

JACEK KORCHOWIEC*, JAKUB LEWANDOWSKI*

ELONGATION CUTOFF TECHNIQUE: PARALLEL PERFORMANCE

It is demonstrated that the elongation cutoff technique (ECT) substantially speeds up the quantum-chemical calculation at Hartree-Fock (HF) level of theory and is especially well suited for parallel performance. A comparison of ECT timings for water chains with the reference HF calculations is given. The analysis includes the overall CPU (central processing unit) time and its most time consuming steps.

Keywords: *order- N methods, fragmentation techniques, elongation method*

METODA WYDŁUŻANIA W WARIANCIE ZRÓWNOLEGLONYM

Pokazano, że technika obcięcia zastosowana w metodzie wydłużania istotnie przyspiesza obliczenia kwantowo-chemiczne w formalizmie Hartreego-Focka (HF). Metoda ta sprawdza się wyjątkowo dobrze w obliczeniach przeprowadzanych na wielu procesorach. W pracy porównano zarówno całkowity czas obliczeń, jak i jego najbardziej czasochłonnych części, z tradycyjnym schematem obliczeniowym metody HF.

Słowa kluczowe: *metody klasy $O(N)$, techniki fragmentacyjne, metoda wydłużania*

1. Introduction

Huge progress in computer technology brings about conditions for ab initio computational methods to be of practical value in chemistry, biochemistry and material science. The expense of a given method is measured by its scaling property. This is usually expressed as βN^α , where N , α and β are the number of basis functions applied in the calculations, scaling exponent and scaling prefactor, respectively. The application of Hartree-Fock (HF) and post-HF methods to routine calculation on huge molecular systems is limited due to undesired scaling behavior. The HF method scales formally as N^4 . However, many of two-electron integrals (TEI) are negligible and the upper bounds to these integrals can be easily computed. In such a case, the cost can be reduced to N^2 [1]. Much higher scaling behavior is typical for electron correlation methods: conventional MP2, CCSD, and CCSD(T) scale as fifth, sixth, and seventh

* K. Guminski Department of Theoretical Chemistry, Faculty of Chemistry, Jagiellonian University, Krakow, Poland

power of N , respectively. The scaling properties of Kohn-Sham (KS) method is the same or lower than HF one.

In the last two decades one can observe a growing interest in developing so-called order- N methods for electronic structure calculations [2, 3, 4, 5, 6, 7, 8, 9, 10, 11] i.e., methods which scale linearly with N . The common feature of these methods is that the total energy is computed without solving the eigenvalue problem of the whole system. This goal is achieved by dividing the entire system into smaller molecular fragments. By solving fragment HF or KS equations one can compute the system total energy via assumed matching rules. The size of the fragment is a parameter controlling the accuracy of the calculations. Apart from fragmentation methods, the efforts of many researches are directed towards more efficient techniques that allow to linearize every step of HF or KS method. Here, the best examples are the fast multiple moments for continuous charge distributions [12], near-field-exchange [13], order- N exchange [14], density-matrix minimization techniques [15].

In the elongation cutoff technique (ECT) which is described in this paper [9], the electronic structure is synthesized by stepwise adding a monomer unit to a starting oligomer. Therefore, the whole process resembles the mechanism of polymerization/copolymerization reaction. Every elongation step is followed by molecular orbital (MO) localization procedure [6]. The localized MOs (LMOs) that are far away from the chain propagation center are kept frozen. This picture of interactions is in agreement with traditional chemistry concepts; the properties of bonds between given atoms are determined by these atoms and their nearest neighbors. ECT diminishes dimension of the variational space that remains practically constant. Moreover, it substantially reduce the number of TEI that have to be evaluated in the SCF process. In this paper, a comparison of ECT timings for water chains with the reference HF calculations is given. Analysis is focused on all time consuming steps: generation of initial guess, computation of TEI and solving HF equation. The paper is organized as follows. First, we describe ECT. Then the information concerning computational details are given. Next, the results are presented and, finally, the conclusions are given along with the future prospects.

2. Elongation cutoff technique

The HF calculations performed for a starting cluster M_1 followed by MOs localization into A_1 and B_1 regions initialize the elongation scheme [2, 7]. The fragment A_1 is defined by atoms that are far away from the chain-propagation center. The remaining atoms belong to B_1 . More precisely, the definition of the fragment extends to all its atomic orbitals (AOs). In the first elongation step, the system is enlarged $M_2 = M_1 + C_1$ by adding a new monomer C_1 . MOs assigned to A_1 region are kept frozen while those assigned to B_1 and C_1 define a new working space S_1 . The whole system can be denoted as $M_2 = (A_1|B_1+C_1) \equiv (A_1|S_1)$, where the vertical solid line separates frozen and active MOs. After SCF is converged, the active space MOs are localized into A_2 and B_2 regions. The remaining steps follow in the same manner, therefore, in

the n th step $M_n = M_{n-1} + C_{n-1} = (A_1 + A_2 + \dots + A_{n-1} | B_{n-1} + C_{n-1}) \equiv (A^{n-1} | S_{n-1})$. The process is continued until desired length of the polymer is obtained.

The idea of ECT is simple. Let us consider the transformation narrowing the variational space:

$$\mathbf{F}^{\text{MO}}(S_1) \equiv (\mathbf{C}_{S_1})^\dagger \mathbf{F}(M_2) \mathbf{C}_{S_1} \quad (1)$$

where $\mathbf{F}(M_2) \equiv \mathbf{F}(A_1 | S_1) \equiv \mathbf{F}(A_1 | B_1 + C_1)$ is the Fock matrix of system M_2 . Notice that only active space MOs \mathbf{C}_{S_1} are involved.

In the limit of perfect localization, the frozen orbitals \mathbf{C}_{A_1} have no tails in the active part and vice versa, the active orbitals have no tails in the frozen region. In other words, some of the blocks in $\mathbf{C} = (\mathbf{C}_{A_1}, \mathbf{C}_{S_1})$ are filled with zeros. The zero block in the active part simply cuts $\mathbf{F}_{A_1 A_1}$, $\mathbf{F}_{A_1 S_1}$, and $\mathbf{F}_{S_1 A_1}$ blocks out when we transform $\mathbf{F}(M_2)$ to MO representation. The only contribution that survived in the resultant matrix is $\mathbf{C}_{S_1 S_1}^\dagger \mathbf{F}_{S_1 S_1} \mathbf{C}_{S_1 S_1}$ one. Therefore, at this point we have clear-cut evidence that the construction of full AO Fock matrix is a waste of time. Instead we can compute only $\mathbf{F}_{S_1 S_1}$ block [9].

This very ideal situation never takes place due to MO tails beyond the localized area. Therefore, we have to introduce a threshold value that reflects the effective coupling between frozen and active orbitals. If the coupling is below the threshold we can exclude a given subset of frozen MOs from direct calculations. Of course, it does not mean that the interaction between the cut and remaining fragments are switched off. Therefore, the density matrix of the whole system $\mathbf{D}(M_2) \equiv \mathbf{D}(A_1) \oplus \mathbf{D}(S_1)$ is required for constructing $\mathbf{F}_{S_1 S_1}$. Its cutoff part $\mathbf{D}_{\text{cutoff}} \equiv \mathbf{D}(A_1)$ is invariable in current and subsequent steps. One should also notice that the cut MOs introduce a correction $\delta \mathbf{D}$ to $\mathbf{D}(S_1)$. In practice, after at least two elongation steps we can perform the first cutoff step. It is worth mentioning here, that we compute less TEI than for normal elongation or HF run and this difference is getting bigger and bigger with every subsequent cutoff step.

3. Computational Details

The elongation and cutoff elongation methods have been implemented and linked to the GAMESS program package [16]. Test calculations were performed for linear water chains. The conventional type of calculations, with integrals stored on a disc were carried out. Two basis sets, namely, STO-3G [17] and 6-31G [18] were applied. Several chains built from $N = 20$ to 290(200) water units were taken into account for STO-3G (6-31G) basis set. The threshold for the density matrix was assumed to be 10^{-6} while for cutoff calculations was assumed to be less than 10^{-8} . Only one type of partitioning ($m/n = 20/15$) was considered, i.e., the starting cluster contained 20 water molecules and 15 molecules were added and frozen in a time. The cutoff calculations were initiated for 50 water molecules. In order to speed up the SCF convergence, the second-order method for orbital optimization was adopted for HF and ECT calculations [19]. The translational symmetry was not taken into account.

The HF and ECT calculations were carried out on 2.33 GHz Xeon architecture with two 4-cores processors. The step (τ_i^{ECT}) and overall [$\tau(\text{ECT}) = \sum_{i=1}^N \tau_i^{\text{ECT}}$] CPU time of elongation calculations were compared with the HF reference time [$\tau(\text{HF})$].

4. Results and Discussion

Let us first compare ECT with HF reference calculations. In Figure 1, the CPU time savings can be seen, i.e., the ratio $\tau(\text{ECT})/\tau(\text{HF})$ is plotted against the system size. Part a corresponds to STO-3G while b to 6-31G. At the beginning this ratio is higher than one for both basis sets. It is connected with the additional costs we have to pay in ECT (MO localization and elongation procedure). The next observation is that all curves tend to the asymptotic values. These values depend on the number of nodes (CPU) used in the calculations. Lines connecting circles, squares and triangles correspond to 1, 4 and 8 nodes, respectively. We would like to stress here that ECT as well as HF reference calculations were performed on the same number of nodes.

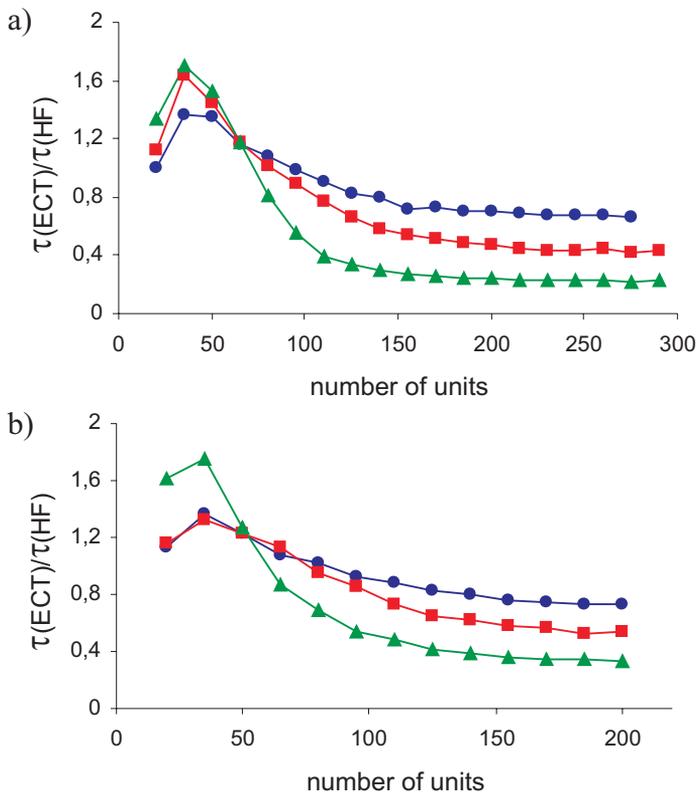


Fig. 1. CPU time savings $\tau(\text{ECT})/\tau(\text{HF})$ for STO-3G (a) and 6-31G (b) basis set. The circles, squares, and triangles correspond to 1, 4, and 8 nodes (CPU), respectively

Roughly speaking, we reduce the total CPU time for STO-3G by 34, 56, and 77% when the number of nodes increases. In the case of 6-31G basis, the reduction of CPU time is 27, 46, and 67%, respectively. In contrast to STO-3G, the usage of CPU time in the HF calculation for 6-31G is lower than 99% due to huge number of TEI (read/write operations). The usage for ECT is greater than 99%. Therefore, the real clock time should be plotted. Then, the CPU time savings increase to 32, 60, and 77%.

In applications of HF calculations to huge molecular systems the following steps are especially time consuming: preparation of initial guess, computation of TEI, and SCF process. In real, it is diagonalization which limits routine quantum-chemical calculations to moderate size molecules (about 200 atoms) since it depends as third power of N . However, due to very small prefactor in the scaling relation this undesired behavior is manifested for big molecules. The diagonalization is performed several times in the initial guess and SCF steps. Here, we demonstrate that this problem does not exist in ECT calculations (Fig. 2).

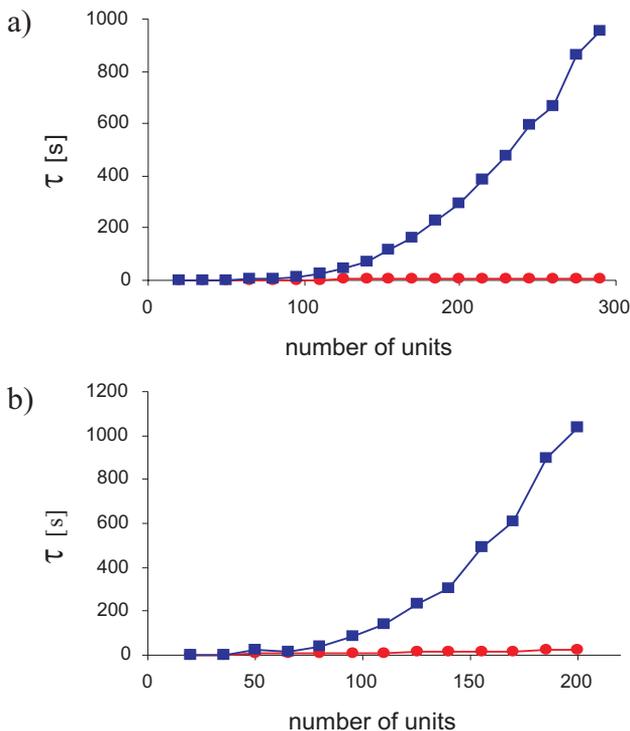


Fig. 2. CPU time required for generation of initial guess for STO-3G (a) and 6-31G (b) basis sets. The circles correspond to overall ECT while squares corresponds to HF reference. Calculations were performed at four nodes

Figure 2 concerns initial guess preparation, i.e., the overall CPU time for ECT is compared with HF reference one. Again parts a and b correspond to STO-3G and 6-31G basis sets, respectively. This step is not well suited for parallel performance. The time savings, when going from one to eight nodes, are negligible, therefore, we have plotted the curves corresponding to calculations performed on four nodes. It is clear from the picture that this step is very cheap for ECT calculations and practically doesn't depend on the system size. It is connected with the fact that we are working in the subspace containing only 35 water units. Therefore, time per step is constant. In consequence, the overall time for generating the initial guess depends linearly on the number of water molecules and the prefactor in this relation is very small. In contrast, HF time in this step increases quickly with the system size. In the case of STO-3G (290 water molecules), ECT for initial guess part is 108 times faster than the reference HF method. For extended 6-31G basis (200 water molecules), this ratio is equal to 46.

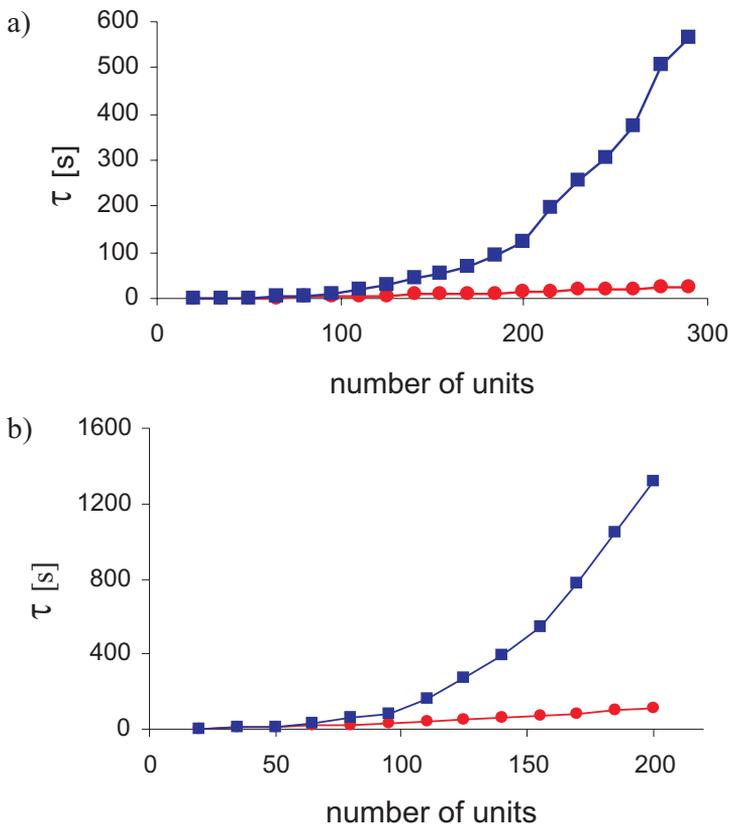


Fig. 3. CPU time required for solving SCF equations for STO-3G (a) and 6-31G (b) basis sets. The circles correspond to overall ECT while squares corresponds to HF reference. Calculations were performed at four nodes

The influence of diagonalization on CPU time is also strongly manifested in SCF process. The comparison of the overall SCF time for ECT with HF reference is shown in Figure 3a for STO-3G and Figure 3b for 6-31G. The curves show the same qualitative trends as those of Figure 2. Again, the overall ECT time depends linearly on the number of water molecules (number of AO). However, here the prefactor is greater. It is connected with formation of the Fock matrix that is more time consuming than in semiempirical methods used in initial guess routines. The formation of Fock matrix is well suited for parallelization, therefore, the CPU time reduction is more pronounced when going from one to eight nodes. Nevertheless, we have plotted the results obtained on 4 nodes, in order to have direct reference to Figure 2.

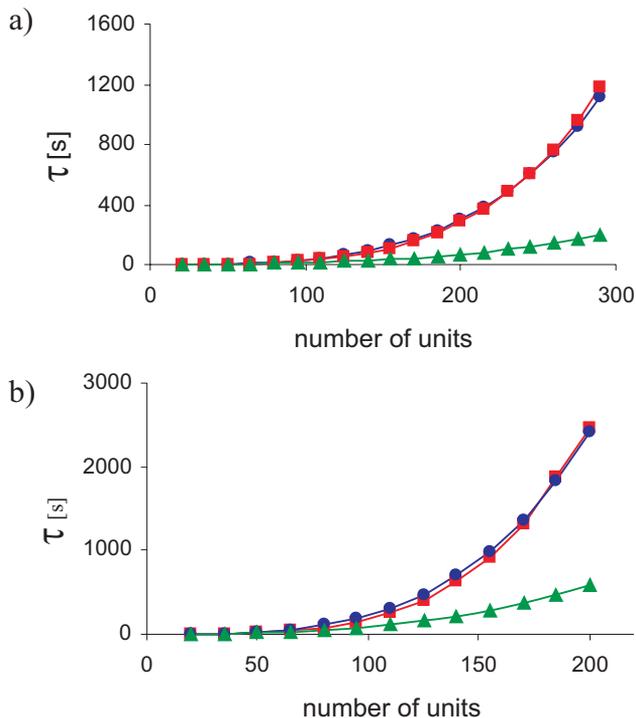


Fig. 4. CPU time required for computing TEI. The lines connecting circles, triangles and square correspond to step ECT, the overall ECT, and HF timings, respectively. Calculations were performed at four nodes. Parts (a) and (b) corresponds to STO-3G and 6-31G, respectively

The computation of TEI is especially well prepared for parallel performance. Every subset of TEI can be computed on different nodes. This is a great advantage for direct type of calculations since TEI are recalculated several times during the SCF process. Nevertheless, also in conventional type of calculations this can be a source of huge time savings. In Figure 4, we have compared the CPU time of TEI step for both

methods (ECT and HF). Part a (b) of Figure 4, illustrates results obtained during calculations performed on 4 nodes for STO-3G (6-31G).

The overall CPU time for ECT calculations (circles) is almost the same as for HF calculations (squares). It means that during the elongation process we are computing nearly the same number of TEI as in the HF scheme, though in a single step the number of TEI (triangles) is much lower than in HF calculations. Therefore, ECT calculations proceed longer than HF reference. Of course, it is connected with a hard disc capacity. The same quadratic dependency for both methods suggests that we have to limit the number of TEI in ECT scheme or to simplify the picture of interactions, otherwise the ECT method will never be linear. One possible solution is the interaction radius. It means that all TEI beyond this radius will be neglected. Another possibility is connected with introducing the classical multipoles for very distant interactions. In direct mode, application of quantum fast multipole method for continuous charge distribution is linear. However, this technique has a huge prefactor and huge memory requirements.

5. Conclusions and Future Prospects

In this paper we have presented a cutoff elongation technique and investigated its scaling properties. The elongation takes the advantage of the localized molecular orbitals. This allows to efficiently reduce the computational time. Such behavior of the elongation cutoff technique should be important in direct SCF calculations when the integrals are evaluated for every iteration step. The ECT method significantly reduces the prefactor in the quadratic scaling relation; the more nodes are used in calculations, the lower is prefactor.

The ECT removes the diagonalization bottleneck from the computational scheme. Therefore, generation of initial guess and solving SCF equations are linear with the system size. To achieve linear scaling in the whole ECT procedure, the problem of computing two-electron integrals should be overcome. Due to the fact that in a single ECT step much less integrals are computed than in the reference HF calculations, we can extend conventional type of calculations (TEI stored on a disc) to bigger molecular systems. By introducing interaction radii, a real linear scaling in the total CPU time can be obtained. Another way towards linear scaling is application of quantum fast multipole method to the elongation cutoff scheme in direct type of calculations or combined conventional/direct calculations. Taking into account that the overlap Gaussian distributions belonging to cut region are well separated from the active space distributions, such scheme should be very efficient (time and memory requirements). We are planning to incorporate both mentioned techniques to the elongation cutoff method.

The work was supported by Polish Ministry of Science and Higher Education (Project No. 1486/B/H03/2007/33). Computational grant No MNiSW/IBM_BC_HS21/UJ/074/2007 is greatly acknowledged.

References

- [1] Halgaker T., Jorgensen P., Olsen J.: *Molecular Electronic-Structure Theory*. Chichester, John Wiley & Sons LTD, 2000
- [2] Aoki Y., Imamura A.: *Local Density of States of Aperiodic Polymers using the Localized Orbitals from an ab initio elongation method*. J. Chem. Phys., vol. 97, 1992, 8432–8440
- [3] Deev V., Collins M. A.: *Approximate ab initio energies by systematic molecular Fragmentation*. J. Chem. Phys., vol. 122, 2005, 154102–154111
- [4] Exner T. E., Mezey P. G.: *The Field-Adapted ADMA Approach: Introducing Point Charges*. J. Phys. Chem. A, vol. 108, 2004, 4301–4309
- [5] Fedorov D. G., Kitaura K.: *On the Accuracy of the 3-Body Fragment Molecular Orbital Method (FMO) Applied to Density Functional Theory*. Chem. Phys. Lett., vol. 389, 2004, 129–134
- [6] Gu F. L., Aoki Y., Korchowiec J., Imamura A., Kirtman B.: *A New Localization Scheme for the Elongation Method*. J. Chem. Phys., vol. 121, 2004, 10385–10391
- [7] Imamura A., Aoki Y., Maekawa K.: *A Theoretical Synthesis of Polymers by Using Uniform Localization of Molecular Orbitals: Proposal of an Elongation Method*. J. Chem. Phys., vol. 95, 1991, 5419–5431
- [8] Jiang N., Ma J., Jiang Y.: *Electrostatic Field-Adapted Molecular Fractionation with Conjugated Caps for Energy Calculations of Charged Biomolecules*. J. Chem. Phys., vol. 124, 2006, 114112–114119
- [9] Korchowiec J., Gu F. L., Imamura A., Kirtman B., Aoki Y.: *Elongation Method with Cutoff Technique for Linear SCF Scaling*. Int. J. Quantum Chem., vol. 102, 2005, 785–794
- [10] Li S., Li W., Fang T.: *An Efficient Fragment-Based Approach for Predicting the Ground-State Energies and Structures of Large Molecules*. J. Am. Chem. Soc., vol. 127, 2005, 7215–7226
- [11] Makowski M., Korchowiec J., Gu F. L., Aoki Y.: *Efficiency and Accuracy of the Elongation Method as Applied to the Electronic Structure of Large Systems*. J. Comput. Chem., vol. 27, 2006, 1603–1619
- [12] White C. A., Johnson B. G., Gill P. M. W., Head-Gordon M.: *Linear Scaling Density Functional Calculations via the Continuous Fast Multipole Method*. Chem. Phys. Lett., vol. 253, 1996, 268–278
- [13] Burant J. C., Scuseria G. E., Frisch M. J.: *A Linear Scaling Method for Hartree-Fock Exchange Calculations of Large Molecules*. J. Chem. Phys., vol. 105, 1996, 8969–8972
- [14] Schwegler E., Challacombe M.: *Linear Scaling Computation of the Hartree-Fock Exchange Matrix*. J. Chem. Phys., vol. 105, 1996, 2726–2734
- [15] Salek P., Hyst S., Thygersen L., Jyrgensen P., Manninen P., Olsen J., Jansik B., Reine S., Pawlowski F., Tellgren E., Helgaker T., Coriani S.: *Linear-Scaling*

- Implementation of Molecular Electronic Self-Consistent Field Theory.* J. Chem. Phys., vol. 126, 2007, 114110–114116
- [16] Schmidt M. W., Baldrige K. K., Boatz J. A., Elbert S. T., Gordon M. S., Jensen J. H., Koseki S., Matsunaga N., Nguyen K. A., Su S., Windus T. L., Dupuis M., Montgomery J. A. Jr: *General Atomic and Molecular Electronic Structure system.* J. Comput. Chem., vol. 14, 1993, 1347–1363
- [17] Hehre W. J., Stewart R. F., Pople J. A.: *Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals.* J. Chem. Phys., vol. 51, 1969, 2657–2664
- [18] Ditchfield R., Hehre W. J., Pople J. A.: *Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules.* J. Chem. Phys., vol. 54, 1971, 724–728
- [19] Chaban G., Schmidt M. W., Gordon M. S.: *Approximate Second Order Method for Orbital Optimization of SCF and MCSCF Wavefunction.* Theoret. Chem. Acc., vol. 97, 1997, 88–95