



# A novel extremely localized molecular orbitals based technique for the one-electron density matrix computation

Alessandro Genoni, Michela Ghitti, Stefano Pieraccini, Maurizio Sironi

## ► To cite this version:

Alessandro Genoni, Michela Ghitti, Stefano Pieraccini, Maurizio Sironi. A novel extremely localized molecular orbitals based technique for the one-electron density matrix computation. Chemical Physics Letters, 2005, 415 (4-6), pp.256-260. 10.1016/j.cplett.2005.09.011 . hal-02196451

**HAL Id: hal-02196451**

**<https://hal.univ-lorraine.fr/hal-02196451>**

Submitted on 27 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# **A novel Extremely Localized Molecular Orbitals (ELMOs) based technique for the one-electron density matrix computation**

Alessandro Genoni <sup>(a)</sup>, Michela Ghitti <sup>(a)</sup>, Stefano Pieraccini <sup>(a)</sup>, Maurizio Sironi <sup>(a, b)</sup>.

(a) Dipartimento di Chimica Fisica ed Elettrochimica, Università degli Studi di Milano,  
Via Golgi 19, 20133 Milano, Italy.

(b) Centre for Biomolecular Interdisciplinary Studies and Industrial Applications, CISI,  
Via Fratelli Cervi 93, Palazzo LITA, 20090 Segrate (MI), Italy.

## **Abstract**

The “nearsightedness” of electronic structure is an underlying principle in many of the linear scaling methods recently developed to study large systems. Among them, there are strategies based on the transferring of orbitals strictly localized on molecular fragments, such as the Extremely Localized Molecular Orbitals (ELMOs). Unfortunately, due to the nonorthogonal nature of these orbitals, the density matrix calculation is computationally demanding, so preventing a straightforward application to very large molecules. In this letter, we show how this problem can be overcome by a proper application of the “Divide and Conquer” strategy to the ELMO approach.

## **Introduction**

The “nearsightedness” of electronic structure [1, 2], that is the weak interactions between very distant regions of a large system, plays a crucial role in theoretical chemistry. Actually, exploiting this principle, several researchers have devised many linear scaling methods [3].

---

· Fax: +39-0250314300.

E-mail address: [maurizio.sironi@unimi.it](mailto:maurizio.sironi@unimi.it) (M. Sironi).

Among them, a possible strategy is the “Divide and Conquer” (D&C) technique, initially developed by Yang [4] for the electron density and afterwards extended to the density matrix within the Hartree-Fock and the semiempirical MO approaches [5-7]. In these cases, after subdividing a large molecule into a set of relatively small subsystems with overlapping buffers, a proper electronic problem is solved for each region and then the total electron density (or density matrix) is assembled. A similar method, the Molecular Tailoring Approach devised independently by Gadre et al. [8-10], consists in three steps: 1) fragments determination from the parent molecule, 2) *ab initio* density matrix calculation at the Hartree-Fock (or MP2) level for every subunit and 3) construction of the overall density matrix using the constituent fragment density matrices. In this context the Adjustable Density Matrix Assembler (ADMA) strategy proposed by Exner and Mezey [11-13] is also noteworthy. This technique, which can be considered as an extension of the MEDLA approach [14], allows to obtain one-electron properties for macromolecules combining wisely the subsystem density matrices saved in a database previously determined by computations on smaller systems.

Another way of exploiting the nearsightedness of electronic structure to scale the computational cost is based on physically motivated reductions of the variational space, as the the ones recently performed by Janesco and Yaron in the construction of their Functional Group Basis Sets [15].

Finally, it is important to consider the approach based on the transferring of Localized Molecular Orbitals (LMOs) [16] which can be obtained by means of *a posteriori* [17] or *a priori* [18-22] techniques. The former consist in unitary transformations of canonical Hartree-Fock orbitals and preserve “orthogonalization tails” that have to be deleted in order to make the resulting orbitals transferable [23]. The latter need the definition of a localization scheme before the calculation so that the molecular orbitals are expanded on the only basis functions centred on the atoms belonging to preselected fragments. In this way the presence of tails is

avoided *a priori* and orbitals strictly localized on few atoms, namely Extremely Localized Molecular Orbitals (ELMOs), are obtained.

In our laboratory, we have developed an algorithm to determine ELMOs [24] mainly based on the Stoll theory [18]. Furthermore, by means of extensive studies [20, 24-28], we have shown their reliable transferability that especially allows to reproduce very well one-electron properties at the Hartree-Fock level (e.g. electron density and Molecular Electrostatic Potential).

It is worthwhile to note that both ELMOs and LMOs with deleted tails (LMOs-TD) are nonorthogonal orbitals. Hence, the calculation of the corresponding density matrix is not trivial and, for large systems, is very expensive in terms of CPU time. In order to reduce the computational cost, it is possible either to consider the orbitals as they were orthogonal or to orthogonalize them. However, in the first case a loss of accuracy naturally occurs, while the orthogonalization procedure becomes more and more demanding than using directly the nonorthogonal orbitals as the molecular size grows up.

So, in this paper, we will propose a new strategy for an efficient density matrix computation for large systems. In particular, we will take advantage of the extremely localized nature of the ELMOs combined with the D&C philosophy to obtain an approximate density matrix expression that permits a considerable saving of the computational resources retaining the accuracy of the results.

## Theory

Let us consider a  $2N$  electrons closed-shell system described by a single Slater determinant built up with a set of  $N$  transferred ELMOs  $\{\phi_i\}_{i=1}^N$ . The generic  $i$ -th ELMO is a linear combination of a preselected subset of all the atomic orbitals, namely

$$\phi_i(\mathbf{r}) = \sum_{\mu=1}^{NAO} C_{\mu i} \chi_{\mu}(\mathbf{r})$$

where, according to the adopted localization scheme, some  $C_{\mu i}$  coefficients are constrained to be zero while others have been variationally determined on model molecules. Details are thoroughly described in [18, 24] and references therein.

The corresponding ELMO electron density is given by :

$$\rho(\mathbf{r}) = \sum_{\mu, \nu=1}^{NAO} \chi_{\mu}(\mathbf{r}) \mathbf{D}_{\mu \nu} \chi_{\nu}(\mathbf{r})$$

where, due to the nonorthogonality of the orbitals, the density matrix element  $\mathbf{D}_{\mu \nu}$  is:

$$\mathbf{D}_{\mu \nu} = 2 \sum_{i, j=1}^N C_{\mu j} [\mathbf{S}^{-1}]_{ji} C_{\nu i}^* \quad (1)$$

with  $\mathbf{S}$  as the overlap matrix between the ELMOs.

In order to greatly reduce the computational cost associated with equation (1), we divide the molecule into  $N_s$  disjoint subunits so that their union provides the whole system. Owing to this fragmentation, a local basis set  $\beta_k$  is automatically assigned to the generic  $k$ -th subsystem ( $k = 1, \dots, N_s$ ).

According to the D&C strategy, we introduce the partition matrix  $\mathbf{P}^k$  for the  $k$ -th subunit:

$$\mathbf{P}_{\mu \nu}^k = \begin{cases} 1 & \text{if } \mu \in \beta_k \text{ and } \nu \in \beta_k \\ 1/2 & \text{if } (\mu \in \beta_k \text{ and } \nu \notin \beta_k) \text{ or } (\mu \notin \beta_k \text{ and } \nu \in \beta_k) \\ 0 & \text{if } \mu \notin \beta_k \text{ and } \nu \notin \beta_k \end{cases}$$

which satisfies the normalization condition

$$\sum_{k=1}^{N_s} \mathbf{P}_{\mu \nu}^k = 1 \quad \forall \mu, \nu$$

and allows to assemble the density matrix as a sum of contributions arising from the different subsystems:

$$\mathbf{D}_{\mu\nu} = \sum_{k=1}^{N_s} \mathbf{D}_{\mu\nu}^k \quad \text{with} \quad \mathbf{D}_{\mu\nu}^k = \mathbf{P}_{\mu\nu}^k \mathbf{D}_{\mu\nu}$$

The ‘divide’ step allows to introduce proper approximations in the evaluation of the  $\mathbf{D}_{\mu\nu}^k$  contributions, taking into the right account the extremely localized nature of the ELMOs. In particular, the elements  $\mathbf{D}_{\mu\nu}^k$  can be efficiently computed without considering the complete set of the  $N$  ELMOs, but involving only the orbitals localized on the  $k$ -th subsystem (“core ELMOs”) and the ones that significantly overlap with them (“buffer ELMOs”), i.e. orbitals that generally belong to the near subunits in the real space.

Therefore, for the  $k$ -th subsystem, we have

$$\mathbf{D}_{\mu\nu}^{k \text{ (approx)}} = 2 \mathbf{P}_{\mu\nu}^k \sum_{i,j=1}^{n_k} C_{\mu j} [\mathbf{S}_k^{-1}]_{ji} C_{\nu i}^* \quad (2)$$

with  $n_k$  as the total number of considered ELMOs (i.e. the core and buffer ELMOs) for the  $k$ -th subunit and  $\mathbf{S}_k$  as their overlap matrix. It is important to note that  $n_k$  is generally much lower than  $N$ , so the inversion of  $N_s$  matrices  $\mathbf{S}_k$  (with a mean dimension  $\langle n_k \rangle \ll N$ ) is clearly less demanding than a single inversion of the complete,  $N$  dimension and sparse matrix  $\mathbf{S}$  in equation (1).

In this way we realize the ‘conquer’ step to efficiently compute the electron density for large molecules, using transferred orbitals determined on model molecules. Of course, the present approach would also be applicable to any sets of Localized Nonorthogonal Orbitals, as the LMOs subjected to a tail deletion.

## Calculations

We will report some calculations performed on polypeptides to show the reliability and the capabilities of our strategy. To accomplish this, we have compared “reference” electron

densities, namely charge distributions obtained by means of an exact calculation of the density matrix elements (i.e. using equation (1)), with those deriving from the new approximate expression. In order to have a quantitative comparison, we have considered the similarity index  $L(a, a')$  introduced by Walker and Mezey [14], which compares point by point two charge distributions in the space bounded by two electron density isosurfaces characterized by the values  $a$  and  $a'$ .

In all the computations performed with our technique, the subsystems have been chosen to correspond to the aminoacids constituting the polypeptides (see Figure 1), in order to have a scheme which will be easily applicable also to a generic protein.

The ELMOs transferring has been carried out using DENPOL [29, 30], a program that allows to determine *ab initio* quality electron densities for polypeptides through a database of model molecules, suited to represent all the aminoacids, and a database of ELMOs determined on the model molecules. We have chosen the STO-4G basis set in order to limit the number of atomic orbitals for the very large systems studied in this paper.

To determine the set of buffer ELMOs for each subsystem, we have introduced a criterion based on the overlap between orbitals. In particular, a generic ELMO non-belonging to the  $k$ -th subunit is considered a buffer ELMO for this subunit if it significantly overlaps (i.e. the overlap integral is, in absolute value, greater or equal to an Overlap Threshold [OT]) with at least a non-negligible number of core ELMOs, number represented by the Significant Overlap Threshold (SOT). In order to fix the values for OT and SOT, we have carried out test calculations on the small polypeptide Gly<sub>10</sub> using two conformations, namely the  $\alpha$ -elix and the  $\beta$ -sheet structures, to allow for possible effects arising from different extents of overlap between ELMOs of different subsystems. Hence, analyzing Tables I and II, we have chosen SOT and OT equal to 6 and  $1 \cdot 10^{-3}$ , respectively. This choice reflects the need to have the

best agreement with the reference electron density keeping the number of buffer ELMOs as low as possible.

Furthermore, due to the relatively small size of this system (310 electrons and 237 basis functions), we have also performed Restricted Hartree-Fock (RHF) calculations at the STO-4G level to compute the similarity indexes between the RHF and the reference ELMO electron densities. As it can be observed from Table III, the charge distributions show a significant similarity that confirms the reliability of the ELMOs transferring, as already reported for other systems [20, 24-28].

We have then considered the case of the Gly<sub>100</sub> polypeptide (3010 electrons and 2307 basis functions) in the  $\alpha$ -elix,  $\beta$ -sheet and globular conformations. The last one (see Figure 2A) has been obtained from the first one by means of molecular dynamics. In particular, using the AMBER package [31], we have performed a 1 ns simulation at 400 K, in implicit solvent, exploiting the GB/SA model with a 20 Å cutoff distance and freezing all the bonds with the SHAKE algorithm. This conformation allowed us to evaluate the capabilities of the proposed strategy also in the case of very compact structures. As it can be clearly seen from Table IV, our method provides very good agreements with the reference electron densities and permits to greatly reduce the computational cost.

Finally, to test the reliability of the technique also in the case of polypeptides with side chains characterized by conformational freedom, we have taken into account the Leu<sub>50</sub> (3110 electrons and 2557 basis functions) and the Leu<sub>100</sub> (6210 electrons and 5107 basis functions) molecules. These two polypeptides have been considered in their  $\alpha$ -elix structure and in another conformation obtained after subjecting the molecules to the molecular dynamics simulation protocol previously used for the Gly<sub>100</sub> system. The resulting structures (see Figures 2B and 2C) are, for Leu<sub>50</sub>, a short coil staggered helix-turn-helix sequence and, for Leu<sub>100</sub>, a more globular conformation constituted by helix-coil and helix-turn motifs. Also for



these polypeptides we have a similarity always greater than 99 % with respect to the reference electron density and we have a formidable reduction in CPU time for the density matrix calculation, in particular for the Leu<sub>100</sub> conformations (see Table IV).

It has to be noted that, in Tables I, II and IV, we have reported only the index  $L(0.001,10)$  because it gives a global similarity measure between two charge distributions. During our test calculations, we have also computed the  $L$  index using different isosurface values [e.g.  $L(0.1,10)$ ,  $L(0.01,0.1)$ ,  $L(0.001,0.01)$ ], which take into account more localized regions of real space. Also in these cases, for all the molecules considered in Table IV, we have always achieved values greater than 99 %.

## Conclusion

The transferring of ELMOs, previously determined on model molecules, is a way to assemble *ab-initio* quality electron densities for very large molecules. However, the ELMOs nonorthogonality introduces a non-trivial task in the exact evaluation of the density matrix elements. In this letter, we have proposed an approximate expression for the computation of these elements, combining the extreme localization of the involved orbitals with the “Divide & Conquer” philosophy. Furthermore, from preliminary tests, we have observed that this approach does not affect the accuracy of the results and allows to obtain sensitive reductions in terms of CPU time. The calculations on some model polypeptides showed that the approach can be also applied to large systems such as proteins.

## Acknowledgement

We would like to thank Dr. Arianna Fornili, Dr. Monica Civera and Dr. Fausto Cargnoni for helpful discussions.

## References

1. W. Kohn, Phys. Rev. Lett. 76 (1996) 3168.
2. W. Kohn, Rev. Mod. Phys. 71 (1999) 1253.
3. S. Goedecker, Rev. Mod. Phys. 71 (1999) 1085.
4. W. Yang, Phys. Rev. Lett. 66 (1991) 1438.
5. W. Yang, T.-S. Lee, J. Chem. Phys. 103 (1995) 5674.
6. S. L. Dixon, K. M. Merz Jr., J. Chem. Phys. 104 (1996) 6643.
7. S. L. Dixon, K. M. Merz Jr., J. Chem. Phys. 107 (1997) 879.
8. S. R. Gadre, R. N. Shirsat, A. C. Limaye, J. Phys. Chem. 98 (1994) 9165.
9. K. Babu, S. R. Gadre, J. Comput. Chem. 24 (2003) 484.
10. K. Babu, V. Ganesh, S. R. Gadre, N. E. Ghermani, Theor. Chem. Acc. 111 (2004) 255.
11. P. G. Mezey, J. Math. Chem. 18 (1995) 141.
12. T. E. Exner, P. G. Mezey, J. Phys. Chem. A 106 (2002) 11791.
13. T. E. Exner, P. G. Mezey, J. Comput. Chem. 24 (2003) 1980.
14. P. D. Walker, P. G. Mezey, J. Am. Chem. Soc. 116 (1994) 12022.
15. B. G. Janesko, D. Yaron, J. Chem. Theory Comput. 1 (2005) 267.
16. W. Hierse, E. B. Stechel, Phys. Rev. B 54 (1996) 16515.
17. J. Pipek, P. G. Mezey, J. Chem. Phys. 90 (1989) 4916.
18. H. Stoll, G. Wagenblast, H. Preuss, Theoret. Chim. Acta 57 (1980) 169.
19. Z. Szekeres, P. R. Surján, Chem. Phys. Lett. 369 (2003) 125.
20. M. Sironi, A. Famulari, Theor. Chem. Acc. 103 (2000) 417.
21. G. Nàray-Szabò, Comput. Chem. 24 (2000) 287.
22. G. F. Smits, C. Altona, Theoret. Chim. Acta 67 (1985) 461.
23. G. Alagona, S. Campanile, C. Ghio, A. Giolitti, S. Monti, Theor. Chem. Acc. 110 (2003) 446.

24. A. Fornili, M. Sironi, M. Raimondi, J. Mol. Struct. (Theochem) 632 (2003) 157.
25. M. Sironi, A. Famulari, M. Raimondi, S. Chiesa, J. Mol. Struct. (Theochem) 529 (2000) 47.
26. E. Burrelli, M. Sironi, Theor. Chem. Acc. 112 (2004) 247.
27. A. Genoni, M. Sironi, Theor. Chem. Acc. 112 (2004) 254.
28. A. Genoni, A. Fornili, M. Sironi, J. Comput. Chem. 26 (2005) 827.
29. M. Ghitti, A. Genoni, M. Civera, A. Fornili, S. Pieraccini, M. Sironi, Proceedings of the International Karlsruhe Nanoscience Workshop, University of Karlsruhe, 2005.
30. M. Ghitti, A. Genoni, M. Sironi, in preparation.
31. D. A. Case et al., AMBER 7, University of California, San Francisco, CA, 2002.

## Figures

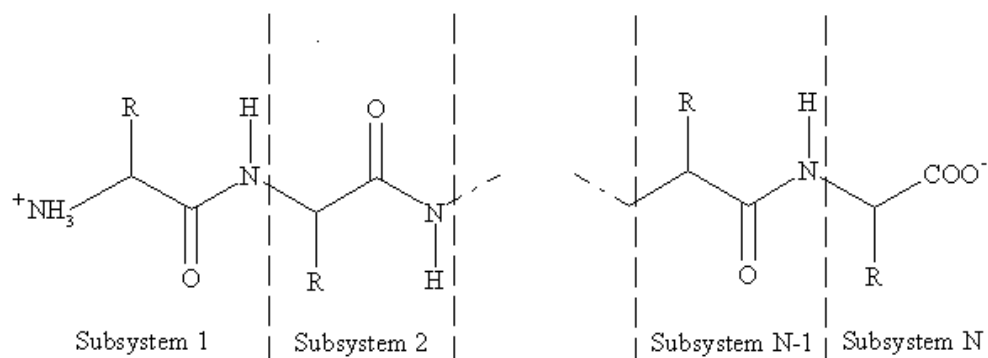


Fig. 1. Partition of the polypeptides in subsystems.

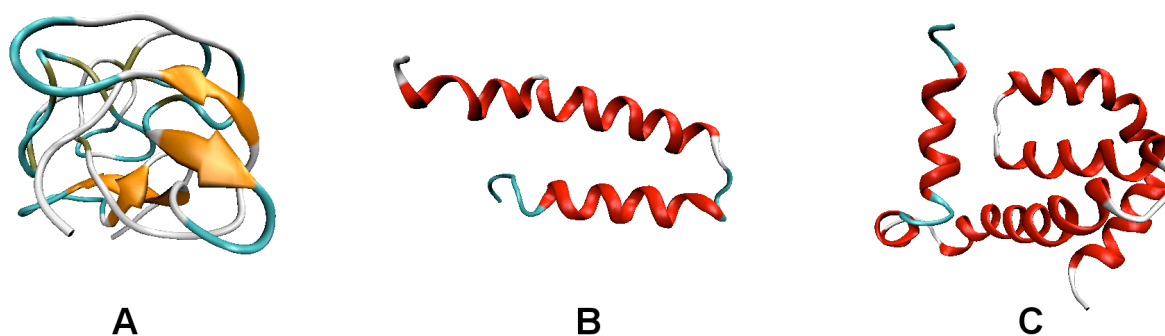


Fig. 2. Conformations obtained after the molecular dynamics simulations described in the text: A) Gly<sub>100</sub>, B) Leu<sub>50</sub>, C) Leu<sub>100</sub>.

Table I - Gly<sub>10</sub> ( $\alpha$ -elix): similarity index values beetwen the reference electron density and the approximate charge distributions (the acronyms are described in the text).

OT	SOT=6		SOT=8		SOT=10	
	L (0.001, 10) %	$\langle n_k \rangle^{(a)}$	L (0.001, 10) %	$\langle n_k \rangle^{(a)}$	L (0.001, 10) %	$\langle n_k \rangle^{(a)}$
$5 \cdot 10^{-5}$	100.0	77	100.0	68	99.98	59
$5 \cdot 10^{-4}$	99.98	53	99.92	43	99.45	33
$1 \cdot 10^{-3}$	99.91	43	99.66	32	99.13	26
$1 \cdot 10^{-2}$	98.46	19	98.23	17	97.84	16

(a) Average number of ELMOs (core plus buffer) associated to each subunit.

Table II - Gly<sub>10</sub> ( $\beta$ -sheet): similarity index values between the reference electron density and the approximate charge distributions (the acronyms are described in the text).

OT	SOT=6		SOT=8		SOT=10	
	L (0.001, 10) %	$\langle n_k \rangle^{(a)}$	L (0.001, 10) %	$\langle n_k \rangle^{(a)}$	L (0.001, 10) %	$\langle n_k \rangle^{(a)}$
$5 \cdot 10^{-5}$	100.0	39	99.99	37	99.98	33
$5 \cdot 10^{-4}$	99.93	33	99.93	32	99.89	26
$1 \cdot 10^{-3}$	99.93	33	99.93	29	99.82	25
$1 \cdot 10^{-2}$	99.19	21	99.09	20	98.17	16

(a) Average number of ELMOs (core plus buffer) associated to each subunit.

Table III - Gly<sub>10</sub>: similarity index values between the Restricted Hartree-Fock electron density and the reference ELMO charge distribution.

Molecule	L ( 0.001, 10 ) %	L ( 0.1, 10 ) %	L ( 0.01, 0.1 ) %	L ( 0.001, 0.01 ) %
Gly <sub>10</sub> ( $\alpha$ -elix)	98.31	98.64	98.41	98.08
Gly <sub>10</sub> ( $\beta$ -sheet)	98.83	98.90	98.85	98.80

Table IV – Gly<sub>100</sub>, Leu<sub>50</sub> and Leu<sub>100</sub> polypeptides: similarity index values and comparison of CPU times for the density matrix calculation when different approaches are used.

Molecule	L (0.001, 10) %	$\langle n_k \rangle^{(a)}$	CPU time (s)		
			Approximation <sup>(c)</sup>	Nonorthogonal Orbitals <sup>(d)</sup>	Orthogonalized Orbitals <sup>(e)</sup>
Gly <sub>100</sub> ( $\alpha$ -elix)	99.89	49	74	295	695
Gly <sub>100</sub> ( $\beta$ -sheet)	99.93	34	68	299	694
Gly <sub>100</sub> (MD) <sup>(b)</sup>	99.86	39	71	305	699
Leu <sub>50</sub> ( $\alpha$ -elix)	99.87	95	102	429	924
Leu <sub>50</sub> (MD) <sup>(b)</sup>	99.91	65	83	500	923
Leu <sub>100</sub> ( $\alpha$ -elix)	99.86	97	563	6658	22604
Leu <sub>100</sub> (MD) <sup>(b)</sup>	99.89	64	521	6559	22853

(a) Average number of ELMOs (core plus buffer) associated to each subunit; (b) Conformation obtained by means of the molecular dynamics simulation described in the text; (c) Density matrix computed exploiting the new strategy; (d) Density matrix calculated traditionally; (e) The ELMOs are Löwdin orthogonalized before the density matrix computation.