ORIGINAL PAPER

Minimal set of molecule-adapted atomic orbitals from maximum overlap criterion

Dariusz Szczepanik · Janusz Mrozek

Abstract The criterion of maximum overlap with the canonical free-atom orbitals is used to construct a minimal set of molecule-intrinsic orthogonal atomic orbitals that resemble the most their promolecular origins. Partial atomic charges derived from population analysis within representation of such molecule-adopted atomic orbitals are examined on example of first-row hydrides and compared with charges from other methods. The maximum overlap criterion is also utilized to approximate the exact free-atom orbitals obtained from ab initio calculations in any arbitrary basis set and the influence of the resulting fitted canonical atomic orbitals on properties of molecule-adopted atomic orbitals is briefly discussed.

Keywords Minimal basis set · Atomic orbitals · Maximum overlap criterion · Partial atomic charges · Molecule-intrinsic orbitals

1 Introduction

It has recently been argued [1] within the framework the Orbital Communication Theory (OCT) [2–4] of the chemical bond that the minimum basis (MB) of atomic orbitals (AO) occupied in the promolecular system of non-interacting atoms gives rise to the most intuitively "chemical" account of the bond covalency/ionicity and gives understanding of diverse factors conditioning the efficiency of the AO interactions. It has turned out that, in the case of molecules with typical covalent bonds, amount of information about electron localization calculated within the MB-set of atomic orbitals is (to a certain degree) invariant with respect to basis-set enlargement [1].

D. Szczepanik (✉) · J. Mrozek
Department of Computational Methods in Chemistry, Faculty of Chemistry,
Jagiellonian University, R. Ingardena St. 3, 30-060 Cracow, Poland
e-mail: szczepad@chemia.uj.edu.pl

The maximum overlap criterion (MOC) [5,6] has been proposed to numerically con-firm that the MOC-approximated (fitted) small basis and the reference minimum set of AOs indeed generate practically identical OCT-indices.

The criterion of maximum overlap has also been used [7] to approximate the wave function calculated within a general (extended) basis set by the molecular orbitals (MO) calculated within the external minimal basis (EMB), i.e. Huzinagas MINI basis functions [8]. It has been demonstrated that the electron population analysis (EPA) based on the resulting EMB-orbitals gives rise to definitely more reliable and sensible partial atomic charges (PICs) than the corresponding Mulliken's and Löwdins popula-tion analyses (MPA [9–12] and LPA [13], respectively), especially if calculated within extended sets of basis functions. Moreover, EMB-charges have turned out to be compa-rable to PACs calculated within representation of natural atomic orbitals (NAO) [14], as well as preserve the proper convergence profile when one systematically enlarges the basis set.

The main goal of this paper is to utilize the MOC-scheme (in the form presented in [7]) to generate the molecule-intrinsic minimal-basis (IMB) orbitals that resembles the most (in the least-squares sense) the canonical free-atom orbitals (FAO) of the relevant system promolecule.

2 Method details

Within the framework of molecular-orbital theory the one-determinant wavefunction of the closed-shell groundstate molecular system of bonded atoms is fully determined by the subspace of occupied MOs, $|\psi^o\rangle$,

$$|\psi^o\rangle = |\chi\rangle \mathbf{C}^o_\chi,\tag{1}$$

calculated within a general set of non-orthogonal basis functions (not necessarily centered on particular atoms), $|\chi\rangle$, i.e. with overlap matrix $\mathbf{S}_\chi = \langle\chi|\chi\rangle \neq \mathbf{1}$. In the above equation rectangular matrix $\mathbf{C}^o_\chi$ collects the relevant linear-combination (LC) coefficients for occupied molecular orbitals. The complementary subspace of virtual MOs, $|\psi^v\rangle$, is therefore defined as follows:

$$|\psi^v\rangle = |\chi\rangle \mathbf{C}^v_\chi.\tag{2}$$

The dimension of the Hilbert space of all molecular orbitals, n, is the simple sum of dimensions of occupied and virtual MO-subspaces, $n = n^o + n^v$; furthermore, n^o can be expressed by the number of core-MOs (n^o_c) and valence-MOs (n^o_v), $n^o = n^o_c + n^o_v$. The number of basis functions $|\chi\rangle$ we will denote by n_χ.

Now, let us consider the corresponding promolecular system given by the rele-vant non-bonded atoms at their molecular positions, calculated using the ROHF/GVB method with explicitly assumed fractional occupation numbers to assure sphericall symmetry of all atoms. For separate atom X, the subspace of the resultant occupied canonical free-atom orbitals, $|\varphi^0_X\rangle$,

$$|\varphi^0_X\rangle = |\chi\rangle \mathbf{B}^0_X.\tag{3}$$

determine the minimal set of atomic orbitals strictly assigned to the atom in question; here, matrix $\mathbf{B}_X^0$ collects the linear-combination coefficients related to the subspace of occupied FAOs. Therefore, grouping the relevant LC-matrices of all constituent atoms in the molecule gives rise to the promolecular minimal basis of canonical atomic orbitals, $|\varphi^0\rangle$, within representation of $|\chi\rangle$,

$$|\varphi^0\rangle = |\chi\rangle\mathbf{B}_\chi^0, \quad \mathbf{B}_\chi^0 = (\mathbf{B}_{X_1}^0|\cdots|\mathbf{B}_{X_N}^0). \tag{4}$$

The number of all promolecular FAOs, b, can be regarded as the simple sum of the number of core-orbitals $|\varphi_c^0\rangle$, b_c, and valence-orbitals $|\varphi_v^0\rangle$, b_v, $b = b_c + b_v$. At this point it has to be mentioned that, generally, $b_c = n_c^o$ but $b_v > n_v^o$ and hence $b > n^o$, i.e. the number of FAOs is always greater than the number of occupied MOs in the molecule and therefore the definition of molecular intrinsic AOs requires using also some orbitals from virtual MO-subspace. However, the choice of appropriate $n_v^v = b - n^o$ orbitals from n_χ-dimensional subspace $|\psi^v\rangle$ is somewhat ambiguous.

Recently, has been proposed an efficient algorithm for determining the n_v^v-dimensional valence-subspace $|\varphi_v^0\rangle$ within the framework of quasi-atomic minimal basis orbitals (QUAMBO) [15,16] which constitute the molecule-intrinsic minimal basis of valence atomic orbitals, $|\varphi_v\rangle$. However, the use of the maximum overlap criterion allows one to easily generalize the original procedure to cover also the core atomic orbitals. In short, the alternative algorithm that transforms occupied non-orthogonal FAOs directly to orthogonal IMB-orbitals, $|\varphi^o\rangle \to |\varphi\rangle$, can be summarized in the following two steps:

1. Evaluating the missing $n_\chi \times n_v^v$ LC-matrix $\bar{\mathbf{C}}_\chi^v$ from the virtual MO-subspace $\mathbf{C}_\chi^v$ as follows:

$$\bar{\mathbf{C}}_\chi^v = \mathbf{C}_\chi^v \left(\mathbf{U}_\varphi^v\right)_{1 \to n_v^v}, \tag{5}$$

where the unitary matrix $\mathbf{U}_\varphi^v$ diagonalize the appropriate metric matrix $\mathbf{W}_\varphi^v$,

$$\mathbf{W}_\varphi^v = \mathbf{V}_\varphi^{v\dagger}\mathbf{V}_\varphi^v = \mathbf{U}_\varphi^v\mathbf{w}_\varphi^v\mathbf{U}_\varphi^{v\dagger}, \quad \text{and} \quad \mathbf{V}_\varphi^v = \mathbf{B}_\chi^{o\dagger}\mathbf{S}_\chi\mathbf{C}_\chi^v. \tag{6}$$

More details about the virtual valence MO-subspace given by $\bar{\mathbf{C}}_\chi^v$ one can find in the original work [15]. The subspaces of occupied and virtual valence-MOs can be collectively written as $|\bar{\psi}\rangle$ and the corresponding $n_\chi \times b$ matrix $\bar{\mathbf{C}}_\chi = (\mathbf{C}_\chi^o|\bar{\mathbf{C}}_\chi^v)$.

2. Using the maximum overlap criterion,

$$\text{tr}\langle\varphi^0|\varphi\rangle = \text{maximum}, \tag{7}$$

to find the molecule-intrinsic orthogonal atomic orbitals $|\varphi\rangle$ that resemble the most canonical atomic orbitals $|\varphi^0\rangle$. Then, according to [7] we get:

$$|\varphi\rangle = |\chi\rangle\mathbf{B}_\chi, \quad \mathbf{B}_\chi = \mathbf{T}\left(\mathbf{T}^\dagger\mathbf{T}\right)^{-1/2}, \tag{8}$$

where

$$\mathbf{T} = \mathbf{B}_\chi^{0\dagger}\mathbf{S}_\chi\bar{\mathbf{C}}_\chi. \tag{9}$$

The presented procedure gives rive to the set of molecule-adapted atomic orbitals preserving the most the inner-shell AOs and with the effective valence-AOs that reveal the influence of the net of chemical bonds in the molecule. In accordance to (8), one can express the contracted set of molecular orbitals $|\bar{\psi}\rangle$ within the representation of atomic orbitals $|\varphi\rangle$ as follows:

$$|\bar{\psi}\rangle = |\varphi\rangle\langle\varphi|\bar{\psi}\rangle = |\varphi\rangle\mathbf{B}_\chi^\dagger\langle\chi|\chi\rangle|\bar{\mathbf{C}}_\chi = |\varphi\rangle\mathbf{B}_\chi^\dagger\mathbf{S}_\chi\bar{\mathbf{C}}_\chi = |\varphi\rangle\bar{\mathbf{C}}_\varphi, \tag{10}$$

and

$$\bar{\mathbf{C}}_\varphi = \left(\bar{\mathbf{C}}_\varphi^o | \bar{\mathbf{C}}_\varphi^v\right). \tag{11}$$

Construction of such set of IMB-orbitals every-time involves additional calculations of orbitals $|\varphi^0\rangle$ for separated atoms. However, the maximum overlap criterion can also be successfully utilized to smooth the way of the whole procedure. In particular, matrices $\mathbf{B}_X^0$ for the consecutive elements X from the periodic table, calculated with high accuracy within the extended basis set $|\chi\rangle$ can either reside on disk or can be stored in the quantum-chemistry program. Then, according to the original MOC-scheme [6], the required LC-coefficients within arbitrary set of basis functions $|\chi'\rangle$ one can calculate (exact to basis set superposition error) as follows:

$$\mathbf{B}_{X'}^0 = \mathbf{S}_{X'}^{-1/2}\mathbf{U}_{X'}^0\left(\mathbf{U}_{X'}^{0\dagger}\mathbf{U}_{X'}^0\right)^{-1/2}, \tag{12}$$

where

$$\mathbf{U}_{X'}^0 = \mathbf{S}_{X'}^{-1/2}\boldsymbol{\Delta}_{X',X}\mathbf{B}_X^0 \tag{13}$$

In the foregoing equations labels X and X$'$ conventionally refer to atom-X-centered functions $|\chi_X\rangle$ and $|\chi_X'\rangle$, respectively, and $\boldsymbol{\Delta}_{X',X}$ stands for the rectangular matrix of overlap integrals between both these basis subsets,

$$\boldsymbol{\Delta}_{X',X} = \{\Delta_{\sigma,\varrho} = \langle\chi_\sigma'|\chi_\varrho\rangle\}. \tag{14}$$

3 Numerical results

Numerical results presented in this section were obtained using the special program written by authors to perform MOC-calculations as well as the ab initio quantum chemistry package GAMESS [17], the Natural Bond Orbitals software NBO 5.0 [18] and the molecular visualization program MOLDEN [19]. Molecular calculations were carried out for species diversified by means of chemical bond character, i.e. the following first-row hydrides: LiH, BeH$_2$, BH$_3$, CH$_4$, NH$_3$, H$_2$O and HF (for comparison calculations for LiF were performed as well) at the RHF theory level and using experimental geometries whereas calculations for free atoms were performed using the ROHF/GVB method with explicitly assumed fractional occupation numbers to assure

sphericall symmetry of atoms (more details can be found in "Further Information" section of GAMESS documentation [17]).

3.1 IMB-orbitals from MOC

First, let us analyze sensitivity to the size of basis sets and convergence properties of partial atomic charges involving IMB-orbitals,

$$Q_X^{IMB} = Z_X - N_X^{IMB} = Z_X - 2 \sum_{\sigma \in X} \sum_{i=1}^{n^o} \left| \left(\bar{\mathbf{C}}_\varphi^o \right)_{\sigma,i} \right|^2, \tag{15}$$

(here Z_X stands for the atomic number of atom X and N_X^{IMB} is the relevant electron population on atom X). The above are compared with PACs from the following standard electron population analyses (-PA):

- Mulliken-PA (Q_X^{MPA}) [9–12],
- Löwdin-PA (Q_X^{LPA}) [13],
- *Natural*-PA (Q_X^{NPA}) [14],
- electrostatic potential (ESP)-derived atomic charges (Q_X^{ESP}) [20],
- the recently proposed population analysis involving the EMB of AOs from the criterion of maximum overlap with the Huzinagas MINI basis functions, (Q_X^{EMB}) [7].

The list of basis sets used involves: STO-3G ($\mathfrak{a}$), 6-31G ($\mathfrak{b}$), 6-311G ($\mathfrak{c}$), 6-311G** ($\mathfrak{d}$), 6-311++G** ($\mathfrak{e}$), 6-311++G**^{d} ($\mathfrak{f}$), 6-311++G**^{t} ($\mathfrak{g}$); here superscripts d and t respectively stand for doubled and tripled sets of relevant polarization functions on each atom.

Table 1 collects partial atomic charges of non-hydrogen atoms for all hydrides under consideration and PACs of fluorine atom in LiF molecule, calculated using various PA-schemes and the most extended basis set $|\chi^\mathfrak{g}\rangle$. Even a cursory analysis of Q_X^{LPA} seems to reaffirm the well known shortcoming of population analysis based on symmetrically

Table 1 Partial atomic charges of central atoms of first-row hydrides and LiF molecule calculated at their experimental geometries and using various population-analysis schemes

Molecule	Q_X^{MPA}	Q_X^{LPA}	Q_X^{NPA}	Q_X^{ESP}	Q_X^{EMB}	Q_X^{IMB}
LiH	0.4372	0.1938	0.8468	0.8444	0.6315	0.6226
BeH$_2$	0.4766	0.1004	1.2289	0.7282	1.0475	1.2172
BH$_3$	0.1025	−0.0415	0.4192	0.6295	0.0062	0.0452
CH$_4$	−0.0481	−0.0203	−0.7255	−0.6015	−0.7099	−0.5660
NH$_3$	−0.4729	0.0403	−1.0342	−0.9266	−1.0978	−0.7883
H$_2$O	−0.5704	0.0004	−0.9276	−0.7379	−1.0603	−0.7663
HF	−0.4080	−0.0510	−0.5609	−0.4565	−0.6680	−0.5000
LiF	−0.7547	−0.4594	−0.9781	−0.8849	−0.9932	−0.9450

Method: RHF/6-311++G**(with triple set of polarization functions)

Fig. 1 Basis-set dependence of partial atomic charges from various methods calculated at the RHF theory level for carbon atom in CH$_4$ molecule at experimental geometry
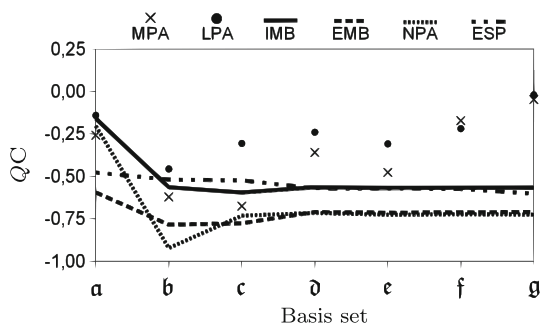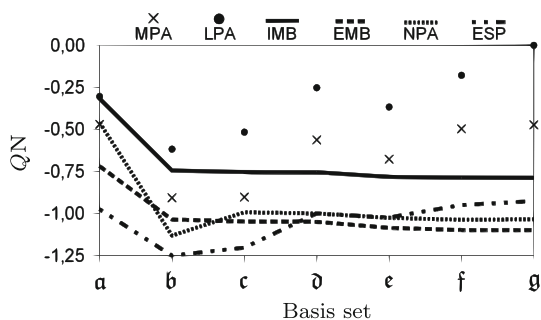

Fig. 2 Basis-set dependence of partial atomic charges from various methods calculated at the RHF theory level for nitrogen atom in NH$_3$ molecule at experimental geometry


orthogonalized atomic orbitals [21], i.e. within extended basis sets (especially with diffuse and polarization functions) LPA gives rise to completely unreliable atomic charges (usually Q_X^{LPA} converges to nearly-zero values as the basis set is enlarged). Although PACs calculated within the framework of MPA assume more sensible values (with the exception of CH$_4$ molecule), they still seem to be somewhat underestimated with regard to the reference atomic charges Q_X^{NPA} and Q_X^{ESP}. Moreover, atomic charges Q_X^{MPA} are much more sensitive to the basis set choice than Q_X^{LPA} [22] and, as proved by Ruedenberg [23], in the limit of a complete basis set they can actually exhibit any value between $\pm\infty$. Details of characteristic basis-set dependencies of Q_X^{LPA}, Q_X^{MPA} as well as charges from other methods for two representative cases, CH$_4$ and NH$_3$, can be investigated in Figs. 1 and 2.

For the most part atomic charges Q_X^{NPA}, Q_X^{ESP}, Q_X^{EMB} and Q_X^{IMB} sharply differ from MPA- and LPA-derived ones; they converge to definitely more reasonable values and thus they seem to display adequate picture of molecular electronic structure in atomic resolution. This is evident especially if one compares results for LiF; only atomic charges from NPA, ESP, EMB and IMB correctly predict predominatingly ionic (90–100 %) character of the chemical bond Li–F. However, as follows from a thorough analysis of results from Table 1, in most cases partial atomic charges Q_X^{NPA} and Q_X^{EMB} reveal quite good correlation and both tend to predict slightly more polarized bond densities than Q_X^{IMB} and Q_X^{ESP}.

This tendencies were also probed using quantities δQ_H involving charges on hydrogen atoms from various PA-methods, Q_H, and electronegativity differences in Pauling scale [24], $\Delta\xi_{XH}$:

Fig. 3 Distances between relative bond-ionicity descriptors of X–H based on Pauling's electronegativity differences and those derived from partial atomic charges calculated using various schemes. Method: RHF/6-311++G** (with triple set of polarization functions), experimental geometries
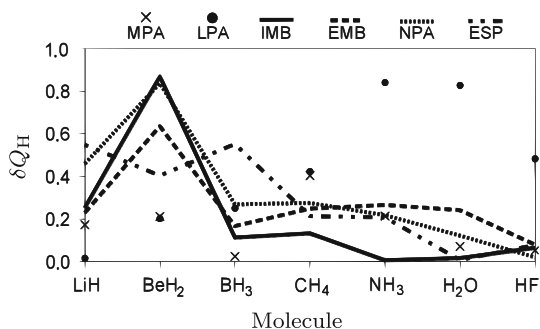
$$\delta Q_{\mathrm{H}} = \left| |(Q_{ref})^{-1} Q_{\mathrm{H}}| - (\Delta\xi_{ref})^{-1}\Delta\xi_{\mathrm{XH}} \right|. \tag{16}$$

Here quantities Q_{ref} and $\Delta\xi_{ref}$ refer to diatomic with possibly the most polarized chemical bond which in this case is LiF. Therefore, for particular molecule, δQ_{H} measures the distance between relative polarizations of the chemical bond H–X evaluated using calculated partial atomic charges and Pauling's electronegativity differences. Results are presented in Fig. 3.

Since natural atomic charges are known to be quite well-correlated with electronegativity differences [14] let us confine to compare results only for NPA- and IMB-charges. It follows directly from Fig. 3 that electronic structures predicted by the newly proposed IMB-orbitals are generally closer to those resulting from Pauling's electronegativity differences ($\overline{\delta Q_{\mathrm{H}}}$, for IMB and NPA assume 0.2072 and 0.3140, respectively), especially for nonmetal hydrides ($\overline{\delta Q_{\mathrm{H}}}(\mathrm{IMB}) = 0.0661$ and $\overline{\delta Q_{\mathrm{H}}}(\mathrm{NPA}) = 0.1805$). Just as natural atomic charge, IMB-charge display a far more ionic bond Be-H than predicted by other methods while in the case of LiH the latter is closer to predictions from $\Delta\xi_{\mathrm{LiH}}$. However, as follows also from Table 1, the most conspicuous difference between atomic charges involving maximum overlap with the minimum set of AOs (IMB and EMB) and PACs from NPA (as well as ESP) concerns BH_3 molecule. In particular, within description involving IMB-orbitals boron atom seem to exhibit almost purely nonmetalic character (since $\Delta\xi_{\mathrm{BH}} \approx 0.16$ and δQ_{H} is comparable with CH_4, NH_3, H_2O and HF) whereas within ESP and NPA approach boron atom reveals rather metalic/semimetalic character. A similar conclusion can be drawn after analysis of quantity $\delta Q'_{\mathrm{H}}$ (Fig. 4) being an analogue of (16) but measuring difference between relative polarizations of bond H–X evaluated from various PACs, relative Pauling's electronegativity differences (ELN) and relative "charges" from ESP-approach:

$$\delta Q'_{\mathrm{H}} = \left\| (Q_{ref})^{-1} Q_{\mathrm{H}} \right| - \left| \left(Q_{ref}^{\mathrm{ESP}}\right)^{-1} Q_{\mathrm{H}}^{\mathrm{ESP}} \right| \right\|. \tag{17}$$

Table 2 collects electron populations for three different sets of minimal-basis AOs, i.e. those from IMB, EMB and NPA formalisms. One should notice that within subset of minimal basis natural atomic orbitals Rydberg-type functions are excluded and thus orbital populations usually do not sum up exactly to the overall number of electrons,
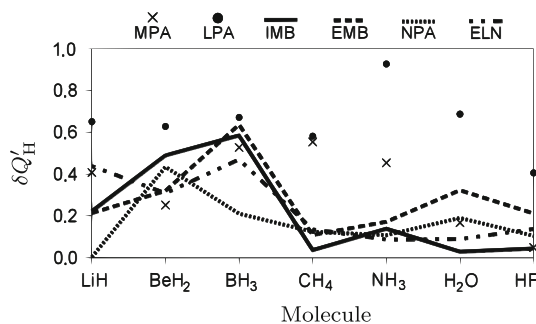
Fig. 4 Distances between relative bond-ionicity descriptors of X–H based on ESP-derived atomic charges and those involving PACs from various population-analysis schemes as well as Pauling's electronegativity differences (ELN). Method: RHF/6-311++G** (with triple set of polarization functions), experimental geometries

Table 2 Electron populations of IMB- and EMB-orbitals as well as the minimal-basis subset of NAOs for central atoms of first-row hydrides calculated at their experimental geometries

Molecule	N_{1s}^{IMB}	N_{2s}^{IMB}	$N_{2p_x}^{IMB}$	$N_{2p_y}^{IMB}$	$N_{2p_z}^{IMB}$
LiH	2.0000	0.3774	0.0000	0.0000	0.0000
BeH$_2$	2.0000	0.7828	0.0000	0.0000	0.0000
BH$_3$	2.0000	0.9705	0.9922	0.9922	0.0000
CH$_4$	2.0000	1.1119	1.1514	1.1514	1.1514
NH$_3$	2.0000	1.4134	1.2641	1.2641	1.8467
H$_2$O	2.0000	1.6502	2.0000	1.3850	1.7312
HF	2.0000	1.8365	2.0000	2.0000	1.6635
Molecule	N_{1s}^{EMB}	N_{2s}^{EMB}	$N_{2p_x}^{EMB}$	$N_{2p_y}^{EMB}$	$N_{2p_z}^{EMB}$
LiH	1.9794	0.3891	0.0000	0.0000	0.0000
BeH$_2$	1.9796	0.9729	0.0000	0.0000	0.0000
BH$_3$	1.9731	1.2225	0.9054	0.9054	0.0000
CH$_4$	1.9790	1.5548	1.0587	1.0587	1.0587
NH$_3$	1.9919	1.8672	1.2069	1.2069	1.8250
H$_2$O	1.9969	1.9601	2.0000	1.3714	1.7319
HF	1.9988	1.9893	2.0000	2.0000	1.6798
Molecule	N_{1s}^{NPA}	N_{2s}^{NPA}	$N_{2p_x}^{NPA}$	$N_{2p_y}^{NPA}$	$N_{2p_z}^{NPA}$
LiH	1.9995	0.1324	0.0000	0.0000	0.0000
BeH$_2$	1.9998	0.6709	0.0000	0.0000	0.0000
BH$_3$	1.9997	0.9549	0.8085	0.8085	0.0000
CH$_4$	1.9997	1.1545	1.1857	1.1857	1.1857
NH$_3$	1.9997	1.4960	1.3358	1.3358	1.8398
H$_2$O	1.9999	1.7434	1.9925	1.4531	1.7168
HF	2.0000	1.9041	1.9963	1.9963	1.6536

Method: RHF/6-311++G**(with triple set of polarization functions)

$\sum_{\mu}^{MB} N_{\mu}^{NAO} \approx N$. At the first glance it is clear that representation of IMB-orbitals assures the best separation of lone-pair and core orbitals; comparable results we obtain using NAOs. But it is somehow puzzling that within representation of EMB-orbitals lone-pairs exhibit better separation than core-orbitals. Furthermore, electron populations of orbital 2s, N_{2s}^{EMB}, assume significantly greater values than N_{2s}^{IMB} and N_{2s}^{NPA} (this feature is characteristic for "physically"-orthogonalized atomic orbitals [25]). In the case of "p"-type orbitals NAO-populations sharply differ from those based on minimal-basis AOs involving MOC-approach only for borane molecule. Below we present averaged (over all molecules) differences between the relevant orbital populations from two different MB-representations, $\overline{\Delta N}_{\mu}$, that concisely recaps Table 2:

– $\overline{\Delta N}_{\mu}^{IMB/EMB}$ = 0.0145e(1s), 0.2590e(2s), 0.0412e(2p),
– $\overline{\Delta N}_{\mu}^{NPA/IMB}$ = 0.0002e(1s), 0.0941e(2s), 0.0485e(2p),
– $\overline{\Delta N}_{\mu}^{NPA/EMB}$ = 0.0142e(1s), 0.2714e(2s), 0.0657e(2p).

3.2 MOC-approximated FAOs

Let us investigate the quality of free-atom AOs within arbitrary basis set (from $\mathfrak{a}$ to $\mathfrak{g}$) obtained using the MOC-method from canonical AOs calculated at ROHF/GVB theory level within three reference set of functions, $\mathfrak{a}$, $\mathfrak{d}$ and $\mathfrak{g}$. Table 3 presents averaged (over all orbitals and all available basis sets $\mathfrak{a}$–$\mathfrak{g}$) overlap integrals between reference atomic orbitals calculated using ab initio method, $|\chi^{ref}\rangle$, and those obtained using maximum overlap criterion, $|\chi'^{x}\rangle$, $\mathcal{S}_{\mu}^{ref} = \langle \chi_{\mu}'^{x} | \chi_{\mu}^{ref} \rangle$. Averaged quantities with subscripts were calculated only for basis sets less ($|\chi'^{x}\rangle = \mathfrak{a}, \mathfrak{b}, \mathfrak{c}$) or more accurate ($|\chi'^{x}\rangle = \mathfrak{e}, \mathfrak{f}, \mathfrak{g}$) than the reference set of functions $|\chi^{ref}\rangle = \mathfrak{d}$.

In accordance to expectations, results from Table 3 clearly indicate that using the maximum overlap criterion with the reference FAOs calculated within set of functions $\mathfrak{g}$ allows one to reproduce canonical orbitals for free-atoms within any basis set $\mathfrak{a}$–$\mathfrak{f}$ with the highest accuracy; typical deviation of MOC-derived orbitals from ab initio FAOs is 10^{-5}. Using the reference basis-set with only one set of polarization funtions and

Table 3 Average (over all basis sets $\mathfrak{a}$–$\mathfrak{g}$) overlap integrals between MOC-approximated free-atom orbitals and the corresponding canonical FAOs obtained from ab initio calculations using reference basis sets $\mathfrak{d}$ and $\mathfrak{g}$

Atom	$\bar{\mathcal{S}}^{\mathfrak{g}}$	$\bar{\mathcal{S}}^{\mathfrak{d}}$	$\bar{\mathcal{S}}_{\mathfrak{a}-\mathfrak{c}}^{\mathfrak{d}}$	$\bar{\mathcal{S}}_{\mathfrak{e}-\mathfrak{g}}^{\mathfrak{d}}$	$\bar{\mathcal{S}}^{\mathfrak{a}}$
H	1.00000	1.00000	1.00000	0.99999	0.98295
Li	0.99995	0.99995	0.99991	0.99998	0.98907
Be	0.99995	0.99994	0.99991	0.99997	0.99140
B	0.99994	0.99982	0.99993	0.99971	0.97287
C	1.00000	0.99991	0.99999	0.99984	0.98912
N	0.99999	0.99995	1.00000	0.99989	0.99438
O	0.99999	0.99992	1.00000	0.99984	0.99306
F	0.99999	0.99993	1.00000	0.99986	0.99276
Atom-averaged	0.99998	0.99993	0.99997	0.99988	0.98820

Table 4 Comparison of partial atomic charges calculated within representation of exact IMB-orbitals and using various basis sets with the relevant PACs calculated within representation of IMB-orbitals involving MOC-fitted FAOs from reference basis sets ∂ and 𝔤

Molecule	PIC	𝔞	𝔟	𝔠	∂	𝔢	𝔣	𝔤
LiH	Q_{Li}^{IMB}	0.4764	0.5797	0.6185	0.6190	0.6216	0.6227	0.6226
	$\mathcal{Q}_{Li}^{\mathfrak{g}}$	0.4764	0.5847	0.6189	0.6193	0.6213	0.6225	0.6226
	$\mathcal{Q}_{Li}^{\partial}$	0.4764	0.5848	0.6188	0.6190	0.6190	0.6194	0.6206
BeH$_2$	Q_{Be}^{IMB}	1.0223	1.1754	1.1998	1.2126	1.2154	1.2168	1.2172
	$\mathcal{Q}_{Be}^{\mathfrak{g}}$	1.0223	1.1823	1.2016	1.2144	1.2153	1.2169	1.2172
	$\mathcal{Q}_{Be}^{\partial}$	1.0223	1.1821	1.2001	1.2126	1.2125	1.2141	1.2145
BH$_3$	Q_{B}^{IMB}	0.1491	0.0013	0.0034	0.0301	0.0407	0.0435	0.0452
	$\mathcal{Q}_{B}^{\mathfrak{g}}$	0.1490	−0.0101	0.0087	0.0349	0.0411	0.0435	0.0452
	$\mathcal{Q}_{B}^{\partial}$	0.1490	−0.0098	0.0038	0.0301	0.0311	0.0341	0.0356
CH$_4$	Q_{C}^{IMB}	−0.1603	−0.5646	−0.5957	−0.5644	−0.5689	−0.5662	−0.5660
	$\mathcal{Q}_{C}^{\mathfrak{g}}$	−0.1604	−0.5908	−0.6000	−0.5695	−0.5679	−0.5666	−0.5660
	$\mathcal{Q}_{C}^{\partial}$	−0.1604	−0.5862	−0.5946	−0.5644	−0.5644	−0.5631	−0.5625
NH$_3$	Q_{N}^{IMB}	−0.3174	−0.7444	−0.7541	−0.7545	−0.7829	−0.7878	−0.7883
	$\mathcal{Q}_{N}^{\mathfrak{g}}$	−0.3174	−0.7471	−0.7590	−0.7597	−0.7824	−0.7882	−0.7883
	$\mathcal{Q}_{N}^{\partial}$	−0.3174	−0.7431	−0.7535	−0.7545	−0.7748	−0.7808	−0.7809
H$_2$O	Q_{O}^{IMB}	−0.2604	−0.7102	−0.7178	−0.7326	−0.7596	−0.7659	−0.7663
	$\mathcal{Q}_{O}^{\mathfrak{g}}$	−0.2604	−0.7123	−0.7215	−0.7359	−0.7592	−0.7662	−0.7663
	$\mathcal{Q}_{O}^{\partial}$	−0.2604	−0.7099	−0.7177	−0.7326	−0.7524	−0.7595	−0.7596
HF	Q_{F}^{IMB}	−0.1553	−0.4585	−0.4738	−0.4841	−0.4972	−0.4995	−0.5000
	$\mathcal{Q}_{F}^{\mathfrak{g}}$	−0.1553	−0.4600	−0.4753	−0.4851	−0.4971	−0.4996	−0.5000
	$\mathcal{Q}_{F}^{\partial}$	−0.1553	−0.4596	−0.4738	−0.4841	−0.4932	−0.4960	−0.4963
Averaged	$\bar{\Delta}_{X}^{IMB}$	0.3447	0.0379	0.0168	0.0156	0.0025	0.0003	0.0000
Over all	$\bar{\Delta}_{X}^{\mathfrak{g}}$	0.3447	0.0305	0.0140	0.0127	0.0029	0.0001	0.0000
Molecules	$\bar{\Delta}_{X}^{\partial}$	0.3447	0.0325	0.0170	0.0156	0.0074	0.0046	0.0043

All molecules at their experimental geometries

without diffuse ones, ∂, gives rise to very similar results provided that $|\chi'^x\rangle = 𝔞, 𝔟, 𝔠$; approximation of more accurate FAOs ($|\chi'^x\rangle = 𝔢, 𝔣, 𝔤$) leads to slightly worse results, with deviation assuming 10^{-4}. Obviously, for reference FAOs calculated with very poor set o functions the MOC-procedure does not allow one to obtain atomic orbitals of satisfying quality (atom-averaged deviation is up to 10^{-2}).

It was of our special interest to probe to what extent using MOC-approximated FAOs in construction of molecule-adopted AOs influences the results of EPA. Table 4 collects partial atomic charges Q_X^{IMB}, calculated within the representation of IMB-orbitals involving exact FAOs and various basis sets (𝔞–𝔤), as well as corresponding PICs, $\mathcal{Q}_X^{\partial}$ and $\mathcal{Q}_X^{\mathfrak{g}}$, involving MOC-approximated FAOs calculated using extended basis sets ∂ and 𝔤, respectively. Quantities $\bar{\Delta}_X$ stand for averaged (over all molecules excluding BH$_3$) differences between IMB-charges calculated within particular basis set and the corresponding atomic charge calculated using exact FAOs and within basis set 𝔤; for negative exact PACs the respective differences were multiplied by factor −1.

As follows from Table 4, using approximated free-atom orbitals in construction of IMB-orbitals has no significant influence on calculated IMB-atomic charges, especially for calculations involving the minimal set of basis functions. Indeed, the first column of numbers in Table 4 clearly indicates that within very poor basis set $\mathfrak{a}$ atomic charges are particularly insensitive to quality of FAOs used in MOC-procedure (it becomes obvious if one recalls the results from Table 3).

On average, atomic charges $\mathcal{Q}_X^{\mathfrak{d}}$ and $\mathcal{Q}_X^{\mathfrak{g}}$ deviate from the corresponding exact charges Q_X^{IMB} by $0.0029e$ and $0.0020e$, respectively and thus, in the view of population-analysis accuracy, they give rise to the same conclusions about the electronic structure of molecules. However, a more discerning comparison of atomic charges Q_X^{IMB} and $\mathcal{Q}_X^{\mathfrak{g}}$ reveals that the latter ones are usually closer to the corresponding exact IMB-charges (calculated within set of functions $\mathfrak{g}$). On the other hand, atomic charges $\mathcal{Q}_X^{\mathfrak{d}}$ deviate significantly from the exact charges only in the case of more accurate calculations (if we exclude BH_3 molecule and take into consideration only basis set of TZV-type the average difference between the most exact values of $\mathcal{Q}_X^{\mathfrak{d}}$ and $Q_{mathrm X}^{\mathrm{IMB}}$ assumes $0.0043e$, e.i. about 0.6 %). Indeed, convergence profiles that emerge from values of $\bar{\Delta}_X$ allow one to draw the conclusion that the more accurate are canonical FAOs used to construct a set of IMB-orbitals the closer to the exact values (in a sense of basis-set completeness) are the resulting IMB-charges calculated within any arbitrary basis set.

4 Summary

In this work we have introduced and briefly examined a simple method of generating a minimal set of molecule-adopted atomic orbitals. Contrary to the previously proposed method involving a reference set of (external) minimal-basis orbitals (e.g. Huzinaga's MINI basis set) [7], in this approach we have used the criterion of maximum overlap to the set of free-atom orbitals obtained from ab initio calculations (using the same set of basis functions) for the corresponding system promolecule. Hence, the resulting minimal-basis orbitals are intrinsic for individual molecules and consequently exhibit appropriate convergence properties as the number of basis functions used in calculations increases.

It has also been demonstrated that the MOC-scheme can be successfully utilized to approximate canonical free-atom orbitals within any arbitrary basis set using one-off calculated FAOs of high quality (by default stored on disk); therefore, generating of IMB-orbitals does not require every-time calculations of promolecular systems. Moreover, partial atomic charges calculated within representation of such molecule-adapted atomic orbitals tend to converge noticeably faster to their exact values in the limit of complete basis set.

References

1. R.F. Nalewajski, D. Szczepanik, J. Mrozek, J. Math. Chem. **50**, 1437 (2012)
2. R.F. Nalewajski, *Information Theory of Molecular Systems* (Elsevier, Amsterdam, 2006)
3. R.F. Nalewajski, *Information Origins of the Chemical Bond* (Nova Science Publishers, Hauppauge, 2010)
4. R.F. Nalewajski, D. Szczepanik, J. Mrozek, Adv. Quantum Chem. **61**, 1 (2011)
5. A. Golebiewski, Trans. Faraday Soc. **57**, 1849 (1961)
6. R.F. Nalewajski, A. Golebiewski, Chem. Phys. Lett. **29**, 441 (1974)
7. D. Szczepanik, J. Mrozek, Comput. Theor. Chem. **996**, 103 (2012)
8. S. Huzinaga, J. Andzelm, M. Klobukowski, E. Radzio-Andzelm, Y. Sakai (eds.), *Gaussian Basis Sets for Molecular Calculations* (Elsevier Science Ltd, Amsterdam, 1984)
9. R.S. Mulliken, J. Chem. Phys. **23**, 1833 (1955a)
10. R.S. Mulliken, J. Chem. Phys. **23**, 1841 (1955b)
11. R.S. Mulliken, J. Chem. Phys. **23**, 2338 (1955c)
12. R.S. Mulliken, J. Chem. Phys. **23**, 2343 (1955d)
13. P.O. Löwdin, J. Chem. Phys. **18**, 365 (1950)
14. A.E. Reed, R.B. Weinstock, F. Weinhold, J. Chem. Phys. **83**, 735 (1985)
15. W.C. Lu, C.Z. Wang, M.W. Schmidt, L. Bytautas, K.M. Ho, K. Ruedenberg, J. Chem. Phys. **120**, 2629 (2004)
16. W.C. Lu, C.Z. Wang, M.W. Schmidt, L. Bytautas, K.M. Ho, K. Ruedenberg, J. Chem. Phys. **120**, 2638 (2004)
17. M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S. Su, T.L. Windus, M. Dupuis, J.A. Montgomery, J. Comput. Chem. **14**, 1347 (1993)
18. E.D. Glendening, J.K. Badenhoop, A.E. Reed, J.E. Carpenter, J.A. Bohmann, C.M. Morales, F. Weinhold, *NBO 5.0* (Theoretical Chemistry Institute, University of Wisconsin, Madison, 2001). 2001
19. G. Schaftenaar, J.H. Noordik, J. Comput.-Aided Mol. Des. **14**, 123 (2000)
20. C. Chipot, B. Maigret, J.L. Rivail, H.A. Sheraga, J. Phys. Chem. **96**, 10276 (1992)
21. F. Jensen, *Introduction to Computational Chemistry*, 2nd edn. (Wiley, West Sussex, 2007)
22. E.R. Davidson, S. Chakravorty, Theor. Chim. Acta **83**, 319 (1992)
23. R.S. Mulliken, W.C. Ermler, *Diatomic Molecules: Results of Ab Initio Calculations* (Academic, New York, 1977)
24. L. Pauling, J. Am. Chem. Soc. **54**, 3570 (1932)
25. D. Szczepanik, J. Mrozek, Chem. Phys. Lett. **512**, 157 (2012)