

Nitrogen-rich atmospheres, DFT study of a precursor molecule to life's building block molecules: HCN

Damien Caliste

May 26, 2025

Abstract

The objectives of these exercises are to explore molecular-related characteristics, like molecular orbitals, thermodynamic equilibrium, reaction heat and vibrational modes, by means of DFT.

1 Introduction

Hydrogen cyanide (HCN) is a pivotal molecule in the context of astrochemistry and the formation of nitrogen-rich atmospheres on exoplanets and moons within our solar system. HCN is recognized as a key precursor for the synthesis of life's building blocks, including nucleotides, amino acids, and lipids. Its formation in nitrogen-rich atmospheres is significant for understanding the potential for prebiotic chemistry on other planetary bodies.

The formation of HCN in an atmosphere rich in nitrogen (N_2) requires the breaking of the strong triple bond between nitrogen atoms and the subsequent combination of atomic nitrogen with carbon. This process can occur through various mechanisms, including photochemical reactions, lightning, impacts, or volcanic activity. The presence of HCN is particularly notable in environments where the carbon to oxygen ratio (C/O) is greater than or equal to one, which facilitates the production of significant amounts of HCN. For instance, HCN can be readily produced in atmospheres rich in methane (CH_4) or acetylene (C_2H_2), but it can also form in carbon monoxide (CO)-dominated atmospheres [1].

In our solar system, HCN has been detected in the atmospheres of several celestial bodies, including Saturn's moon Titan, which has the most HCN-rich atmosphere [2]. Titan's atmosphere, which is dense and nitrogen-dominated, provides a natural laboratory for studying the complex atmospheric chemistry that could lead to the formation of prebiotic molecules. The presence of HCN in such environments suggests that similar processes

could occur on exoplanets with nitrogen-rich atmospheres, potentially leading to the formation of complex organic molecules necessary for life.

2 From atomic orbitals to molecular orbitals

Let's consider the simple HCN molecule.

2.1 The self-consistent field (SCF) loop

The BIGDFT code takes parameters from the `hcn.yaml` file (overriding values from the `default.yaml` file, if present). In this first calculation, the file `hcn.yaml` is simple and mainly defines the position of the three H, C and N atoms of the molecule.

```
1 | $ cat hcn.yaml
2 | dft:
3 |   nplot: 5
4 | posinp:
5 |   units: angstroem
6 |   positions:
7 |     - C: [9.81019854686240422E-01, 0., 0.]
8 |     - H: [-9.34210604090287555E-02, 0., 0.]
9 |     - N: [2.13800116223523951E+00, 0., 0.]
10 | $ bigdft -l no -n hcn | tee hcn.log
11 | [...]
```

Browsing quickly through the log, one can see the convergency loop of the DFT calculations. The system converged after 15 iterations, during which the Kohn-Sham energy (labelled EKS) is bring to its final value of -1.619 Ha.

The DFT formalism is a way to obtain the electronic wavefunctions of a system defined by an hamiltonian containing the forces created by the nuclei on the electrons, the repulsive forces of the electrons themselves and their kinetic energy. It is done self-consistently since all parts of the hamiltonian depends on its solutions (*i.e.* the electronic wavefunctions).

The numerical implementation of the problem follows an iterative approach called self-consistent field (SCF) loop. During this convergency loop, the electronic wavefunctions of the system are optimised from an input guess. Various DFT implementations are using various ways to create input guesses.

Here, BIGDFT is constructing a first hamiltonian from the atomic orbitals of the atoms of the system (the s and p atomic orbitals...) and diagonalise it. Then, it is entering the SCF loop where the wavefunctions are optimized up to a tolerance. Then a final diagonalisation is done. The eigenvalues of this diagonalisation are visible after the SCF loop.

```
1 | #Eigenvalues and New Occupation Numbers
```

```

2 |   Orbitals: [
3 |   {e: -8.453433822250E-01, f:  2.0000},   # 00001
4 |   {e: -5.549772248026E-01, f:  2.0000},   # 00002
5 |   {e: -3.348156347862E-01, f:  2.0000},   # 00003
6 |   {e: -3.302070349411E-01, f:  2.0000},   # 00004
7 |   {e: -3.302069307372E-01, f:  2.0000}] # 00005

```

We computed a super-cell containing 3 atoms, providing $1 + 6 + 7 = 14$ electrons. There should be 14 occupied levels for this molecule. But actually, we're considering that the HCN molecule is non-magnetic and each level is occupied by two electrons, one with a spin-up quantum number and one with a spin-down. Spin-up and spin-down electrons are degenerated and their wavefunctions have the same energies. So this system should have $14/2 = 7$ levels. But to reduce the computational load, we're applying an additional approximation: the electrons with the lowest energies are electrons with a strong link to the nuclei and with a reduced spatial expansion. We call them the *core* electrons. They are the two 1s electrons for the carbon or the nitrogen. We're using the approximation that their wavefunctions are only influenced by the nuclei they belong to and not by the environment of these nuclei. Their wavefunctions are frozen and can be computed once for every chemical species. The remaining electrons are called *valence* electrons. They feel a modified potential from the nuclei, screened by the core electrons. This approximation is called *pseudo-potential* approximation and helps reducing the computation load by reducing the number of wavefunctions to optimize to the valence electrons. In our case, carbon atoms are represented by 4 valence electrons and nitrogen by 5, thus $\frac{1+4+5}{2} = 5$ levels.

What I've learnt 2.1

Compute the total energy of an isolated molecule, dealing with its valence electron wavefunctions.

2.2 Molecular orbitals

The 5 occupied levels of the HCN molecule are defined by their eigen-value energies (with respect to the vacuum energy set at 0).

The eigen-values present a deep level at -23 eV. This level is associated to a *s*-like orbital resulting from the hybridisation of the two *s* orbitals of the C and N atoms. It is a strong bonding molecular orbital.

There is then another quite deep level at -15 eV. It corresponds to a wavefunction delocalised on the full molecule, encompassing the hydrogen and the nitrogen atoms.

Finally, there are three high levels around -7 eV, the two last being degenerated, as visible on Fig. 1. The two degenerated wavefunctions come from the hybridisation of the p_z and p_y orbitals of the C and N atoms,

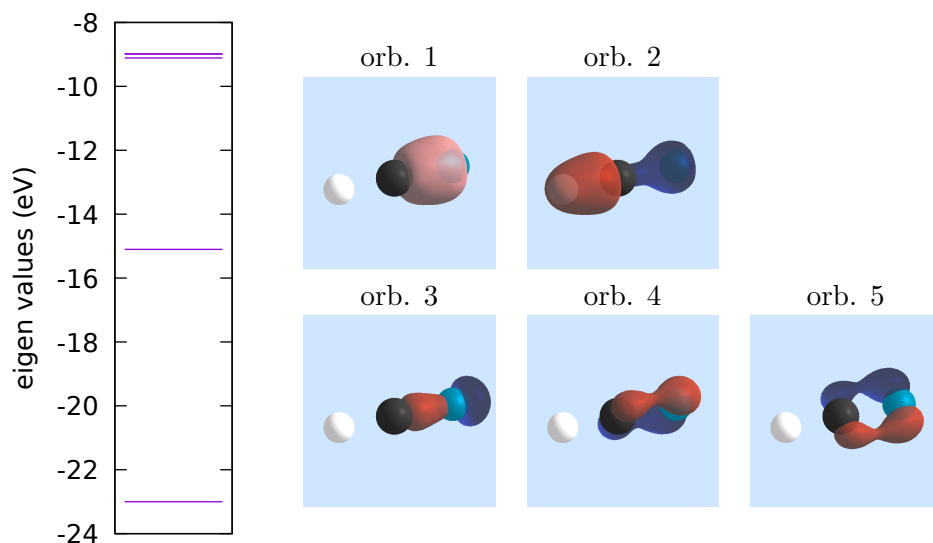


Figure 1: Eigen energies for HCN molecule and the associated eigen states plotted as isosurfaces at 50%.

while the p_x being along the bond direction got a different hybridisation and energy.

After running the calculation, the code generated CUBE files for each of the 5 occupied orbitals, thanks to the `nplot: 5` entry in the input file. Let's move these files into a separated directory (and remove the axis averagings):

```
1 | $ mkdir data-hcn
2 | $ mv *.cube data-hcn/
3 | $ rm -f *.dat
4 | $
```

The CUBE files contain volumetric data on a regular 3D mesh. Here the `orbitalxxxx.cube` correspond to the wavefunction values on the grid, $\Psi_i(\mathbf{r})$. The wavefunctions are complex number objects, but for isolated systems, there is no phase information and the wavefunctions are simply real numbers.

One can visualise 3D volumetric data using various methods, like colour mapping in a plane or on a surface, or also displaying iso-surfaces of the volumetric data. Here is an example using V_Sim software:

```
1 | $ v_sim -f . -v auto data-hcn/orbital0004.cube
2 | $
```

What I've learnt 2.2

Look at the eigen-values and eigen-functions of a DFT calculation.

2.3 Wavefunctions orthogonality

As electrons are fermions, they need to respect an orthogonality constraint. If $|\Psi_i\rangle$ is a wavefunction of the system, we must have $\langle \Psi_i | \Psi_j \rangle = 0$ for $i \neq j$. This constraint is applied during the SCF loop at each iteration when the eigen values are modified.

We can check by hand that this orthogonality is verified on the output wavefunctions with a simple Python script:

```
1  #!/usr/bin/env python3
2
3  import numpy
4  import cube
5
6  Norb = 5
7
8  orbs = []
9  for i in range(Norb):
10     filename = "data-hcn/orbital%04d.cube" % (i+1)
11     orbs.append(cube.read_file(filename))
12
13  overlap = numpy.zeros((Norb, Norb))
14  for i in range(Norb):
15     for j in range(i + 1):
16         overlap[i, j] = numpy.sum(orbs[i] * orbs[j])
17         overlap[j, i] = overlap[i, j]
18  dev = overlap - numpy.identity(Norb)
19  print("deviation from identity:",
20        numpy.sqrt(numpy.sum(dev**2)))
```

Running this script gives a deviation from the identity of the overlap matrix around 10^{-8} , which is a bit poor, but reasonable. Indeed, CUBE files are storing the volumetric data with 6 precision digits only. They are meant for visualisation purposes, not actual calculation. Internally, the wavefunctions are stored as coefficients on a Daubechie wavelet basis-set that can give machine precision accuracy.

What I've learnt 2.3

Basic properties of the DFT solutions, like wavefunction orthogonality.

3 Studying thermodynamics with DFT

We are interested in the following chemical reaction:



DFT calculations provide the total energy of each constituent of the reaction, in their stable conditions, as infinitely separated from each others. Let's also assume here that the constituents are in a gaseous phase, surrounded by vacuum. Reactions in solution where constituents are surrounded by a dielectric can also be treated by numerical approaches but they are out of scope of the present exercise.

3.1 Thermodynamic equilibrium of a molecule

We have seen that the total energy of a molecule is a parameter of the atomic constituents of the molecule, and of their respective positions in space. At equilibrium position, the forces (*i.e.* the derivative of the total energy with respect to the atomic positions) are zero. But if the atoms are slightly not at their equilibrium position, forces can be calculated and used in a minimisation algorithm to reach the equilibrium position.

We can try this with H₂ molecule.

```
1  $ cat h2.yaml
2  dft:
3    nplot: 1
4  geopt:
5    method: FIRE
6    ncount_cluster_x: 75
7    frac_fluct: 5.
8  posinp:
9    units: angstroem
10   positions:
11     - H: [0., 0., 0.]
12     - H: [0.7, 0., 0.]
13  $ bigdft -l no -n h2 | tee h2.log
14 [...]
```

At the end of the SCF loop, forces are computed. Then, a geometry optimisation algorithm is displacing the atoms according to the forces and a new SCF loop is performed on the new positions.

In the dihydrogen molecule, the algorithm converged to the equilibrium position within 22 iterations. The interatomic distance is computed as $d(\text{H}_2) = 0.75 \text{ \AA}$. This value is reasonably close to the experimental value of 0.74 \AA [3], within the approximations used in the calculation (spin is averaged, exchange-correlation functional is PBE...).

What I've learnt 3.1

Perform a geometry optimisation to get the equilibrium positions.

3.2 Heat exchange in a reaction

In full generality, the heat exchange when a reaction occurs is the variation of the Gibbs free energy, defined as:

$$G = E + PV - TS \quad (2)$$

The DFT is providing the internal energy E . The additional term PV to get the enthalpy from the internal energy, is zero in the case of gaseous phases (or more precisely isolated molecules). The entropic term is difficult to compute and arise from the vibrational entropy, but not only. At low temperatures, it can often be neglected.

To compute the ΔE for reaction 1, let's compute the total energy of the product, the methanimine. We're using already an already converged geometry for this molecule, to save time.

```
1 | $ cat cnh3.yaml
2 | dft:
3 |   nplot: 6
4 | posinp:
5 |   units: angstroem
6 |   positions:
7 |     - C: [5.15779360E-002, 0.58522520, 0.]
8 |     - N: [6.18051595E-002, -0.68784819, 0.]
9 |     - H: [-0.86401468, 1.20074404, 0.]
10 |    - H: [ 1.00606196, 1.12696521, 0.]
11 |    - H: [-0.90173037, -1.04918626, 0.]
12 | $ bigdft -l no -n cnh3 | tee cnh3.log
13 | [...]
```

The total energies of the three molecules in the reaction are:

HCN	Energy (Hartree):	-1.61897504689772020E+01
H ₂	Energy (Hartree):	-1.16625983814376810E+00
CNH ₃	Energy (Hartree):	-1.73895022289925265E+01

Which allows to compute the ΔE , as the difference between the product internal energies with the reactant internal energies. Here, $\Delta E = -911$ meV. The reaction is exo-thermic, so it may happen spontaneously. Nevertheless, this approach, using the thermodynamic equilibriums does not account for the required activation energy to get the reaction to occur.

Computing the activation barrier of a reaction is an interesting problem that requires to investigate the transition state (or the various transition states) that links the reactants to the products.

What I've learnt 3.2

Compute and discuss the energetics of a reaction based on the variation of the internal energy.

4 Vibrational spectra of a molecule

Let's now look at the vibrational modes of the HCN molecule. At its equilibrium geometry, the first derivatives of the total energy with respect to the atomic positions are zero (the forces are null). The second derivatives give insights on the vibrations.

Second derivatives could be computed in a perturbative way, like forces. But simple finite differences also work well. That's the approach used here. All atoms are slightly displaced along x , y and z directions, one at a time, to compute the second derivative by finite differences. This creates the Hessian matrix defined as $\frac{\partial^2 E}{\partial x_i \partial x_j}$ where x_i is the dependance on atom α along direction ν . The Hessian matrix is thus of dimension $(3N_{\text{atoms}})^2$. In a finite difference approach, each coefficient is obtained by computing the forces twice, one for each plus and minus displacement on a given axis. In total, it thus requires to run $2 \times 3 \times N_{\text{atoms}}$ calculation of forces.

When the Hessian matrix is calculated, mass of the nuclei are added to obtain the dynamical matrix. Its eigen-values represent the frequencies of the vibrational modes, and its eigen-states the direction of the atomic displacements.

4.1 Obtaining the Hessian matrix by finite differences

BIGDFT automatises the computation of the Hessian matrix, using another executable.

```
1  $ cat hcn.yaml
2  posinp:
3      units: angstroem
4      positions:
5      - C: [9.81019854686240422E-01, 0., 0.]
6      - H: [-9.34210604090287555E-02, 0., 0.]
7      - N: [2.13800116223523951E+00, 0., 0.]
8  $ cat hcn.freq
9  1/64 #finite diff step = alpha*hx, alpha*hy, alpha*hz
10 2    #order finite difference scheme (2 or 3)
11 1    #1 - systematic moves over each atoms
12 $ frequencies -l no -n hcn | tee freq.log
13 [...]
14 (F) Frequencies (cm-1): [ 3345.30,  2086.55,  731.89,
15                          728.69,   0.26, -0.30,
16                          -0.43, -19.29, -22.63]
17 [...]
18 $
```

There are three modes of almost zero frequency. They correspond to a rigid translation of the molecule and are not vibrational modes.

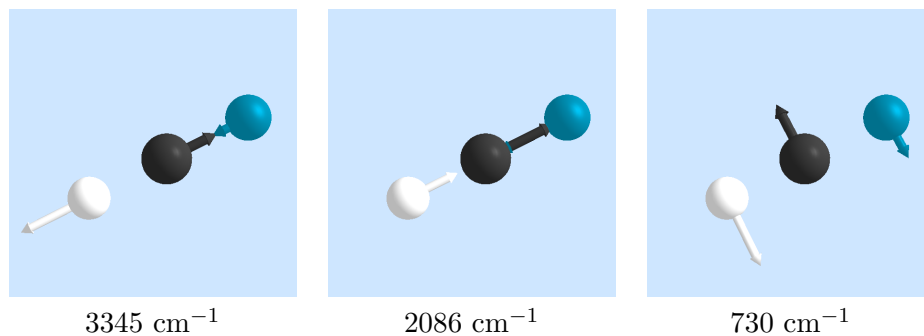


Figure 2: Vibrational modes of the HCN molecule.

4.2 Analysing the vibrational modes

There are four modes with positive frequencies (Fig. 2):

- one mode at 3345 cm⁻¹, which is a stretching movement mainly within the C-H bond,
- one mode at 2086 cm⁻¹, which is another stretching movement within the C-N bond,
- two degenerated modes around 730 cm⁻¹, which are bending modes perpendicular to the molecule axis.

These modes compare well to the three experimentally reported ones at 3311 cm⁻¹, 2097 cm⁻¹ and 712 cm⁻¹ [4].

References

- [1] P.B. Rimmer, S. Rugheimer, [Hydrogen cyanide in nitrogen-rich atmospheres of rocky exoplanets](#) *Icarus* **329** 124-131 (2019)
- [2] M. Rengel, D. Shulyak, P. Hartogh, H. Sagawa, R. Moreno, C. Jarchow, and D. Breitschwerdt, [Ground-based HCN submillimetre measurements in Titan’s atmosphere: an intercomparison with Herschel observations](#) *A&A* **658** A88 (2022)
- [3] Huber, K.P.; Herzberg, G., [Molecular Spectra and Molecular Structure. IV. Constants of Diatomic Molecules](#), Van Nostrand Reinhold Co., 1979
- [4] G Herzberg, [Molecular Spectra and Molecular Structure II. Infrared and Raman Spectra of polyatomic molecules](#) 1945 Van Nostrand.