

CP2K Tutorial

Exercises

Normal Mode Analysis

Problem: Calculation of the effect of I₂ binding to the SCN-group on the IR-spectrum of the N3 dye.

For this tutorial some input and output files are given in order to present a complete procedure to solve the given problem. In addition some tips and tricks to speed up the procedure will be introduced as well.

The tasks to be completed are:

- Optimize the geometry of the N3-dye in gas-phase.
- Compute the frequency of the SCN stretching vibrations.
- Optimize the structure of the SCN-bound N3-I₂.
- Recompute the stretching SCN-vibrations for this system.

1 First geometry optimization of N3-dye

An initial structure can be obtained by any tool of choice (e.g. marvin sketch). In general, these structures do not provide a good initial guess for DFT calculations. Therefore it is a reasonable choice to perform a preliminary optimization using semiempirical methods. The files from these calculation can be found in **NMA/GEO_OPT-N3/SE**.

In the QS section of the input file, parameters for the semiempirical calculations are set.

```
&QS  
  METHOD PM6  
  &SE  
    PERIODIC NONE  
  &END  
&END
```

In the SUBSYS section the cell parameters and the coordinates are specified.

```

&CELL
  ABC 22.000 22.000 22.000
  PERIODIC NONE
&END CELL
&TOPOLOGY
  &CENTER_COORDINATES
  &END
&END

```

Again the PERIODIC NONE keyword has to be specified in the CELL section. This input will run very quickly producing various files. The wavefunction restart files, which will be produced by all the runs have been deleted because of their size. As you can see, a second set of restart (SE_OPT-1.restart) files is written, containing all the information of the input file and additional default parameters. The file SE_OPT-BFGS.Hessian contains the approximate Hessian from the BFGS optimization. The full output is provided in the file out. Furthermore, a file containing the optimized positions can be found (SE_OPT-pos-1.xyz). Files containing coordinates are always named *-pos-*.xyz by CP2K. You can use VMD or other visualization programs to have a look at the trajectory. To extract the final geometry from this file you can use the command `tail -n 'head -n 1 SE_OPT-pos-1.xyz' SE_OPT-pos-1.xyz > final.xyz`. These positions will now provide a reasonable initial guess for the DFT calculations.

The files for the DFT optimizations can be found in `NMA/GEO_OPT-N3/QS`. For a normal mode analysis pseudopotential calculations will provide a sufficient accuracy. Therefore GPW will be used for all DFT calculations. As a functional PBE will be employed.

```

&XC
  &XC_FUNCTIONAL PBE
  &END XC_FUNCTIONAL
&END

```

Even it might be desirable to use tighter convergence criteria on the geometry for normal mode analysis, here the defaults of CP2K will be used.

The important settings in this input file are again related to the nonperiodic nature of these calculations. As the electrostatics in GPW are computed via plane waves, a special Poisson solver is needed to decouple the periodic interactions. CP2K provides several different possible options therefore, of which all have their advantages and drawbacks. The Martyna-Tuckerman solver requires large box sizes (twice the size the charge density needs to decay) but can be used for noncubic cells. The nonperiodic wavelet solver requires less vacuum ($\approx 3\text{\AA}$ in every direction) but is restricted to cubic cells at the moment and the molecule has to be centered in the box. In case of this solver, a warning will be printed if the cell is too small. As the N3 dye is a cubic molecule in first approximation (12\AA in every direction), the wavelet solver is suitable for this case and a cell of 22\AA will be sufficient. In the input file this reads as:

```

&POISSON
  PERIODIC NONE

```

```

        POISSON WAVELET
    &END

```

in the DFT section, and

```

&CELL
    ABC 22.000 22.000 22.000
    PERIODIC NONE
&END CELL
&TOPOLOGY
    &CENTER_COORDINATES
    &END
&END

```

in the SUBSYS section. Note again, that periodic none is specified in two places. Furthermore, it is crucial to use center_coordinates in combination with this Poisson solver.

The files types produced are the same as before (except of the prefix n3-geo). By a modification of the command above, the final positions can be extracted from the trajectory file. These positions will be used in the following calculation.

2 Normal mode analysis of N3

Next to a full normal mode analysis, which would require 6xNatoms force calculations, CP2K provides the option of mode tracking. This is especially useful if only a small subset of frequencies out of the full spectrum is of interest. As this is an iterative procedure (Davidson subspace algorithm), the initial guess is important for convergence. Moreover, there is no guaranty that the created subspace will contain the desired modes in case of a bad initial guess.

Various methods can be used to create an initial guess. The most obvious method is the use of normal modes obtained from lower quality methods. Here, PM6 is used to create this guess. Even though the PM6 calculation is fast enough to run a full NMA, a mode selective calculation will be used to demonstrate other ways of constructing an initial guess and selecting modes. The files related to the initial SE calculation can be found in **NMA/NMA-N3/SE**.

For all NMA calculations it is essential, that the geometry used is well optimized. This can be taken from the previous optimization **NMA/GEO_OPT-N3/SE/final.xyz**. The most important changes in the input file compared to the geometry optimizations are the change of the run type in the GLOBAL section

```

&GLOBAL
    PROJECT n3-geo
    RUN_TYPE NORMAL_MODES
&END GLOBAL

```

and the additional VIBRATIONAL_ANALYSIS section:

```

&VIBRATIONAL_ANALYSIS
  &MODE_SELECTIVE
    INITIAL_GUESS BFGS_HESS
    RESTART_FILE_NAME SE_OPT-BFGS.Hessian
    &INVOLVED_ATOMS
  INVOLVED_ATOMS 27 28 29 30 31 32
    RANGE 1500 3000
  &END
&END
DX 0.01
NPROC_REP 1
&END

```

A mode selective vibrational analysis is performed by simply adding the corresponding section. If the section `MODE_SELECTIVE` is absent, a full normal mode analysis will be performed. Here the BFGS Hessian is chosen as an initial guess (which is obtained by the initial geometry optimization). BFGS constructs only an approximate Hessian. Hence, this guess will not be very good, but is still better than an arbitrary guess. The section `involved atoms` activates a mode tracking procedure, which searches for the normal modes in which the motion of the specified atoms contribute the most. For these calculations these have to be the atoms of the SCN-groups. This selection rule might not be sufficient, as other modes might involve motions of these atoms as well. In order to obtain the correct modes a frequency range (in cm^{-1}) can be specified in addition to the involved atoms, to bias the search. The common frequency for the SCN-stretch is about 2100 cm^{-1} . Nevertheless, as SE methods tend to give poor spectra, the range is extended to both sides to ensure, the modes of interest will be converged.

Here, two keywords are used, which are common for both MS and full NMA. `DX` specifies the norm of the displacement vector for the numerical derivatives. In case of full NMA, this value is simply the displacement of the atoms along the Cartesian axes. For most calculations, values between 0.01 and 0.001 give reasonable results. Nevertheless, depending on the frequency range systematic errors are introduced. As a guideline, it is valid that for low frequencies `DX` should be slightly larger, while for highfrequent vibrations a lower value is recommended.

The last keyword `NPROC_REP` has a special meaning, depending on whether a full or a MS-NMA is performed. For full NMA it simply allows to create several parallel environments to compute more than one derivative at a time. As an example imagine an mpi run on 8 processors with `NPROC_REP` equals two. This will run 4 force calculations at the same time, each running on 2 CPU ($8/2=4$). If a setup like this is chosen in combination with MS, 4 force calculations will run in parallel as well. In contrast to the full NMA, this requires the evaluation of 4 displacement vectors first. Therefore instead of converging a single mode a block Davidson algorithm is used to converge multiple modes, which are closest to the selection in the input file.

For this system it is known, that there are two SCN groups. Therefore it makes sense to track two modes at the same time. According to the discussion above, this can be done by running the calculation on two CPU with the setting `NPROC_REP 1`.

In addition to the usual restart files, the file `MS_SE-MS_RESTART.bin` is created. This file

contains all information about displacements and forces which are necessary to restart a non-converged MS calculation. Moreover, two different types of output files are created. The *-r*.out files contain the output for the SCF/force calculation. As in this example, two force calculations are carried out at the same time, two of these files are created (MS_SE-r-0.out and MS_SE-r-1.out). Information about convergence of the Davidson algorithm can be found in the file out. This file contains all information related to the progress of the iterative procedure can be found, namely which modes are closest to the selection in the approximate Hessian and are therefore used to create the new residuum vectors. Because of the block Davidson scheme employed, two informations are printed for one iteration step. As you can see in the first step, the first mode is already close to the desired mode, while the second mode is completely off. Nevertheless, as increasing the subspace introduces more modes each iteration step, the second desired mode appears from the 5th iteration step and starts to improve until convergence is reached. The final bit of output prints all modes which are present in the subspace and their convergence. By increasing the subspace it is likely that modes which have not been explicitly tracked converge as well. This can be understood by the fact, that the method will be exact in the limit of a $3 \times \text{Natoms}$ dimensional subspace.

For visualization the file MS_SE-VIBRATIONS-1.mol is created. This contains the geometry and information about every mode. You can use molden to visualize the vibrations and check whether the converged modes are really the desired SCN vibrations. In case intensities are computed molden even produces a graph of the IR-spectrum as computed.

After having computed the modes using PM6, the output can be used as initial guess for the DFT calculation. For MS calculations the .mol file can be used to define the initial displacement vector. This is not sensitive to the absolute positions of the atoms but simply takes the displacement vector from a selected normal mode and applies it to the actual geometry. As both, the DFT and the PM6 geometries are reasonably close, the SCN vibrations from SE can serve as an initial guess for the QS calculation. This is specified by the initial guess and the restart file name:

```
&VIBRATIONAL_ANALYSIS
  NPROC_REP 24
  DX 0.01
  INTENSITIES
  &MODE_SELECTIVE
    RANGE 2000 2300
    INITIAL_GUESS MOLDEN_RESTART
    RESTART_FILE_NAME MS_SE-VIBRATIONS-1.mol
  &END
&END
```

Here, a different way of selecting frequencies is introduced. Instead of picking modes according to the displacement of atoms, all modes in the specified frequency range (in cm^{-1}) will be converged. To be precise, all modes in the spanned subspace which match this frequency range will be converged. This is important, as there is no guaranty, that all modes in this range will be found. By specifying the keyword INTENSITIES CP2K will compute the IR-intensities of

the modes. IR-intensities are defined as the derivative of the dipole vector with respect to the normal mode. Therefore it is necessary to activate the computation of the dipoles. This is done in the DFT section via:

```
&PRINT
  &MOMENTS
    PERIODIC FALSE
  &END
&END
```

Again it has to be specified, that the calculation is nonperiodic and the dipoles will be computed accordingly. Finally to speed up the calculation it is useful to use the wavefunction restart from the geometry optimization. All forces will be computed by displacements around the equilibrium geometry and the wavefunction from this geometry will be a better guess than the one obtained from previous displaced geometries. In order to always use the initial guess as a starting point and not to extrapolate the wavefunction, the following settings are used.

```
&QS
  EXTRAPOLATION USE_GUESS
&END
```

This run will produce the same type of output files as the SE run. The only difference is, that the .mol file will contain an additional section containing the intensities. If this file is opened with molden in addition to the list of frequencies a spectrum will be shown. To obtain the desired information with a full NMA, 354 force computations would have been necessary. In comparison, the MS converged after 9 steps, which corresponds in the block Davidson algorithm used to 18 points or 36 force calculations. This clearly shows, that for problems in which only a subset of frequencies is of interest the MS scheme provides a powerful tool to speed up the calculations. The advantage of using MS becomes even more evident going to larger systems. Contrary to NMA, the MS scheme converges almost independent of the system size in a few steps for a single mode and is only influenced by the initial guess.

3 Geometry optimization of the N3-I₂ system

For the second part of the vibrational analysis an initial geometry has to build again. As this will be just an I₂ molecule binding to the SCN group, this can be done by hand in vmd. The initial guess will be again preoptimized using PM6. The related files can be found in NMA/GEO_OPT-N3-I2/SE/ The procedure will be the same as for the first set of calculation and nothing new will be introduced. The files for the DFT optimization are located in NMA/GEO_OPT-N3-I2/QS/

4 Normal mode analysis of N3-I₂

For the NMA the same strategy could be used as before, which is likely to be the most efficient for this problem. Nevertheless, to give an insight in the performance of the MS scheme in case

no initial guess is present the normal modes are computed directly via DFT. The files for these calculations can be found in **NMA/NMA-N3-I2/QS** If little is known about the modes of interest (e.g. a frequency range and contributing atoms) an initial guess can be build by a random displacement of atoms. For these calculations a random displacement for the SCN-groups is chosen. The remaining settings are known from before. It is unlikely for I₂ to shift the SCN frequencies more than 150 wavenumbers. Therefore, a range of ± 150 wavenumbers around the N3-SCN frequencies is chosen to converge.

```
&VIBRATIONAL_ANALYSIS
  NPROC_REP 24
  DX 0.01
  INTENSITIES
  &MODE_SELECTIVE
    RANGE 2000 2300
    INITIAL_GUESS ATOMIC
    ATOMS 27 28 29 30 31 32
  &END
&END
```

By specifying the ATOMS keyword, the initial displacement will be constructed according to the description above.

Now everything is at hand to compare to the experimental spectrum of an N3-dye attached to an anatase surface once in neat acetonitrile and in an solution of I₂ in acetonitrile shown on the next page.

5 Test Problems

The system above is to large to run the calculations in a reasonable time on the available machines. If you are interested in running NMA calculations in this tutorial, the structures for maleinic and fumaric acid are given in **ACID/**. Here the computation of the COOH frequencies for the different isomers might be of interest.

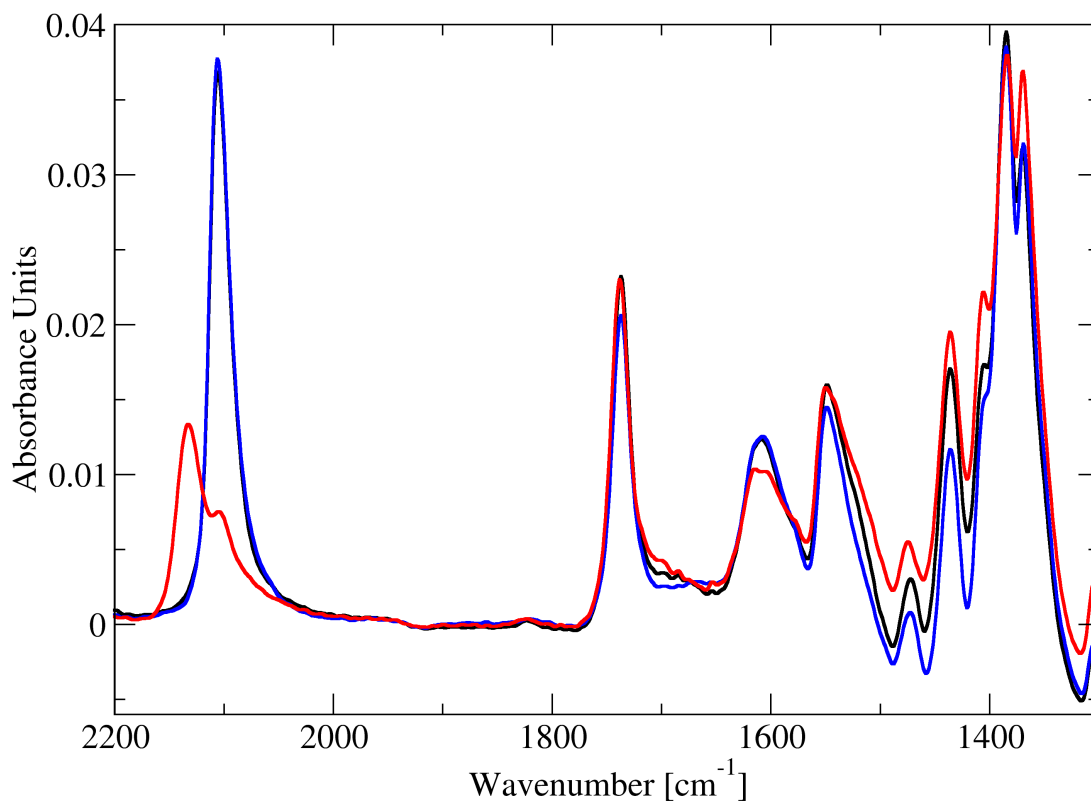


Figure 1: Experimental IR-spectra of the N3 dye attached to TiO_2 in acetonitrile, before adding iodine (black line), with I_2 added (red line), and after washing iodine away (blue line). The pronounced peaks at 2098cm^{-1} and 2130cm^{-1} correspond to vibrations of the SCN and I_2SCN ligands, and support our finding of reversible complex formation.