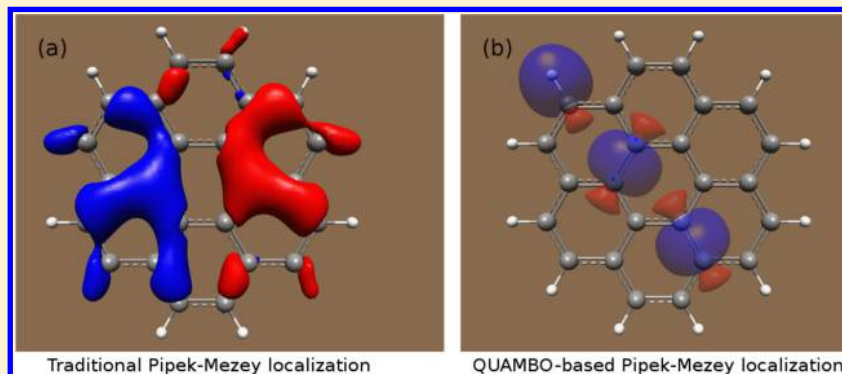# Near Equivalence of Intrinsic Atomic Orbitals and Quasiatomic Orbitals

Tomasz Janowski*

Department of Chemistry and Biochemistry Fulbright College of Arts and Sciences, University of Arkansas, Fayetteville, Arkansas 72701, United States

(a) Traditional Pipek-Mezey localization

(b) QUAMBO-based Pipek-Mezey localization

ABSTRACT: A direct relationship between the Intrinsic Atomic Orbitals (IAO) method (Knizia, G. *J. Chem. Theory Comp.* **2013**, *9*, 4834−4843) and earlier work on the same topic, quasiatomic minimal basis set orbitals (QUAMBO) (Lu, W. C.; Wang, C. Z.; Schmidt, M. W.; Bytautas, L.; Ho, K. M.; Ruedenberg *J. Chem. Phys.* **2004**, 2629) and later modifications (quasiatomic orbitals, QUAO) is investigated. It will be demonstrated mathematically that IAOs are almost identical to the original formulation of QUAMBOs and span the same space as a later QUAO modification. The construction of QUAOs involves minimization of a functional that requires matrix diagonalization, or singular matrix decomposition, while the IAO method provides a direct solution by projections. As a byproduct of this proof, it will be shown that (a) under mild conditions a simpler projection yields identical IAOs and (b) an alternative proof is obtained that IAOs span the full space of molecular orbitals if they are linearly independent. Utilization of QUAMBOs as the defining basis set results in rock-solid numerical stability of Pipek−Mezey localization and Mulliken or Löwdin population analysis in very large systems. The charges do not depend on the basis set used, as already shown by Knizia for smaller systems. In this paper, more difficult cases of large semiperiodic systems with strong linear dependency are tested, and it is shown that QUAMBOs perform extremely well.

1. INTRODUCTION

The interpretation of the wave function is of fundamental importance for all branches of chemistry. Qualitative understanding of the internal workings of molecules is needed to make quick predictions about bond stability, chemical reactivity, and various physical properties, such as local partial charges or multipole moments. Chemists use various one-electron models to interpret molecular wave function such as bonding and antibonding orbitals, lone pairs, hybridization, net charges on atomic centers, etc. Surprisingly, quantum mechanics does not provide any direct recipe for chemical interpretations. In the most popular approximation, a single determinant wave function, only the total N-electron wave function is uniquely determined, but the one-electron orbitals that make up the total wave function can be arbitrarily mixed with each other, only the space spanned by them is well-defined. The most direct solution to the Hartree−Fock equations is the one that diagonalizes the Fock matrix and this solution defines canonical orbitals. These orbitals have well-defined energies but are usually delocalized all over a molecule. They cannot be used for interpreting local quantities,

such as charges, bonds, etc. Canonical orbitals are useful for considering electron detachments, attachments, and excitation. Nonoccupied (virtual) orbitals are almost completely arbitrary, and their interpretative value is even less: not only spatial distribution of charge is arbitrary but also most of the orbital energies, due to the presence of Rydberg states. Various localization techniques have been developed to facilitate interpretation of molecular bonding, but each of them has its own disadvantages. The focus of this paper will be on the Pipek−Mezey localization scheme.[1] It produces one-electron orbitals where σ and π bonds are separated, which is often useful. The disadvantage of this approach is that the localization relies on orbital population analysis that is basis set dependent, and it breaks down if a large basis set is employed. In extreme cases the localization does not converge and the resulting orbitals are just as delocalized as the original canonical orbitals, a case discussed later in this paper. Knizia[2] has already demonstrated that the method can be stabilized using IAOs,

and this will be explored further in this paper for more difficult cases of extended systems.

The most straightforward route to alleviate the interpretative problems that involve properties that use basis functions in their definition is to design a minimal, atomic-like basis set, which spans the same space as the set of occupied orbitals obtained in a large basis set, but which has the same number of basis functions and similar spatial properties as the minimal set of atomic orbitals of free atoms. There has been significant effort to date to come up with such a minimal description of the basis space, many of the methods are quite successful.[2−10] The aim of this short paper is to explore the relationship between two particular methods, the quasiatomic minimal basis orbital (QUAMBO) method,[5] and a method proposed by Knizia,[2] the Intrinsic Atomic Orbitals method (IAO). It will be demonstrated that the sets of orbitals obtained by both methods is quite similar, while the recently published improvement of the original QUAMBO method[10] (QUAO) produces the same space as the IAO approach (although the individual orbitals may be slightly different). In this context, the method by Knizia can be considered a shortcut that produces the same solution as the QUAO recipe. Also, further simplifications to the recipe given by Knizia are possible. In the end, it will be demonstrated that the application of the original QUAMBOs dramatically improves the stability of the Pipek−Mezey (PM) localization. This was already demonstrated by Knizia for small molecules,[2] who first used QUAOs (IAOs) as defining orbitals in the PM localization. Here, more difficult cases are tested in which not only the stability with respect to the choice of the basis set but also with respect to linear dependency in the basis set is explored.

2. THEORY

The QUAMBO method developed by Lu et al.[5] is based on two simple conditions: (a) to build a set of K functions $|\tilde{\phi}_i\rangle$ that are as similar as possible to K atomic orbitals (AOs) of free atoms, denoted here as $|\phi_i\rangle$, and (b) they are built as a linear combination of the occupied molecular orbitals $|\psi_i\rangle$ plus some contribution from the virtual space ($|\bar{\psi}_a\rangle$). The virtual contribution is needed to ensure that the dimensionality of QUAMBO space is the same as the dimensionality of the original AOs:

$$|\tilde{\phi}_\mu\rangle = \sum_{i=1}^{n} X_{i\mu}|\psi_i\rangle + \sum_{a=n+1}^{K} X_{a\mu}|\bar{\psi}_a\rangle \tag{1}$$

where n is the number of molecular orbitals (MOs) and K is the number of AOs. In order to keep the development simple and to avoid introduction of additional projection operators, we will assume that all the entities involved, that is, MOs, virtual orbitals and AOs are expressed in some common, presumably very large basis set $|\mu\rangle$ of dimensionality $N \gg K$. The orbitals $|\bar{\psi}_a\rangle$ will be expressed as a linear combination of $N − n$ virtual orbitals $|\bar{\psi}_a\rangle$. From a purely mathematical point of view there is complete freedom in selecting appropriate underlying atomic orbitals, which can be either high level ab initio orbitals of free atoms, or as in ref 2, the MINAO basis set, which is a subset of orbitals defined in cc-pVTZ basis set and consists of spherically averaged free-atom Hartree−Fock orbitals. The same set of AOs will be adopted in this paper, with added simplification that the original MINAO basis set is immediately projected onto the basis set of interest in which the target wave function is expressed. A particular selection of the atomic orbitals will

result in different numerical results, but the general formulas and relationships derived are the same.

Imposing condition (a) enforces the spatial properties of QUAMBO while condition (b) ensures that the space spanned by MO will be contained in QUAMBO, if the transformation matrix $\mathbf{X}$ is not singular. Condition (a) is expressed mathematically as

$$\sum_{\mu=1}^{K} \langle \tilde{\phi}_\mu - \phi_\mu | \tilde{\phi}_\mu - \phi_\mu \rangle = \min \tag{2}$$

where the minimum is being searched with respect to the elements of the transformation matrix $\mathbf{X}$ and virtual orbitals $|\bar{\psi}_a\rangle$.

The minimization is performed in two steps. First, minimization with respect to elements of $\mathbf{X}$ is performed, while $|\bar{\psi}_a\rangle$ is kept fixed and yet undetermined. After taking derivatives of eq 2 and equating the result to zero, one gets the expressions for the matrix elements $\mathbf{X}$. After substituting these elements back to eq 2, one obtains:

$$|\tilde{\phi}_\mu\rangle = \sum_{i=1}^{n} |\psi_i\rangle\langle\psi_i|\phi_\mu\rangle + \sum_{a=n+1}^{K} |\bar{\psi}_a\rangle\langle\bar{\psi}_a|\phi_\mu\rangle \tag{3}$$

This formula is a sum of two projections (a) onto the occupied space and (b) another projection onto subspace of the virtual space, yet to be determined. Introducing the notation $O = \sum_i |\psi_i\rangle \langle\psi_i|$ and $V = \sum_a |\bar{\psi}_a\rangle \langle\bar{\psi}_a|$, the above formula can be written as

$$|\tilde{\phi}_\mu\rangle = O|\phi_\mu\rangle + V|\phi_\mu\rangle \tag{4}$$

Note the similarity of this formula and eq 2 in ref 2, which reads

$$|\tilde{\phi}_\mu\rangle = O\tilde{O}|\phi_\mu\rangle + (1 − O)(1 − \tilde{O})|\phi_\mu\rangle \tag{5}$$

with the following definitions:[2]

$$O = \sum_i |\psi_i\rangle\langle\psi_i|, \quad \tilde{O} = \sum_{ij} |\tilde{\psi}_i\rangle(\mathbf{S}_\psi^{-1})_{ij}\langle\tilde{\psi}_j|, \quad |\tilde{\psi}_i\rangle = \mathcal{P}|\psi_i\rangle \tag{6}$$

and

$$\mathcal{P} = \sum_{\mu\nu} |\phi_\mu\rangle(S_\phi^{-1})_{\mu\nu}\langle\phi_\nu|, \quad (\mathbf{S}_\psi)_{ij} = \langle\tilde{\psi}_i|\tilde{\psi}_j\rangle,$$

$$(\mathbf{S}_\phi)_{\mu\nu} = \langle\phi_\mu|\phi_\nu\rangle \tag{7}$$

In the above, the operator O is a projector onto the space of occupied orbitals, and the operator $\mathcal{P}$ projects onto the space of AOs. Following Knizia,[2] the $\tilde{\psi}_n$ orbitals will be referred to as the "depolarized" orbitals, since they are obtained by projecting the exact solution ψ_n onto the space of atomic orbitals, which does not contain polarization basis functions. The operator $\tilde{O}$ is a projection onto the space of depolarized occupied molecular orbitals.

The proof of equivalence will be completed if we can show that $O\tilde{O} = O$ and $(1 − O)(1 − \tilde{O}) = V$. In the first step, it will be shown that $O\tilde{O} = O$. The $\mathcal{P}$ operator can be expressed using orthogonalized atomic orbitals $\bar{\phi}$:

$$\mathcal{P} = \sum_{\mu=1}^{K} |\bar{\phi}_\mu\rangle\langle\bar{\phi}_\mu| \tag{8}$$

Then, $|\tilde{\psi}_i\rangle = \mathcal{P}|\psi_i\rangle$ and $\tilde{O} = \sum_{ij} \mathcal{P}|\psi_i\rangle(\mathbf{S}^{-1})_{ij}\langle\psi_j|\mathcal{P}$. Here, S_ψ is denoted by $\mathbf{S}$.

The first term in the formula 5 becomes

$$O\tilde{O} = O \sum_{jk} \mathcal{P}|\psi_j\rangle(\mathbf{S}^{-1})_{jk}\langle\psi_k|\mathcal{P}$$

$$= \sum_{ijk} |\psi_i\rangle\langle\psi_i|\mathcal{P}|\psi_j\rangle(\mathbf{S}^{-1})_{jk}\langle\psi_k|\mathcal{P} \tag{9}$$

Using the fact that $\mathcal{P}\mathcal{P} = \mathcal{P}$, we have

$$\sum_{ijk} |\psi_i\rangle\langle\psi_i|\mathcal{P}\mathcal{P}|\psi_j\rangle(\mathbf{S}^{-1})_{jk}\langle\psi_k|\mathcal{P} = \sum_{ijk} |\psi_i\rangle\langle\tilde{\psi}_i|\tilde{\psi}_j\rangle(\mathbf{S}^{-1})_{jk}\langle\tilde{\psi}_k| = \tag{10}$$

$$\sum_{ijk} |\psi_i\rangle S_{ij}(\mathbf{S}^{-1})_{jk}\langle\tilde{\psi}_k| = \sum_{ik} |\psi_i\rangle\delta_{ik}\langle\tilde{\psi}_k| = \sum_i |\psi_i\rangle\langle\tilde{\psi}_i| = O\mathcal{P} \tag{11}$$

We conclude that if the operator $O\tilde{O}$ is restricted to the space of the AOs where $|\phi_\mu\rangle = \mathcal{P}|\phi_\mu\rangle$, we obtain

$$O\tilde{O}|\phi_\mu\rangle = O|\phi_\mu\rangle \tag{12}$$

Note that this condition holds exactly only if the minimum basis set is a subspace of the main basis, which was not assumed to be true in ref 2, or if the minimum basis is directly expressed in the large basis set, as in this work. In the limit of a very large basis set this approximation is very accurate. We can replace eq 5 by a simpler expression:

$$|\tilde{\phi}_\mu\rangle = O|\phi_\mu\rangle + (1 - O)(1 - \tilde{O})|\phi_\mu\rangle \tag{13}$$

Similar considerations as above reduce the virtual part of eq 5 to

$$(1 - O)(1 - \tilde{O})|\phi_k\rangle = (1 - O - \tilde{O} + O\tilde{O})|\phi_k\rangle = \tag{14}$$

$$(1 - O - \tilde{O} + O)|\phi_k\rangle = (1 - \tilde{O})|\phi_k\rangle \tag{15}$$

The second projection $(1 - O)$ is therefore redundant provided that $|\phi_\mu\rangle = \mathcal{P}|\phi_\mu\rangle$.

The QUAMBO method requires finding the optimum set of virtual orbitals that define the operator V. The manipulations in ref 5 lead to a functional that has to be maximized with respect to the set of orbitals $\bar{\psi}_a$:

$$F = \sum_{a=n+1}^{K} \langle\bar{\psi}_a|\tilde{\mathcal{P}}|\bar{\psi}_a\rangle = \max \tag{16}$$

In the equation above, the $\tilde{\mathcal{P}}$ operator is a modified version of the projection operator, where the orthogonalized functions are replaced by nonorthogonal (original) AOs:

$$\tilde{\mathcal{P}} = \sum_\mu |\phi_\mu\rangle\langle\phi_\mu| \tag{17}$$

Recent modification of the QUAMBO method,[10] referenced here as QUAO, replaces this operator by a true projector $\mathcal{P}$, eq 8, where the atomic orbitals are being replaced by their symmetrically orthogonalized counterparts $|\bar{\phi}_i\rangle$:

$$F' = \sum_{a=n+1}^{K} \langle\bar{\psi}_a|\mathcal{P}|\bar{\psi}_a\rangle = \max, \quad \mathcal{P} = \sum_\mu |\bar{\phi}_\mu\rangle\langle\bar{\phi}_\mu| \tag{18}$$

The above equations state that the optimum virtual orbitals are those with the greatest norm after being projected onto the space of atomic orbitals (i.e., depolarized). Intuitively, it appears plausible that they are identical to $(1 - \tilde{O})|\phi\rangle$, as the $1 - \tilde{O}$ operator might involve the depolarized virtual orbitals as well.

The above condition can be recast into another form in which the matrix $\mathbf{P}$ of the operator $\mathcal{P}$ has a block diagonal form, that is, the space of virtual orbitals is divided into a subspace of optimal virtual orbitals $|\tilde{\psi}_a\rangle$ and a subspace of all other virtual orbitals $|\tilde{\psi}_b\rangle$ with

$$\langle\tilde{\psi}_b|\mathcal{P}|\bar{\psi}_a\rangle = 0 \tag{19}$$

The above condition defines a stationary point and does not ensure the maximum. The easiest way of obtaining the maximum is to diagonalize $\mathbf{P}$ in the virtual space. The set of orbitals with the largest eigenvalues will maximize the functional F'. Since this eigenproblem involves a projection operator, only K eigenvalues are nonzero (the size of the AO space). The largest possible eigenvalues involve virtual orbitals that are fully contained in the AO space, in that case $\mathcal{P}|\tilde{\psi}_a\rangle = |\tilde{\psi}_a\rangle$ and the corresponding eigenvalue is one. All the other eigenvalues $\in$ (0,1). It would appear that the fractional eigenvalues violate the idempodency of $\mathcal{P}$. Such eigenvalue spectrum results from incompleteness of the virtual space and fractional eigenvalues correspond to projections that partially lie *outside* of the virtual space. The eigenvectors that are completely orthogonal to the minimal AO space correspond to eigenvalues of zero.

Now, let us determine the virtual orbitals hidden in Knizia's projection operator $1 - \tilde{O}$. The following identity is true in the space of atomic orbitals:

$$\mathcal{P} = \sum_{i=1}^{n} |\tilde{\psi}_i\rangle\langle\tilde{\psi}_i| + \sum_{a=n+1}^{K} |\tilde{\psi}_a\rangle\langle\tilde{\psi}_a| \tag{20}$$

where $\tilde{\psi}_a$ is taken as orthogonalized complement of the space spanned by $\tilde{\psi}_i$ in the space of AOs. Therefore,

$$1 - \tilde{O} = \sum_{a=n+1}^{K} |\tilde{\psi}_a\rangle\langle\tilde{\psi}_a| \tag{21}$$

The above equation is true in our case since the only functions that we act upon with the operator $1 - \tilde{O}$ are AOs. The $|\tilde{\psi}_a\rangle$ orbitals have two properties: they completely lie in virtual space due to the identity $(1 - O)(1 - \tilde{O}) = 1 - \tilde{O}$, and they completely lie in AO space (since $|\tilde{\psi}_a\rangle \in \mathcal{P}|\psi_a\rangle$). Therefore, the orbitals that define $1 - \tilde{O}$ must correspond to eigenvalue 1 of the $\mathcal{P}$ operator and they are the same orbitals that maximize the functional 18, which completes the proof.

The recent modification of the QUAMBO method renders it practically identical to the method proposed by Knizia. Both methods attempt to build maximum correspondence between the space of virtual orbitals. The authors in ref 10 suggest performing singular value decomposition of the overlap matrix between virtual orbitals and orthogonalized atomic orbitals. This is equivalent to the diagonalization of the operator $\mathcal{P}$ in the full virtual space. Although such procedure results in a maximum correspondence between two spaces, in the author's opinion, it is more reasonable to use nonorthogonal orbitals (where only the overlap between AOs from different atoms is nonzero, but AOs within atom are orthogonalized). Atomic orbitals of free atoms provide a natural definition of the atomic boundary and they appear to be more appropriate for the

calculations involving basis set dependent properties that ascribe given basis function to a certain atom. Interatomic orthogonalization introduces artificial deformation of the original AOs, which depends on the molecular geometry. Therefore, the orbitals obtained by diagonalization of the $\mathcal{P}$ operator in place of $\tilde{\mathcal{P}}$ are less similar to the original atomic orbitals, as they are deformed twice, although this difference is subtle as shown by numerical results presented in the following section.

QUAMBOs or QUAOs span the full space of occupied molecular orbitals if they are linearly independent. This can be demonstrated as follows. The set of orbitals produced by projection $O\phi_k$ spans the space of MOs if the dimension of $\{O\phi_k\}$ is the same as the dimension of MOs. This was already discussed by Knizia in ref 2. The dimension of such a projection would be smaller *only if at least* one molecular orbital were orthogonal to all atomic orbitals, which is impossible physically (it is not just unlikely, it simply cannot happen for any occupied MO of nonexcited electronic state of a molecule). Adding vectors from orthogonal subspace can in principle destroy this spanning property, as it happens in the following projection:

$$O\phi_k + (1 - O)\phi_k \tag{22}$$

The above sum of two projections recovers the original space of atomic orbitals, which do not span the MO space. An explicit construction given by eq 1 ensures that we still recover the MO space if $|\tilde{\phi}_k\rangle$ are linearly independent. The final condition that must be fulfilled is that the sum of ranks of the operator O and the operator $1 - \tilde{O}$ must be equal to the dimension of the AO space, a condition that is not fulfilled by the sum of projections given by eq 22. The fact that orbitals defining the projection operator $1 - \tilde{O}$ are contained in the space of AOs, a property demonstrated earlier, ensures that the dimensionality of the set $\{(1 - \tilde{O})|\phi_k\rangle\}$ is exactly $K - n$, as required.

It is possible to demonstrate that if the resulting QUAO/IAO or QUAMBO orbitals are linearly dependent then they might not span the full MO space. Since numerically they are always very similar to minimum basis AOs, it is very unlikely to happen. Nonetheless, the exact spanning property can be assured only if the QUAO/IAO are linearly independent, a test that is easy to perform in all routine calculations.

3. NUMERICAL RESULTS

The aim of this section, in which various aspects of the applications of QUAMBO/QUAO are discussed, is 3-fold:

1. To compare QUAMBOs build around $\tilde{\mathcal{P}}$ vs $\mathcal{P}$ operators, which involve nonorthogonal (QUAMBO) and orthogonal (QUAO) underlying AOs, respectively.
2. To assess the performance of QUAO-based population analysis.
3. To test the convergence of the Pipek−Mezey localization in difficult cases, where basis set is linearly dependent.

Some of the points are extension of the analysis performed in ref 2. The current work focuses on large systems and linearly dependent basis sets.

All the calculations presented in this paper were performed using the PQS program package[11] and a newly implemented QUAO/QUAMBO module. Table 1 contains the values for root-mean-square deviations (RMSD) of the atomic polarized orbitals from the original atomic orbitals, as defined by eq 2. This involves orbitals obtained by diagonalization of $\tilde{\mathcal{P}}$

Table 1. HF/6-311G** Root Mean Square Deviations of QUAOs (Quasiatomic Orbitals) from Free Atom Orbitals[a]

molecule	(QUAO)-AO	(QUAMBO)-AO	(QUAO)-(QUAMBO)
NaCl	0.04695	0.04665	0.00534
Na_4Cl_4	0.03949	0.03890	0.00682
$Na_{32}Cl_{32}$	0.03417	0.03383	0.00478
$Na_{108}Cl_{108}$	0.03265	0.03239	0.00420
SF_6	0.07854	0.07830	0.00658
$NaClO_4$	0.09578	0.09484	0.01396

[a]Two versions of polarized orbitals were tested based on (a) nonorthogonal AO orbitals (QUAMBO)[5] and (b) orthogonalized AO orbitals (QUAO).[2,10] All the geometries are optimized at 6-311G**/B3LYP level. $Na_{108}Cl_{108}$ is not reoptimized and a lattice constant a = 5.64 Å has been used.

(original QUAMBO formulation[5]) and orbitals obtained by diagonalization of the $\mathcal{P}$ operator[10] (orthogonal projection, QUAO). The numbers demonstrate that in all the cases the difference is small and both methods produce similar polarized orbitals. The case of sodium perchlorate is somewhat special. In this molecule QUAOs significantly differ from the AOs, compared to sodium chloride nanocrystals. However, the absolute difference between two versions of the QUAO orbitals is still small. Sodium perchlorate probably involves strong overlap between orbitals of oxygen and chlorine atoms resulting in greater nonorthogonality. In this case the difference between $\tilde{\mathcal{P}}$ and $\mathcal{P}$ is more significant.

Table 2 shows population analysis results for sodium perchlorate based on two sets of orbitals. The numbers slightly

Table 2. Mulliken Population Analysis in Sodium Perchlorate at HF/6-311G** Level Based on Two Sets of Quasiatomic Orbitals, As Described in Table 1[a]

molecule	QUAO	QUAMBO
Na	0.967	0.968
Cl	2.498	2.454
O1	−0.819	−0.807
O2	−0.819	−0.807
O3	−0.914	−0.904
O4	−0.914	−0.904

[a]The geometry was optimized at the 6-311G**/B3LYP level.

differ and QUAMBOs produce slightly smaller partial charges. Pipek−Mezey localization based on QUAMBOs converges faster by about 10%. It might be argued that the projection onto nonorthogonalized atomic orbitals should produce physically more natural framework for the definition of atomic boundaries, since the molecular orthogonalization is a physically artificial constraint. In practice the difference is rather small.

Population analysis and closely related Pipek−Mezey localization scheme are important tools in qualitative interpretation of the wave function or local treatment of the electron correlation. However, they suffer from the basis set dependence, often providing different results in the same molecule if a different basis set is used. Even if the resulting localized orbitals are stable with respect to the basis set, the convergence might be affected as shown in refs 12 and 13 (the authors also suggested improved algorithms and tested applications of IAOs/QUAOs). The performance of population analysis deteriorates significantly if basis set is linearly

3088

dx.doi.org/10.1021/ct500245f | J. Chem. Theory Comput. 2014, 10, 3085−3091

dependent, due to either the presence of diffuse basis functions or periodic character of a system under investigation. Löwdin populations are also sensitive to spatial rotations of a molecule or nonorthogonal mixing of individual basis functions within an atom, even if the space they define is the same. Therefore, one can expect that results can change significantly even in small molecules in extended basis sets if diffuse basis functions are employed, or in crystalline models consisting of hundreds atoms or more, despite using modest-size basis sets like 6-311G**. This deterioration is the consequence of the fact that the definition of the atomic boundary based on basis functions, as done in regular Löwdin or Mulliken analysis, becomes ill-defined due to the presence of basis functions that extend far in space. In case of extended systems, multitude of small contributions from off-diagonal density matrix elements involving distant atoms combine to a large sum that can override the physically relevant picture provided by basis functions located on atoms of interest. As the first example let us demonstrate the performance of the population analysis in models of sodium chloride crystals of increased size, presented in Table 3 where two basic schemes, Mulliken and Löwdin population analysis are presented.

Table 3. Traditional Mulliken and Löwdin Population Analysis for NaCl Clusters of Various Sizes at Hartree−Fock/6-311G** Level[a]

system	Mulliken		Löwdin	
	Na	Cl	Na	Cl
NaCl	0.75	−0.75	0.42	−0.42
Na$_4$Cl$_4$	0.79	−0.79	0.16	−0.16
Na$_{32}$Cl$_{32}$	(0.58, 3.21)	(−1.23, −0.85)	(−0.10, 0.17)	(−0.17, 0.10)
Na$_{108}$Cl$_{108}$	(−0.33, 3.85)	(−1.38, −0.87)	(−0.17, 0.33)	(−0.26, 0.14)
Na$_{256}$Cl$_{256}$[b]	(0.57, 1.40)	(−0.84, −0.70)	(−0.36, 0.18)	(−0.11, 0.34)

[a]Each entry represents either a single number or a range of charges for a given *type* of atomic center. The geometries of the first three systems are optimized at 6-311G**/B3LYP level, while Na$_{108}$Cl$_{108}$ and Na$_{256}$Cl$_{256}$ are not reoptimized and a lattice constant a = 5.64 Å has been used. [b]Due to the size of the system and numerical stability issues, the 6-31G** basis set was used.

The results exhibit the well-known fact that usefulness of the traditional population analysis in extended systems is close to none. Charges on positive ionic centers range from −0.3 to +3.8 in larger crystal models. Decreasing the size of the valence shell in the basis set of the largest system (from triple-ζ to double-ζ) results in less linear dependency and the numbers are more reasonable, for sodium atoms the range is (0.57,1.40). Still, the biggest charges and the scatter are unphysical.

The coronene molecule in aug-cc-pVDZ basis set is a prototypical example of a smaller system in which difficulties arise from the presence of diffuse basis functions. While such functions are usually not needed in calculations involving intramolecular properties, their presence significantly accelerates the convergence toward basis set limit for interaction energies.[14,15] The populations of symmetry unique atoms are presented in Table 5, while the atom numbering scheme is illustrated in Figure 1.

Both cases are easily fixed by the application of the QUAMBO basis. Table 4, Figure 1 and Table 5 contain results for NaCl nanocrystals and the coronene molecule. The partial
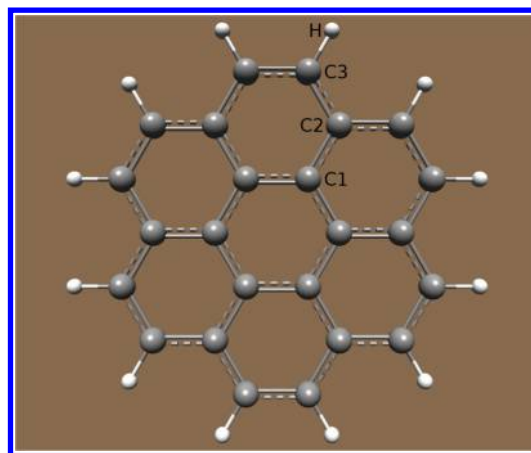

Figure 1. Atom labeling in coronene molecule, as used in Table 5

Table 4. Mulliken and Löwdin Population Analysis in Quasiatomic (QUAMBO) Basis for NaCl Clusters of Various Sizes at Hartree−Fock/6-311G** Level[a]

system	Mulliken		Löwdin	
	Na	Cl	Na	Cl
NaCl	0.93	−0.93	0.88	−0.88
Na$_4$Cl$_4$	0.90	−0.90	0.81	−0.81
Na$_{32}$Cl$_{32}$	(0.84, 0.92)	(−0.89, −0.88)	(0.72, 0.82)	(−0.80, −0.74)
Na$_{108}$Cl$_{108}$	(0.84, 0.92)	(−0.89, −0.86)	(0.71, 0.83)	(−0.81, −0.72)
Na$_{256}$Cl$_{256}$	(0.84, 0.92)	(−0.89, −0.86)	(0.71, 0.83)	(−0.81, −0.72)

[a]Each entry represents either a single number or a range of charges for a given *type* of atomic center. Geometry optimization, see Table 3.

Table 5. Traditional Mulliken and Löwdin Population Analysis and the Quasiatomic (QUAMBO) Based Analysis for Coronene Molecule at Hartree−Fock/aug-cc-pVDZ and B3LYP/aug-cc-pVDZ Levels[a]

atom	traditional (SCF)		QUAMBO-based (SCF)		QUAMBO-based (DFT)	
	Mulliken	Löwdin	Mulliken	Löwdin	Mulliken	Löwdin
C1	0.835	−0.026	0.000	0.002	0.002	0.003
C2	0.113	−0.145	−0.016	−0.003	−0.013	0.000
C3	0.248	0.082	−0.259	−0.158	−0.270	−0.165
H	−0.722	0.003	0.267	0.158	0.276	0.164

[a]The coronene geometry was optimized at aug-cc-pVTZ/B3LYP level.

charges in NaCl are very stable and do not depend on the size of the crystal model. Even the application of the smaller basis set that is numerically more stable, a necessity in such a large system, did not change the populations that appear to be fully converged. This stability of results enables meaningful conclusions about net charges in NaCl crystals. Mulliken population analysis produces similar charges for bulk atoms equal to 0.85, with Cl$^-$ ions having slightly more charge than Na$^+$ ions, resulting in net negative charge in the bulk phase. The surface Na$^+$ ions have +0.88 charge, while Cl$^-$ has −0.87. This results in +0.01 net positive charge for each pair of Na and Cl ions. The corner atoms are even more charged with −0.88 for Cl$^-$ and +0.92 for Na$^+$. While the charges of bulk ions almost exactly cancel with slight negative net charge, the surface is positively charged while corners and edges are even more positively charged. The total charge density near the surface is greater than in the bulk, which means that the outer atoms
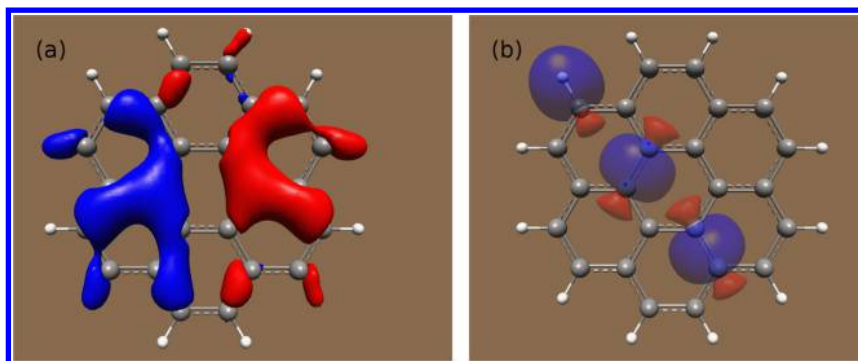
Figure 2. Pipek−Mezey localization in coronene molecule at SCF/aug-cc-pVDZ level. (a) A sample "localized" σ orbital obtained from a traditional scheme based on the density matrix expressed directly in aug-cc-pVDZ basis set; (b) a scheme based on quasiatomic orbitals—here three localized orbitals are superimposed.

produce strong electric fields near the crystal surface. Similar conclusions can be obtained by comparing charges obtained from Löwdin population analysis, although the absolute ionic charges are slightly different. Note that this kind of insight is impossible without QUAMBOs or QUAOs/IAOs.

The results for coronene are reasonable as well. The partial charges inside the molecule are close to zero, as expected, only the carbon atoms in the outer ring are negatively charged, while hydrogen atoms are positive. This is consistent with the chemical intuition based on atomic electronegativities. The central carbon atoms are bonded only to alike atoms and almost no net charge transfer takes place. The outer atoms "consume" some electronic charge from hydrogen, while C2 atoms share in this feast and are slightly negatively charged as well.

The last test involves Pipek−Mezey localization in the coronene molecule. Figure 2 compares two results: (a) traditional localization which utilizes all the basis functions with (b) QUAMBO-based scheme. The π orbitals localize well in both cases and they are not shown, but the σ framework turns out to be difficult. The left panel represents one of the σ orbitals localized in a traditional way. This is compared to a sample of three σ localized orbitals based on QUAMBO decomposition, superimposed on the left panel. The improvement is dramatic, as expected from earlier results, since the PM localization is build on top of the population analysis concept.

4. CONCLUSIONS

The near-equivalence between the most recent quasiatomic orbitals approach as reported in ref 10 and the Intrinsic Atomic Orbitals[2] method was established. The space spanned by orbitals obtained from both methods is the same. As an additional benefit, it was shown that the simpler projection formula in Appendix C of ref 2 is in fact equivalent to eq 2 in ref 2 under the condition that the minimal basis is exactly spanned by the computational basis. Although the Knizia's formula appears to be a shortcut that provides direct solution to the modified methods of West et al., there are some drawbacks as well. First, it is equivalent to the minimization of the differential overlap based on fully orthogonalized atomic orbitals, which in the author's opinion is physically less appealing. Interatomic orthogonalization usually results in more diffuse atomic orbitals, while nonorthogonalized (interatomically) AOs are more compact. Second, although no direct diagonalization is required in Knizia's method, building the projection operator $1 - \tilde{O}$ is not trivial and requires many

projections and inverting the overlap matrix. Therefore, the numerical complexity is not significantly different from ref 10, although if Cholesky decomposition is used then Knizia's scheme will produce the polarized orbitals significantly faster, especially in large systems with more than 10 000 basis functions.

In building the polarized atomic set via the QUAMBO/QUAO approach it is physically well-chosen to perform orthogonalization of the underlying minimal AO basis for each atom independently. This reflects the fact that the orbitals of free atoms are indeed orthogonal, and as an extra benefit, it ensures rotational invariance of the population analysis and the invariance of the valence virtual orbital space with respect to geometrical rotations of a molecule. The MINAO basis set utilized in the IAO scheme is a good choice, but since both the QUAO or QUAMBO methods are generic, that is, in principle an arbitrary minimum basis set can be used, it is worthwhile to ensure such intraatomic orthogonality before performing any projections/minimizations.

Numerical tests demonstrate the extraordinary stability of the standard population analysis schemes when used with QUAMBOs as the defining basis sets. It is important to stress that although QUAMBOs constitute a much smaller set, no loss of information about the molecular orbitals occurs. QUAMBOs provide a good platform to define atomic boundaries for interpretative tools that use basis sets in their definition of atomic contributions. QUAMBOs are very similar to atomic orbitals, and therefore, they are compact and do not produce overlap matrix elements between spatially distant atoms. Bulk of the contributions to the atomic charges originate from next-neighbor in the QUAMBO basis. Additional benefit from employing QUAMBOs (or QUAOs/IAOs) is flexibility in underlying atomic orbitals. In principle, they can be expressed as an arbitrary set of minimal basis functions. Such basis functions can be obtained, for example, from either Hartree−Fock (as done in ref 2) or high level calculations (as in ref 10) on the free atoms or atomic orbitals of free atoms taken directly from high level calculations. In any case, each shell forms an orthogonal irreducible representation with respect to rotations. This property ensures that, for example, the Löwdin population analysis is invariant to the rotations of the molecule.[16] To prevent any trace of such rotational dependence it is advisable to orthogonalize the AOs *within* each atom only (if they are taken as some arbitrary minimum basis set), but to avoid global molecular orthogonalization to prevent loss of locality and to prevent blurring of the atomic boundaries. Note that this

statement requires more extensive testing and is based on a few examples tested by the author so far.

No attempt was made in this paper to validate the physical correctness of obtained results, as they were obtained by means of the Hartree−Fock method. The major focus was on the performance, numerical stability and on establishing the mathematical equivalence between two recently published schemes involving polarized atomic orbitals. A follow-up paper will extensively explore chemical applications of the charge decomposition provided by QUAO, which will include, for example, piezoelectric effect by imposing external forces as in ref 17 or decomposition of multipole moments into contributions from individual atoms. This is especially useful in decomposing molecular contribution in bulk phases, see an example application in ref 18. Decomposition of the molecular electrostatic charge and multipole moments are also very useful in molecular simulations.[19] Such decomposition ensures fast convergence of the multipole expansion even at distances very close to the molecular frame. The QUAO approach will enable applications of all these methods to large molecules and diffuse basis sets.

■ AUTHOR INFORMATION

Corresponding Author
*Email: janowski@uark.edu.

Notes
The authors declare no competing financial interest.

■ REFERENCES

(1) Pipek, J.; Mezey, P. G. *J. Chem. Phys.* **1989**, *90*, 4916−4926.
(2) Knizia, G. *J. Chem. Theory Comput.* **2013**, *9*, 4834−4843.
(3) Lee, M. S.; Head-Gordon, M. *J. Chem. Phys.* **1997**, *107*, 9085−9095.
(4) Lee, M. S.; Head-Gordon, M. *Int. J. Quantum Chem.* **1999**, *76*, 169−184.
(5) Lu, W. C.; Wang, C. Z.; Schmidt, M. W.; Bytautas, L.; Ho, K. M.; Ruedenberg, K. *J. Chem. Phys.* **2004**, *120*, 2629−2637.
(6) Lu, W. C.; Wang, C. Z.; Schmidt, M. W.; Bytautas, L.; Ho, K. M.; Ruedenberg, K. *J. Chem. Phys.* **2004**, *120*, 2638−2651.
(7) Subotnik, J. E.; Dutoi, A. D.; Head-Gordon, M. *J. Chem. Phys.* **2005**, *123*, 114108.
(8) Auer, A. A.; Nooijen, M. *J. Chem. Phys.* **2006**, *125*, 024104.
(9) Laikov, D. N. *Int. J. Quantum Chem.* **2010**, *111*, 2851−2867.
(10) West, A. C.; Schmidt, M. W.; Gordon, M. S.; Ruedenberg, K. *J. Chem. Phys.* **2013**, *139*, 234107.
(11) *PQS version 4.1*; Parallel Quantum Solutions: Fayetteville, AR, see http://www.pqs-chem.com/.
(12) Lehtola, S.; Jonsson, H. *J. Chem. Theory Comput.* **2013**, *9*, 5365−5372.
(13) Lehtola, S.; Jonsson, H. *J. Chem. Theory Comput.* **2014**, *10*, 642−649.
(14) Sinnokrot, M. O.; Sherrill, C. D. *J. Chem. Phys.* **2006**, *110*, 10656−10668.
(15) Janowski, T.; Pulay, P. *Chem. Phys. Lett.* **2007**, *447*, 27−32.
(16) Mayer, I. *Chem. Phys. Lett.* **2004**, *393*, 209−212.
(17) Wolinski, K.; Baker, J. *Mol. Phys.* **2010**, *108*, 1845−1856.
(18) Rusnak, A. J.; Pinnick, E. R.; Calderon, C. E.; Wang, F. *J. Chem. Phys.* **2012**, *137*, 034510.
(19) Janowski, T.; Wolinski, K.; Pulay, P. *Chem. Phys. Lett.* **2012**, *530*, 1−9.