Computational Nanoscience: Exercise Sheet No. 4

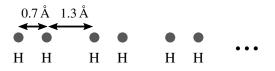## Exercise 4.1: Computational scaling of Hartree-Fock

In this exercise, we investigate how the computational cost of closed-shell Hartree-Fock calculations grows with the number of atoms in a system. For simplicity, suppose we only have the same atom type in the system and we keep the same basis set and the same numerical parameters when increasing the system size. As practical example of such a test system, you can think of a chain of H_2 molecules with increasing number of H_2 molecules. As basis for this exercise, have a look at Fig. 5.2 in the lecture notes.

(a) How does the number of basis functions N_b and the number of occupied molecular orbitals increase with the number of atoms N_{atom} in the calculation?

(b) The computational cost of every single step in the Hartree-Fock algorithm in Fig. 5.2 increases (= scales) as N_{atom}^α, $\alpha \in \mathbb{N}$. Identify the steps where the computational cost increases with the highest exponent on N_{atom} (without reformulating the algorithm).

(c) We introduce the density matrix

$$D_{\lambda\sigma} = \sum_{m=1}^{N/2} C_{m\lambda} C_{m\sigma} \, .$$

How can we use the density matrix to reduce the scaling of Hartree-Fock to N_{atom}^4?

(d) Suppose, we have localized basis functions $\phi_\nu(\mathbf{r})$, for example atom-centered Gaussians as in Turbomole, and suppose we have a large molecule (e.g. system size in one direction $20\,\text{Å}$). What is the condition on the basis functions $\phi_\nu(\mathbf{r})$, $\phi_\mu(\mathbf{r})$, $\phi_\sigma(\mathbf{r})$, $\phi_\lambda(\mathbf{r})$ in four-center Coulomb integrals $(\mu\nu|\lambda\sigma)$ such that the integrals have a small numerical value (for example smaller than 10^{-12})?

(e) Suppose, we neglect all four-center integrals $(\mu\nu|\lambda\sigma)$ with small numerical value (for example: threshold 10^{-12} in the Hartree-Fock calculation). Argue that the number of kept integrals $(\mu\nu|\lambda\sigma)$ scales with N_{atom}^2 for reasonably large molecular systems (e.g. system size in largest direction $\approx 20\,\text{Å}$).

(f) Based on rigorous screening in large molecules, we only keep $O(N_{at}^2)$ four-center integrals $(\mu\nu|\lambda\sigma)$. In such a scenario, how is the computational cost of Hartree-Fock increasing with the system size?

(g) Measure the CPU time of Hartree-Fock for H_2 molecules and fill the table on the next page. You will find the CPU time in the Turbomole output after "total cpu-time". Use a def2-QZVPP basis and a geometry of the H_2 molecules as follows.

Number of H_2 molecules, N_{H_2}	CPU time (s)
1	
2	
4	
8	

Hint 1: Change the xyz file such that the lower geometry is used in the calculation. After creating the xyz file, type `x2t H2.xyz > coord.turbo`. The coordinates in the xyz file are in the unit Angström, while the coordinates in the coord.turbo file are in bohr (1 bohr = 0.529 Å).

Hint 2: It is no problem if an error message appears that you did not define any internal coordinates. Internal coordinates are only used for geometry optimizations.

(h) Assume the CPU time scales as

$$\text{CPU time} = C \cdot N_{H_2}^{\alpha} \, .$$

Determine the exponent α when using your data points (1) 1 H_2 and 2 H_2 and (2) 4 H_2 and 8 H_2.

(i) What is the reason in the Hartree-Fock algorithm for the difference between the exponents you have determined in (h) (1) and (2)?