Timing, memory and file-space usage

The following are questions assessing your knowledge and ability to determine various run-time properties for simple systems, required to ensure that you submit jobs that will not run out of memory, disk-space or time on a given computational resource. An auxiliary file is provided from which you can obtained some parameters needed to answer the questions.

Time

1. What is the time taken for the first diagonalisation of the Hamiltonian? **Take care to include you units.**
2. What is the time taken to calculate the forces?
3. Based upon the time for the initial SCF step, the time for a force, and an assumed average number of 10 SCF steps per energy, estimate the time for a typical structural optimisation step using the conjugate gradients algorithm. State any assumptions you make. Estimate the total run time required for a complete optimisation, again justifying any assumptions made.
4. If the previous timing were for a Γ-point approximation, how would you expect the time to change if the calculation is repeated with an $MP2^3$ sampling scheme? [Assume there is no geometric symmetry beyond the identity and that the sampling is the only significant change to the calculation.]
5. For question 4, indicate how much time might be saved by using a restart-dump. Explain how you arrive at your answer.
6. A calculation of the energy second derivatives is being planned. The cell contains 256 atoms, and symmetry reduces the number of independent sites to just 168. The time per energy and force is estimated from the optimisation of the defect which may be summarised as follows:
    a. $MP2^3$, reducing to 2 points by symmetry.
    b. The initial SCF took 12 steps at approximately 580s per step.
    c. The initial forces took 113s.
    d. k-point parallelism is not used.

   Estimate the total time to obtain the full set of second derivatives, explaining clearly any assumptions made.
7. Explain why the time per energy in a typical structural optimisation decreases with increasing iteration number before saturating to an approximately constant time. [Assume that the optimisation is proceeding using the *optimise{all}* data file specification.]

Memory

1. What, in megabytes, is the peak amount of physical memory used at the point the first step in the SCF is reported in the attached AIMRPO output file? Be sure to express your answer with units.

2. Explain a method for determining the amount of **physical memory** available on one of the computers used in Newcastle. (Asking someone else does not count as a method.)
3. In broad terms, how does the memory used by AIMPRO depend upon:
    a. The size of the Hamiltonian
    b. The super-cell volume
    c. The plane-wave cut-off
    d. The number of special k-points
    e. The use (or not) of k-point parallelism
    f. The number of physical nodes
    g. Whether or not DIIS is used in the SCF
    h. Whether or not the real space build is enabled
4. What is the impact upon run-time when the memory requirements exceed the physical memory available? Explain your answer clearly.
5. What is meant by the term 'swapping' and how can you tell if your job is 'swapping' on Verity or Trueman?

Disk-usage

1. A defect is being analysed by a student to connect it to an optical transition seen in experiment. This requires classification of the defect-states according to symmetry, and an AIMVIEW dump file is required. There are 156 occupied bands for the system, which is being simulated using $MP4^3$ sampling, with the proposed optical transition, as determined from a band-structure plot, being between the highest occupied and lowest empty states. Which of the following would be the most suitable calculation to perform to address the aims of the student? Explain your choice in terms of both **information required** and the **size** of the AIMVIEW dump file.
    a. analysis{bandst,levels=1 156,dump}
    analysis{k-points,gamma}
    b. analysis{bandst,levels=1 256,dump}
    analysis{k-points,path,define=int-p,npt=11}
    c. analysis{bandst,levels=156 157,dump}
    analysis{k-points,grid=4 4 4}
    d. analysis{bandst,levels=150 160,dump}
    analysis{k-points,gamma}
    e. analysis{bandst,levels=1 300,dump}
    analysis{k-points,gamma}
2. Consider the calculation of a simple surface system consisting of a slab with eight surface sites terminated by hydrogen such that the system is chemically satisfied. One surface hydrogen atom is then removed, resulting in a single surface radical with a net electron spin of ½. The aim of the student performing the calculations is to characterise this surface defect in ways that might be directly compared to experiment. Rank the following calculations in terms of the amount of **disk-space** required to store the data generated in each of the following cases:
    a. Structural optimisation
    b. A band-offset analysis calculation
    c. The Mulliken bond populations

    d. A band structure following a path containing 6 branches in the surface Brillouin-zone
    e. EELS
    f. Energy second derivatives
    g. Hyperfine interactions
    h. A NEB calculation for the diffusion of the radical site across the surface

   In each case, note any parameters that influence the rank order, such as the number of energies and broadenings included in the EELS calculations.

3. For the calculations in Q2, identify which run-time output files, in addition to the AIMPRO standard output, you need to preserve (i.e. which ones contain useful information regarding the scientific application). Include in your answer files such as hgh-pots, fort.99, restart-dump, dat and the AIM.sh.* files, as well as any additional outputs, as appropriate.
4. By what proportion is the disk-usage of a typical text-based AIMPRO output file reduced by the use of a compression programme such as gzip or bzip2? What is the proportion for a restart-dump file? Why are these proportions not the same?
5. Write down the LINUX "find" command to locate and compress all files that have not been bzip2'd.
6. Write a similar find command that will delete all restart-dump files that have not been accessed for more than 60 days.
7. The following lists of LINUX commands include information regarding disk usage and file size. In each case explain how they're used, and what useful information you obtain:
    a. quota
    b. du
    c. df
    d. ls –l
8. On some machines there are disk quotas enabled that prevent a user exceeding a specified amount of file space. Explain the impact upon the user and other users if the quota is exceeded in terms of the running of AIMPRO in a batch queue.
9. What is the impact upon users when the user partition on the computer head-node is filled up? Does the use of quotas guarantee that the disk space cannot be completely exhausted?