AMADEUS

installation/user guide

Ab initio materials design using conformational space annealing

In-Ho Lee (ihlee@kriss.re.kr)

2016 FEBRUARY 18

Korea Research Institute of Standards and Science

AMADEUS installation/user guide

AMADEUS (**A**b initio **MA**terials **DE**sign **U**sing conformational **S**pace annealing) is a combined computational approach of conformational space annealing (CSA) for global optimization and first-principles electronic structure calculations. The enthalpy minimization and the analysis of electronic properties were performed, within the framework of density functional theory, via **VASP** code (http://www.vasp.at/). The AMADEUS package is distributed under the BSD license.

AMADEUS installation guide

AMADEUS (**csa_vasp.x**) is written in FORTRAN 90. This installation guide is designed for the compiler **ifort** (Intel FORTRAN compiler) on the Linux system. A gzipped-tar file **amadeus????????.tar.gz** will be the starting object of this AMADEUS installation guide.

1. $gzip –d amadeus????????.tar.gz
2. $tar xvf amadeus.tar
3. $cd csa_vasp
4. $cd source0
5. $vi makefile
6. $make
7. $ls –ltra ../

8. $ls CSA_KING.pbs CSA_SOLDIER.pbs csa.in

9. $ls INCAR_rlx INCAR_rlxall INCAR_bs

10. The step 2 will provide a directory with the name of **csa_vasp**. In step 5, you have to have a configuration setup. Through the command shown in step 7, an execution file **csa_vasp.x** will be present.

AMADEUS user guide

For your specific study, you should have a working directory, e.g., **/home/ihlee/Si14**. This full path name will be an input parameter in the file **csa.in**. **POTCAR** file cannot be distributed without the permission. Thus, **POTCAR** file is not included in the present tar file.

1. In your working directory, e.g., **/home/ihlee/Si14**, you may have to place **VASP** input files such as **POTCAR**, **INCAR_rlx**, **INCAR_rlxall**, and **INCAR_bs**. Edit files **INCAR_rlx**, **INCAR_rlxall**, and **INCAR_bs**. A **POTCAR** should be prepared before the study. There are real-life examples of **AMADUES**, **exam0**, **exam1**, and exam2. It is worthwhile to experiment with these example setups for variables in *csa.in*. Keep in mind that different problems require different settings.

2. Two files **CSA_KING.pbs** and **CSA_SOLDIER.pbs** are waiting to be edited. The file **CSA_SOLDIER.pbs** contains a script for **VASP** run. On the other hand, the file **CSA_KING.pbs** contains a script for the **csa_vasp.x** run. As you expect, **cas_vasp.x** will spawn many **VASP** runs

in a set of independent working directories. Note that **csa_vasp.x** requires a single CPU. **CSA_SOLDIER.pbs** is nothing but a script for **VASP** run. Thus, it is related to many CPUs tightly connected. Actually, **csa_vasp.x** generates many working directories and submits many **VASP** runs in the mode of parallel in parallel.

3. Before you run the **csa_vasp.x** program, you need to generate a directory named by **deposit**.

$mkdir deposit

4. In case you are not using PBS batch file **CSA_KING.pbs**, for submitting **csa_vasp.x**, you may use following Linux command at the prompt.

$nohup nice ~/csa_vasp/csa_vasp.x < csa.in &> csa.out &

5. For a run of **csa_vasp.x** in a PBS batch mode, you will have the following command on Linux system. **CSA_KING.pbs** should be a long-time queue. In addition, **CSA_KING.pbs** should take a single CPU.

$qsub CSA_KING.pbs

6. Monitor the job process, files **csa.out** and **fort.1**. Thus, in general, input files for **csa_vasp.x** contains **csa.in**, **POTCAR**, **INCAR_rlx**, **INCAR_rlxall**, **INCAR_bs**, and **fort.1**.

7. Only for an iterative calculation of **csa_vasp.x**, file **fort.1** will be an additional input file. You have to remove **fort.1** for a new calculation.

8. An input file **csa.in** is key input file for the **csa_vasp.x** program.

- If we call this file **csa.in**, we can run the **csa_vasp.x** by putting it as an input file, and then just place it like so:

  $nohup nice /home/ihlee/csa_vap/csa_vasp.x<csa.in &> csa.out &

- Note that the **extpress** is a dummy variable. It should be expressed in the input files (**INCAR_rlx**, **INCAR_rlxall**, and **INCAR_bs**), if it is necessary. E.g., external pressure (kbar unit).

- Check the file **CSA_SOLDIER.pbs**. It remains to be seen whether or not the contents can be put into practice. It should be compatible with parameter **iobj** in file **csa.in**. **Iobj=0** stands for an enthalpy minimization for a given external pressure.

- **voltol**, this value is in percent unit, tolerance in volume, **cellvol0** (in unit of $Å^3$) initial volume setting, which is different from the relaxed one, in general. **cellvol0** is a reference volume in unit of $Å^3$.

- **lvcs =.true.** means variable cell shape calculations. In this case **voltol** is an important input.

- **cmatrix(3,3)** will be given in Å unit. This is important input in the case of **lvcs=.false.**

- **ndirectory**, this parameter is useful for limiting the usage of the resources, as an intentional action. 1 <= **ndirectory** <= **npop**

- **lnewjob**, this variable can actually be set to **.false.** any cases. You can explicitly remove the file **fort.1** by using the command "rm

fort.1" for a new job.

- **sigmamatrix(:,:)** defines the minimum distance in Å unit between atomic species, before the **VASP** calculations.
- **npop** means the number of populations in CSA.
- **nmate** : **npop** x **nmate** will be the number of crossovers.
- **npert** : **nopo** x **npert** will be the number of mutations.
- **drate** : CSA annealing schedule parameter.
- **nevol** : the number of CSA stages
- **idiff = 3** will be a bond length histogram based discrimination.
- **amp** : a reference amplitude in Å unit for a mutation.
- **iseed10, iseed20** : random number seeds.
- As this example shows, input parameters may not be broken after lines, and a comment may be inserted at the end of each line.
- To check the latest calculations.

```
ls -ltr 00??/stdout.log
ls -ltr 00??/OUTCAR
```

- $/home/ihlee/csa_vasp/readfort1.x < csa.in will provide many POSCAR files such as POSCAR_0001, POSCAR_0002, ... These file names are compatible format to **VESTA** program (http://jp-minerals.org/vesta/en/), a crystal viewer.

- Availability of each directory, the quality of being at hand when needed, can be monitored by using the command.

```
tail -n1 ./00??/STATUS
```

- check list

fort.1 : cp, rm

csa.in : vi

INCAR_rlx : vi

INCAR_rlxall : vi

INCAR_bs : vi

POTCAR : cp

CSA_KING.pbs : vi

CSA_SODIER.pbs : vi

deposit : mkdir deposit

A set of useful commands for job hold, release, and cancel in the PBS system.

```
kill -9 29861; rm 29861; qstat -a|grep SiC18 |awk '{split($0,a,"."); print "qdel ", a[1]}' |sh –x

qstat -a|grep SiC18 |awk '{split($0,a,"."); print "qhold ", a[1]}' |sh –x

qstat -a|grep SiC18 |awk '{split($0,a,"."); print "qrls ", a[1]}' |sh –x
```

Here, we assume that job id is 29861 and VASP jobnames all contain characters SiC18.

9. To start a new study, in your working directory, you will have the following commands.

$mkdir deposit

$ls INCAR_rlx INCAR_rlxall INCAR_bs POTCAR CSA_KING.pbs CSA_SOLDIER.pbs csa.in deposit

$qsub CSA_KING.pbs

In your iterative calculation, you will have a file named **fort.1**.

10. If you have **ndirectory=50** (as an input parameter in **csa.in**) and each **VASP** run takes 20 CPUs through the **CSA_SOLDIER.pbs** then you can use about 1,000 CPUs simultaneously.

11. At the early stage of the **csa_vasp.x** run, **csa_vasp.x** generates many directories and subsequently copies of five files, **CSA_SOLDIER.pbs**, **INCAR_rlx**, **INCAR_rlxall**, **INCAR_bs**, and **POTCAR**. Thus, each directory has above five files. The last input file, **POSCAR**, trial crystal structure for a **VASP** run, will be provided by the **csa_vasp.x** program as it runs. Whenever a new **POSCAR** file is generated, a new **VASP** run is involved through the command 'qsub CSA_SOLDIER.pbs'. The last job submission command is issued by the **csa_vasp.x** program. In each directory, except either starting stage or finalizing stage, one **VASP** run calls another **VASP** run in a successive mode. How can this happen?

Actually, **csa_vasp.x** program monitors every **VASP** job status through the file **STATUS** placed in each directory. In practice, the last line of the **STATUS** file stands for whether current **VASP** job is running (queued) or finished.

12. For a code analysis, you can generate postscript files with the following command in your source code directory.

$make a2ps

13. In most cases, integer variables start with [i, j, k, l, m, n], while double precision (real*8) variables start with [a-h, o-z] following the old convention.

14. A known issue: It is impossible to build a set of trial unit cells under the unphysically large **sigmamatrix(:,:)** values with a smaller unit cell volume. This results in a status of the computer program for not generating working directories, 0001, 0002, 0003, ... and so on, just after issuing the run command. You have to reconfigure the **sigmamatrix(:,:)** values just along with your current crystal unit cell volume size so that the program can generate reasonable random symmetric crystal structures. The best solution to this problem is to make sure the reconfigured values for the symmetric **sigmamatrix(:,:)** are very close to 1.0 Å, in most cases. You will have to run the program again, after the killing process of the execution.

15. **AMADUES** is a fault-tolerant protocol. **AMADUES** can abandon conformations related to any kind of fault from the **VASP** execution,

including artificial job termination and/or abortion. Actually, **AMADUES** gives up the specific conformation as a high energy/enthalpy (objective function) crystal system. A fault recognition in **AMADUES** is made through the investigation of the files (**stdout.log**) scattered over working directories, in addition, in the case of **STATUS** file ends with a script **DONE**.

16. When using **AMADUES**, you are expected to cite the reference papers in your publications.

17. **AMADUES** is not complete. It is intended to be extended. Both providing feedback and contributing to the source code are the ways to make it better.

System requirements

1. Linux/UNIX operating system
2. PBS like job submission system
3. Intel FORTRAN compiler
4. VASP

History

The author tried to develop the protocol during the application period (2013) for Samsung Science and Technology Foundation

under Grant No.SSTF-BA1401-08.

References

1. J. Lee, I.-H. Lee, and J. Lee, Phys. Rev. Lett. **91**, 080201 (2003). DOI: 10.1103/PhysRevLett.91.080201
2. I.-H. Lee, J. Lee, Y. J. Oh, S. Kim, and K. J. Chang, Phys. Rev. B **90**, 115209 (2014). DOI: 10.1103/PhysRevB.90.115209
3. Y. J. Oh, I.-H. Lee, S. Kim, J. Lee, and K. J. Chang, Sci. Rep. **5**, 18086 (2015). DOI: 10.1038/srep18086(2015)
4. Y. J. Oh, S. Kim, I.-H. Lee, J. Lee, and K. J. Chang, Phys. Rev. B **93**, 085201 (2016). DOI: 10.1103/PhysRevB.93.085201
5. I.-H. Lee, Y. J. Oh, S. Kim, J. Lee, and K. J. Chang, Comput. Phys. Commun. **203**, 110 (2016). DOI: 10.1016/j.cpc.2016.02.011

Disclaimer

on or made in reliance on the information, services, or material on or generated from this protocol. The author takes no responsibility for any damage incurred by the user as a result of using this protocol. No reproduction, republication, and redistribution to third parties of any contents is permitted without written permission of the author.

Glossary

- A **genetic algorithm (GA)** is a search heuristic that mimics the process of natural selection. In a GA, a population of candidate solutions to an optimization problem is evolved toward better solutions.
- A **heuristic** is a technique designed for solving a problem more quickly when classic methods are too slow, or for finding an approximate solution when classic methods fail to find any exact solutions.
- An **evolutionary algorithm** (EA) is a subset of evolutionary computation, a generic population-based metaheuristic optimization algorithm. An EA uses mechanisms inspired by biological evolution, such as mutation, crossover, and selection.
- **Evolutionary algorithms** use a population of individuals, where an individual is referred to as a chromosome.
- A **chromosome** (*chromo-* + *-some*) is a packaged and organized

structure containing most of the DNA of a living organism.

- **Mutation** is a genetic operator used to maintain genetic diversity from one generation of a population of GA chromosomes to the next.
- **Crossover** is a genetic operator used to vary the programming of a chromosome or chromosomes from one generation to the next. **Crossover** and **mutation** are known as the main genetic operators.
- **Selection** is the stage of a GA in which individual genomes are chosen from a population for later breeding. A proportion of the existing population is selected to breed a new generation.
- An **objective function** is either a loss function or its negative, in which case it is to be maximized. The survival strength of an individual is measured using an objective function which reflects the objectives and restraints of the problem to be solved.

Figures

```
   1                                          nspecies
   C                                          symbl(:)
   8                                          nelements(:)
  45.378190     0.000d0   0.040d0             cellvol0(A^3), extpress(Mbar), voltol
   1.0000000    0.0000000   0.0000000         cmatrix(1,1), cmatrix(1,2), cmatrix(1,3)
   0.0000000    1.0000000   0.0000000         cmatrix(2,1), cmatrix(2,2), cmatrix(2,3)
   0.0000000    0.0000000   1.0000000         cmatrix(3,1), cmatrix(3,2), cmatrix(3,3)
   1.08                                       sigmamatrix(1,1), sigmamatrix(1,2), ..... in Angstrom unit
.true.     .true.    0                       lvcs, lpbc, iobj
  50                                          ndirectory
/home/ihlee/csa_vasp/exam07                   cwd
   3     5                                    idiff, nevol
  50     50                                   npop, npop1
  80     20        3                          nmate, npert, nfrac
  3.0     0.990d0                            amp(A), drate
457    1442       .false.                    iseed10, iseed20, lnewjob
+------------------------------------------------------------+
|      read(5,*) nspecies                                    |
|      read(5,*) (symbl(i),i=1,nspecies)                     |
|      read(5,*) (nelements(i),i=1,nspecies)                 |
|      read(5,*) cellvol0,extpress,voltol                    |
|      read(5,*) cmatrix(1,1),cmatrix(1,2),cmatrix(1,3)      |
|      read(5,*) cmatrix(2,1),cmatrix(2,2),cmatrix(2,3)      |
|      read(5,*) cmatrix(3,1),cmatrix(3,2),cmatrix(3,3)      |
|      do i=1,nspecies                                       |
|      read(5,*) (sigmamatrix(i,j),j=1,nspecies)             |
|      enddo                                                 |
|      read(5,*) lvcs,lpbc,iobj                              |
|      read(5,*) ndirectory                                  |
|      read(5,'(a280)') cwd                                  |
|      read(5,*) idiff,nevol0                                |
|      read(5,*) npop0,npop10                                |
|      read(5,*) nmate0,npert0,nfrac                         |
|      read(5,*) amp0,drate0                                 |
|      read(5,*) iseed10,iseed20,lnewjob                     |
+------------------------------------------------------------+
```

Figure 1 A sample input file **csa.in** for **csa_vasp.x** is shown. The number of species (**nspecies**), atomic symbol (**symbl**), and the number of atoms for each species (**nelements**) are placed in the first three lines, respectively.

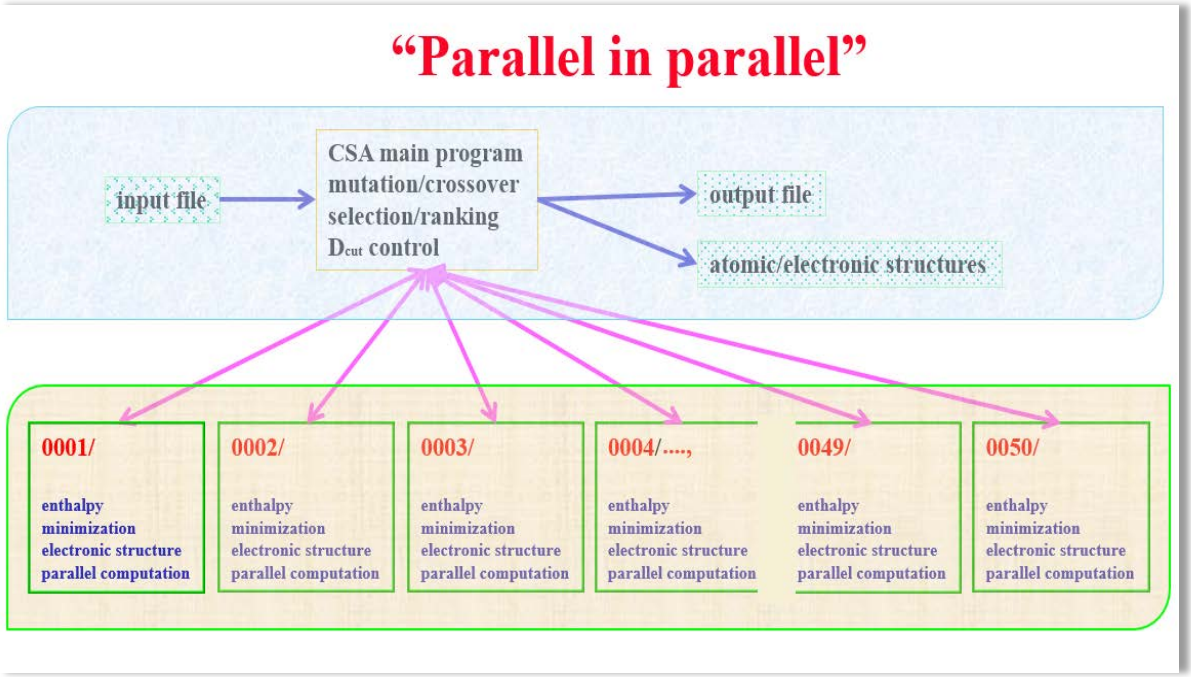


Figure 2 Many parallel **VASP** runs are available in many working directories. The number of configurations was set to e.g., 50 in the population size of CSA.

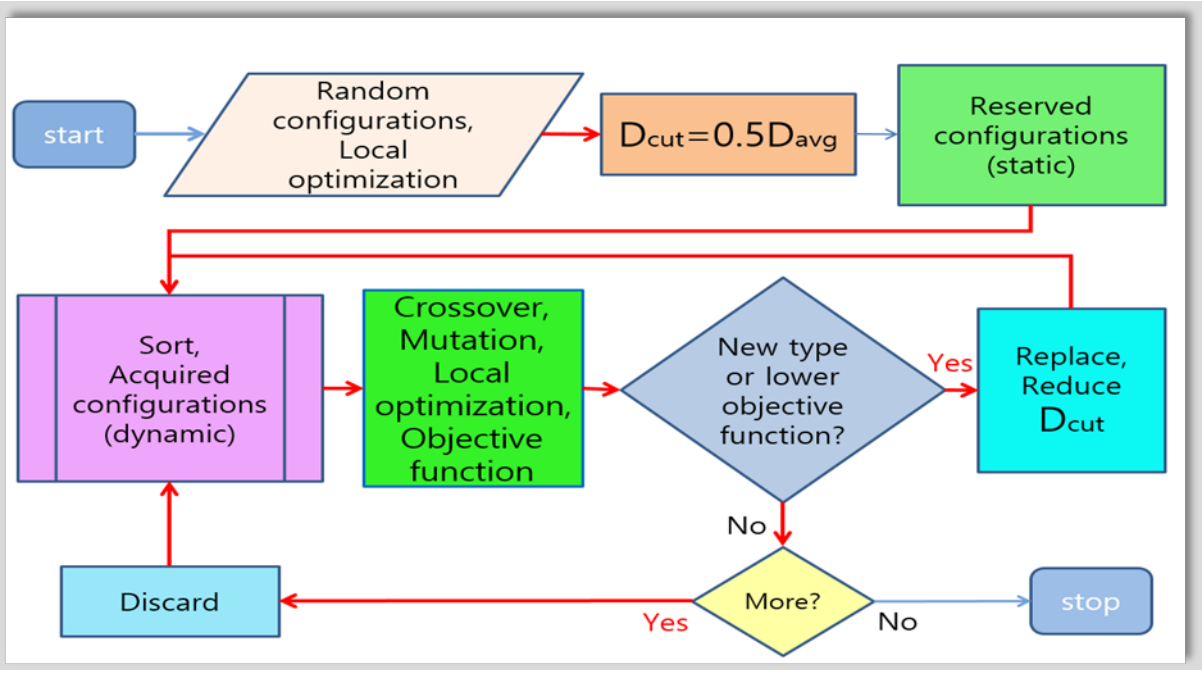


Figure 3 CSA flowchart is shown. A measure 'distance' is defined for a pair of crystal structures. Thus, 'distance' can be compared with current D_{cut} value whenever new

crystal structure is obtained through the local enthalpy minimization process. In **AMADEUS**, the master and slave processes were operated separately, and their communication is executed via disk files. For convenience, **AMADEUS** creates 50 (an arbitrary number) directories to be used repeatedly for first-principles calculations.

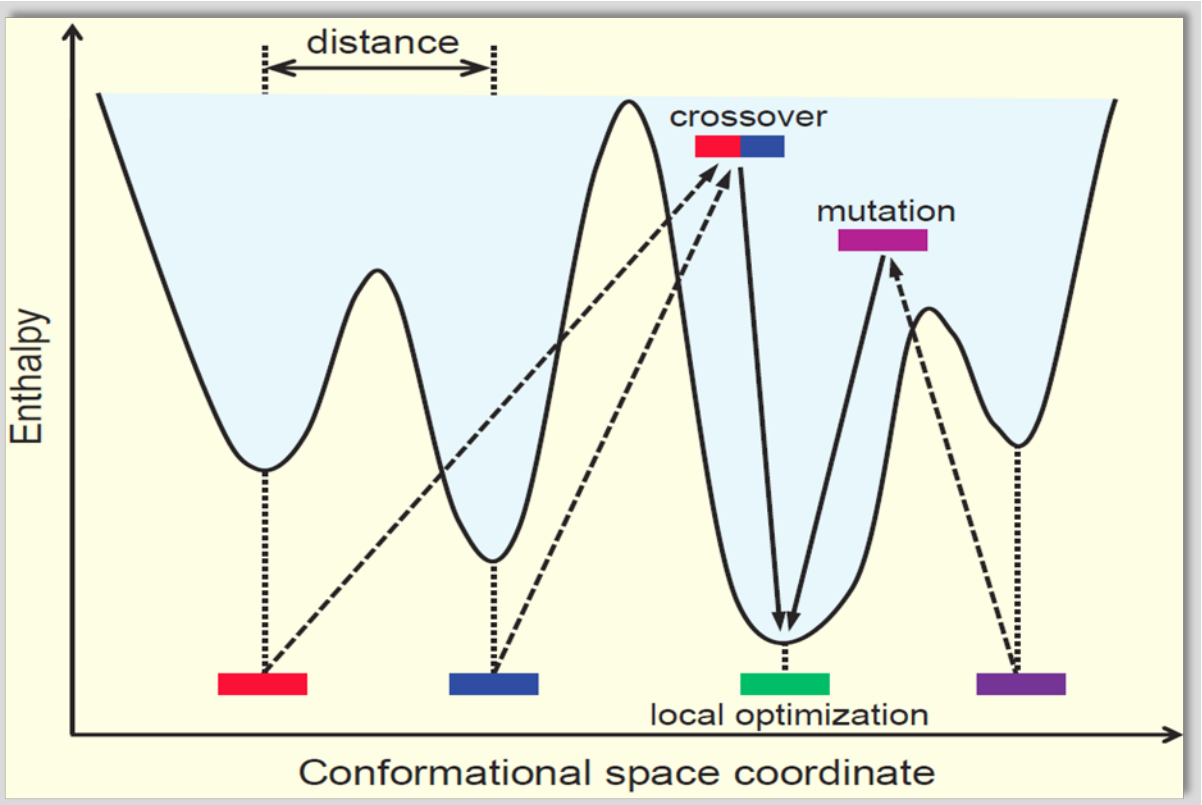


Figure 4 A schematic diagram for the relationship between conformational space and enthalpy. A measure 'distance' is well defined for a pair of crystal structures. Crossover and mutation are the major processes for a new crystal structure generation through local enthalpy minimization process. For a given external pressure, a crystal structure can be determined by a local enthalpy minimization. Stress and force calculations are required to relax cell shape and atomic positions.

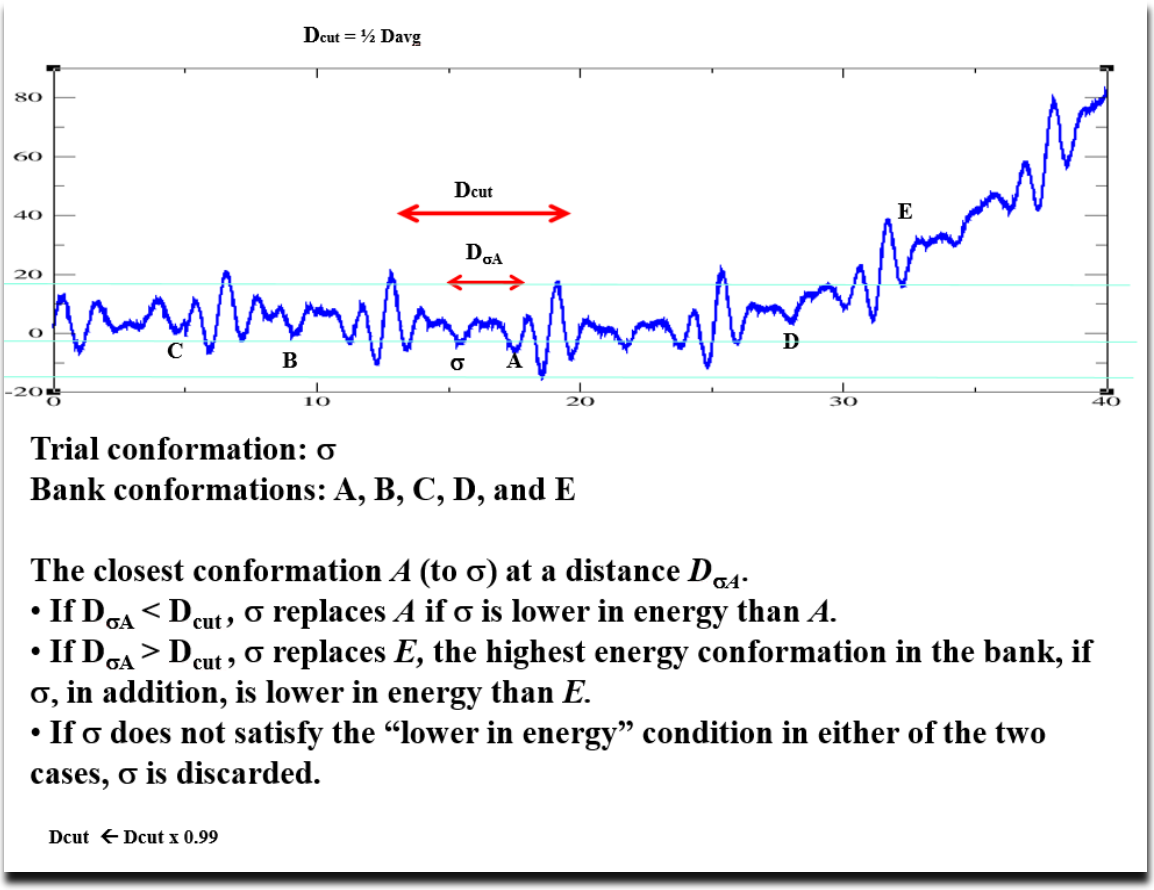


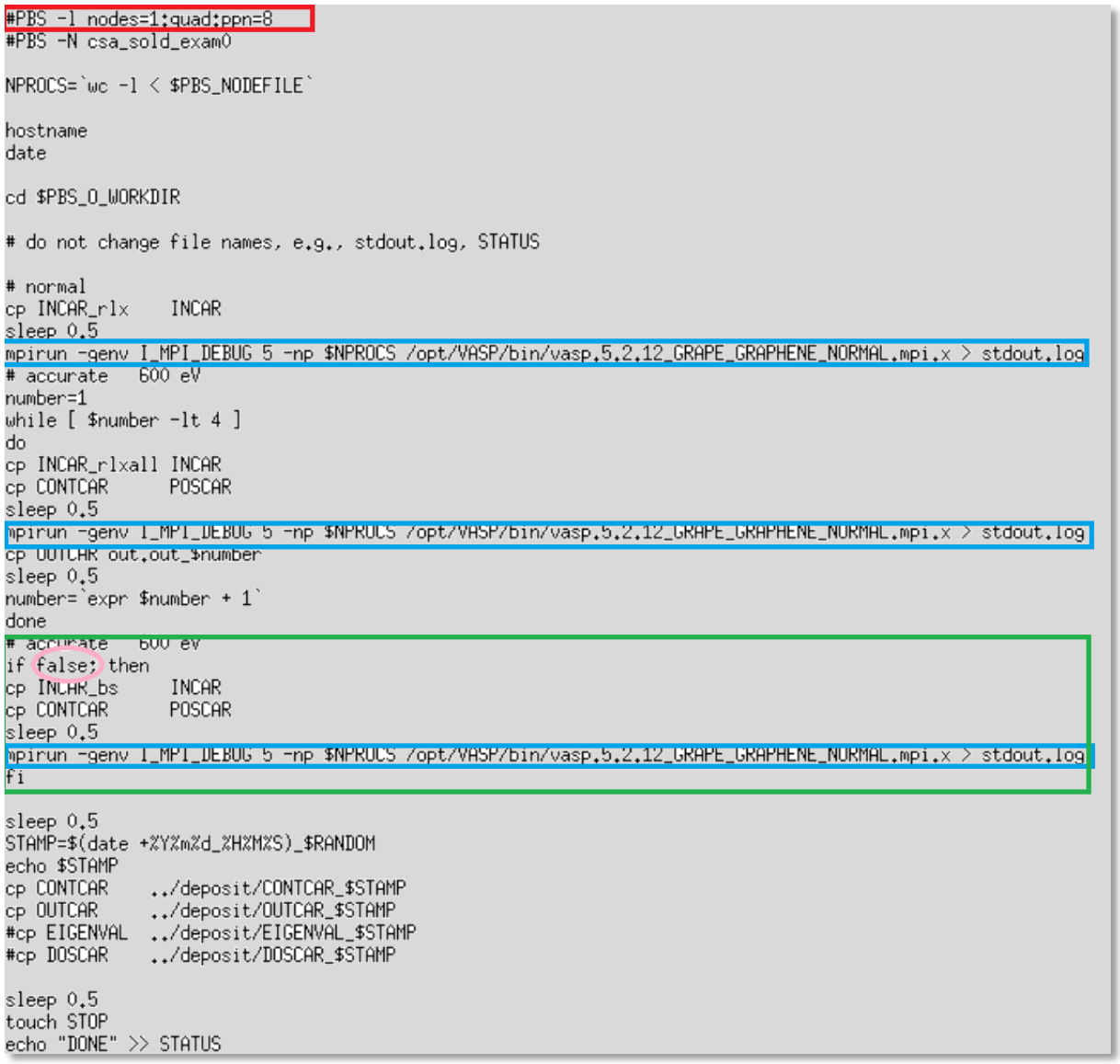
Figure 5 A set of conditions for the replacement operation in the CSA algorithm is shown. A diversity control scheme in the CSA is described.

```
#!/bin/sh
#PBS -l nodes=1:quad:ppn=1
#PBS -N csa_king_exam0

NPROCS=`wc -l < $PBS_NODEFILE`

hostname
date

cd $PBS_O_WORKDIR


#rm fort.?
#rm -rf 0???
#rm -rf /home/ihlee/csa_vasp/exam0/deposit/*
#mkdir  /home/ihlee/csa_vasp/exam0/deposit

# check files INCAR_bs, INCAR_rlx, INCAR_rlxall, POTCAR, csa.in, and CSA_SOLDIER.pbs
# never ending job : kill the PBS job, change STATUS
if [ ! -d deposit ]
then
    mkdir deposit
fi

/home/ihlee/csa_vasp/csa_vasp.x < csa.in &> csa.out

STAMP=$(date +%Y%m%d_%H%M%S)_$RANDOM
cp csa.out   csa.out_$STAMP
cp fort.1     fort.1_$STAMP
```

Figure 6 A sample **CSA_KING.pbs** is shown. **CSA_KING.pbs** must be involved to a single CPU. However, **CSA_SOLDIER.pbs** involved to a set of many CPUs, in general. **CSA_KING.pbs** is closely related to **CSA_SOLDIER.pbs.** An alternative command for 'qsub CSA_KING.pbs' is:

nohup nice /home/ihlee/csa_vasp/csa_vasp.x < csa.in &> csa.out &

Files **INCAR_bs**, **INCAR_rlx**, **INCAR_rlxall**, **csa.in**, **CSA_SOLDIER.pbs**, **POTCAR** must be present in the same directory. Furthermore, file names are fixed. Because they are supposed to be copied to working directories, 0001, 0002, …., .



```
#PBS -l nodes=1:quad:ppn=8
#PBS -N csa_sold_exam0

NPROCS=`wc -l < $PBS_NODEFILE`

hostname
date

cd $PBS_O_WORKDIR

# do not change file names, e.g., stdout.log, STATUS

# normal
cp INCAR_rlx    INCAR
sleep 0.5
mpirun -genv I_MPI_DEBUG 5 -np $NPROCS /opt/VASP/bin/vasp.5.2.12_GRAPE_GRAPHENE_NORMAL.mpi.x > stdout.log
# accurate   600 eV
number=1
while [ $number -lt 4 ]
do
cp INCAR_rlxall INCAR
cp CONTCAR      POSCAR
sleep 0.5
mpirun -genv I_MPI_DEBUG 5 -np $NPROCS /opt/VASP/bin/vasp.5.2.12_GRAPE_GRAPHENE_NORMAL.mpi.x > stdout.log
cp OUTCAR out.out_$number
sleep 0.5
number=`expr $number + 1`
done
# accurate   600 eV
if false; then
cp INCAR_bs     INCAR
cp CONTCAR      POSCAR
sleep 0.5
mpirun -genv I_MPI_DEBUG 5 -np $NPROCS /opt/VASP/bin/vasp.5.2.12_GRAPE_GRAPHENE_NORMAL.mpi.x > stdout.log
fi

sleep 0.5
STAMP=$(date +%Y%m%d_%H%M%S)_$RANDOM
echo $STAMP
cp CONTCAR    ../deposit/CONTCAR_$STAMP
cp OUTCAR     ../deposit/OUTCAR_$STAMP
#cp EIGENVAL  ../deposit/EIGENVAL_$STAMP
#cp DOSCAR    ../deposit/DOSCAR_$STAMP

sleep 0.5
touch STOP
echo "DONE" >> STATUS
```

Figure 7 A sample **CSA_SOLDIER.pbs**. A change false → true is required in case of band-structure calculation. **CSA_SOLDIER.pbs** involved to a set of many CPUs. File **STATUS** must be present so as to **csa_vasp.x** recognizes the status of the electronic structure calculation. For a serial backup of crystal structures, one can use either

"cp CONTCAR /home/ihlee/csa_vasp/exam0/deposit/CONTCAR_$STAMP" or

"cp CONTCAR ../deposit/CONTCAR_$STAMP".

The above code shows the skeleton of a distributed **VASP** calculation, executing atomic position relaxation, lattice parameters relaxation and/or band structure calculation. It includes the code for the calculation parameters and tasks. You should do to put the first thing first. In the first

step, we minimize total energy with respect to the atomic positions. In the second step, we minimize the enthalpy with respect to the cell-edge vectors (stress tensor) and atomic positions (Hellmann-Feynman forces).