.in; text after exclamation marks (!) is included as commentaries to help explain the input:

Line 1: 1 10 !0: Randomly generate 10 configurations; 1: Do the annealing
Two options here for the first field, 0 or 1, both integers. The second field is meaningful only if 0 is chosen in the first field, however the second field must remain even if 1 is selected.

Line 2: 1 g03 !1:Gaussian03, 2:MSINDO1.1.opt
Only 1 (integer) is activated for this version of ASCEC, which means that Gaussian is to be used to calculate the energy. g03 is the command with which gaussian is invoked.

Line 3: 7. !Cube's length
Cube's length in angstroms.

Line 4: 2 !Quenching route 1: linear, 2: geometrical
Selecting 1 (integer) will use a linear quenching route. If 2 is chosen, a geometrical quenching route will be used. Option 1 will make line 5 relevant, while option 2 will make line 6 relevant; both lines 5 and 6 must be present regardless of the selected option.

Line 5. 100. 2. 50 !Linear To/dT/nT
For a linear quenching route: the initial temperature (100. a real number), the constant decrease in temperature (2. a real number), and the total number of temperatures (50 an integer). All temperatures in Kelvins.

Line 6: 500. 5. 100 !Geometrical To/%dism/nT
For a geometrical quenching route: the initial temperature (500. a real number), the constant % decrease in temperature (5. a real number), and the total number of temperatures (100 an integer). All temperatures in Kelvins.

Line 7: 3000 !MaxCycle
Maximum number of energy evaluations at a given temperature; the number is reduced every time it is reached. See the w4.pdf and the lin.pdf papers for details.

line 8: 2 !Number of elements
Number of different elements contained in the system (an integer). For the water pentamer there is H and O, 2 elements.

Line 9: 1 10 !Atomic number of the nucleus, number of nuclei
Line 10: 8 5 !Atomic number of the nucleus, number of nuclei
There must be one line for each element in the system, hence we have lines 9, 10. For the water pentamer we have 10 H atoms, therefore numbers 1 10 (integers) in line 9; there are also 5 O atoms, therefore numbers 8 5 (integers) in line 10.

Line 11: pm3 zdo !Hamiltonian / Basis set
The Hamiltonian and basis set for the energy calculations during the annealing.

Line 12: 0 1 !Charge / Spin multiplicity
Total charge and multiplicity of the entire system (integers).

Line 13: Blank line. Must be there.

Line 14: 5 !Number of molecules (1,2,...)
Total number of molecules (integer). If atomic clusters are studied, then each atom must be considered a molecule. If atoms interacting with molecules are considered, then each atom not being part of a molecule must be considered a molecule.

Line 15: Blank line. Must be there.

Line 16: w1 3
The first field (w1) is a nickname for the individual molecule. The second field (3) is the number of atoms in the molecule (integer).

Lines 17, 18, 19: Contain the atomic number (integer) and the x, y, z, coordinates of the 3 atoms contained in w1. For this particular case, I drew the water molecule in GaussView and then produced the cartesian coordinates.

Lines 20 – 35: This lines identify the remaining water molecules in exactly the same fashion as Lines 16 – 19. No blank lines in between.

Line 36: Blank line. Must be there.

Line 37: 1.0 1. !Max. displacement (A); Max. rotation angle (radians)
Maximum displacement in angstroms for each mass center on a given random motion (1.0 a real number), and maximum rotation in radians of the principal axes of each molecule in a given random motion (1. a real number). If an atomic cluster is studied, the second field must be set to 0.

5 Output

A successful ASCEC run with the annealing option (1 in the first field, firs line) will generate four files, in this case:

5.1 w5-a.out

This file contains all the details of the annealing. To make sure, that the program terminated correctly, issue tail w5-a.out, the ** Normal annealing termination ** message will appear.

5.2 rless_29562010.out

This file contains cartesian coordinates for the lowest energy configuration.

5.3 tvse_29562010.dat

This file contains three columns: the step number, the temperature and the energy. Only information for those accepted structures is included. The information in this file will produce plots like Figures 2 and 3 of the w4.pdf paper and Figures 1 and 2 of the lin.pdf paper.

5.4 result_29562010.xyz

This file contains the cartesian coordinates of all accepted structures (167 in this particular case) in a format suitable for visualization with molekel. I am including ascec-extract.sh, a processing script, that performs the following tasks: (*i*) extracts cartesian coordinates for each candidate structure, (*ii*) creates input files for gaussian optimization and frequency calculations, (*iii*) creates another script, file launch_g.sh, which is actually a queue to send all the gaussian calculations one after the other; in this way, only one batch of calculations is to be monitored.

Very important notes

1. Usually the gaussian optimizations will require two rounds to find the local minima, that is, the first time the optimization will go up to the maximum number of steps allowed in gaussian and crash, then reading the last geometry from the checkpoint file (geom=check keyword) would generally be sufficient. This is because of progressive deterioration of the hessian matrix during the gaussian optimization. Some extreme cases will require a third optimization round; this is not so bad because after visually inspecting the structures, only those which will lead to different geometries will be recalculated.

2. The ascec-extract.sh script will need the ksh shell. To run it, issue ksh ascec-extract.sh result_29562010.xyz b3lyp "6–311++g**" 0 1 (quotation marks are mandatory). This will produce input files for optimization and frequency calculations at the b3lyp level in conjunction with the 6–311++g** basis set.

3. Change the number of requested processors in line 14 of the ascec-extract.sh script as needed.