Theor Chim Acta (1997) 95: 81–95

Theoretica
Chimica Acta

A method for population and bonding analyses in calculations with extended basis sets

Wenjian Liu, Lemin Li
State Key Laboratory of Rare Earth Materials Chemistry and Applications, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China

Received July 2, 1996 / Accepted October 2, 1996

Summary. A method for population and bonding analyses in the calculations with extended basis sets is proposed. The definition and evaluation method of the atomic orbitals in molecular environments (AOIMs) are described. It is shown that the AOIMs can be divided into two subsets, the strongly occupied minimal compact subset $\{\mathrm{AOIM}\}_B$ and the very weakly occupied "Rydberg" subset $\{\mathrm{AOIM}\}_R$, according to the orbital population obtained from Mulliken analysis with AOIMs as basis sets. The viewpoint of "molecular orbitals consisting of minimal atomic orbital sets" can be optimally realized in terms of $\{\mathrm{AOIM}\}_B$. The Mulliken population based on AOIMs is reasonable and fairly stable to changes of basis sets. The Mayer bond orders calculated based on $\{\mathrm{AOIM}\}_B$ are quite stable to the changes of basis sets; therefore they can be used to measure objectively the contribution of individual atomic orbitals to bonding.

Key words: Population analysis – Bond order – AOIM – Basis set

1 Introduction

It is well known that the population and bonding analysis method proposed by Mulliken [1] can transform the complicated quantum chemical calculation results into a simplified physical picture familiar to chemists. Thus it is very popular among chemists. It is based on the "LCAO-MO" formulation, that is, the molecular orbitals are expressed as the linear combination of the minimal orbital set of the constituent atoms. Thus Mulliken analysis method can give clear and reasonable pictures only under the "balanced" minimal basis sets [1]. However, in order to obtain accurate results extended basis sets have to be used in quantum chemical calculations, which results in some difficulties in the population and bonding analyses. Using extended basis sets the Mulliken analysis cannot give results with unequivocal physical meaning, because in that case a basis function does not correspond to an atomic orbital. Furthermore, the results are rather sensitive to

---

Correspondence to: L. Li; Fax: 86-010-62751708; e-mail: lilm@ibmstone.pku.edu.cn

changes in the basis sets. It may give in some cases even unreasonable results. Several approaches have been proposed to remove the defects of Mulliken analysis method [2–9]. However, none of them have completely succeeded due to their own weaknesses.

In this paper, it will be shown that in the calculations with extended basis sets, if the atomic orbitals in molecular environments (AOIMs) are taken as the basis sets of population and bonding analyses the defects of Mulliken analysis method can be overcome to a certain extent, while keeping the advantage of simplicity and being easy to use. Therefore an approach is provided to obtain a clear and simplified physical picture from the results calculated with extended basis sets.

2 Method and principle

It has been recognized for a long time that the orbitals of an atom in free states and in molecular environments may be different [10–14]. The changes of atomic orbitals in molecular environments have to be taken into account for explanation of many chemical phenomena [15–17]. Therefore it seems more reasonable to adopt the atomic orbitals in molecular environments as basis sets in the LCAO-MO formulation. In a free atom the electrons move in a potential field with spherical symmetry. In molecular environments the potential which the electrons experience is non-spherical. Its spherical component with respect to an atom is also different from the potential of that atom in free states. Both the atomic feature and the effect of molecular environments should be taken into account. It has been proposed [18] to define the atomic orbitals in molecular environments as the wave functions of an electron moving in an effective potential which is the spherical average of the molecular potential with the center at the related atomic nucleus. In other words, the AOIMs, $\{\phi_i^{\mathrm{A}}(\boldsymbol{r})\}$, are the eigenfunctions of the following equation:

$$[-\tfrac{1}{2}\nabla^2 + V_{00}^{\mathrm{A}}(r)]\phi_i^{\mathrm{A}}(\boldsymbol{r}) = E_i^{\mathrm{A}}\phi_i^{\mathrm{A}}(\boldsymbol{r}), \tag{1}$$

where the effective potential $V_{00}^{\mathrm{A}}(r)$ is the spherical average of the molecular potential field with the center at the nucleus of atom A. When the spherical coordinate (r, θ, φ) with origin at nucleus A is adopted, then

$$V_{00}^{\mathrm{A}}(r) = (1/4\pi)\int V(\boldsymbol{r})\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\varphi, \tag{2}$$

where $V(\boldsymbol{r})$ is the molecular potential field. E_i^{A} is the corresponding eigenvalue. Let

$$\phi_i^{\mathrm{A}}(\boldsymbol{r}) = (1/r)P_{nl}^{\mathrm{A}}(r)Y_{lm}(\theta, \varphi), \tag{3}$$

where n, l, m are the principal, angular and magnetic quantum numbers. Substituting Eq. (3) into Eq. (1), the radial Schroedinger equation for atom A can be obtained:

$$\left[-\tfrac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}r^2} + l(l+1)/2r^2 + V_{00}^{\mathrm{A}}(r)\right]P_{nl}^{\mathrm{A}}(r) = E_i^{\mathrm{A}}P_{nl}^{\mathrm{A}}(r). \tag{4}$$

Expanding $P_{nl}(r)$ with the basis set $\{\chi_k^{\mathrm{A}}(r)\}$ centered on nucleus A

$$P_{nl}^{\mathrm{A}}(r) = \sum_k C_{ki}\chi_k^{\mathrm{A}}(r) \tag{5}$$

and substituting it into Eq. (4), the matrix equation (6) can be obtained:

$$\boldsymbol{F}_l^{\mathrm{A}} \boldsymbol{C}_l^{\mathrm{A}} = \boldsymbol{S}_l^{\mathrm{A}} \boldsymbol{C}_l^{\mathrm{A}} \boldsymbol{E}_l^{\mathrm{A}}, \tag{6}$$

where the matrix elements of $\boldsymbol{F}_l^{\mathrm{A}}$ are

$$(\boldsymbol{F}_l^{\mathrm{A}})_{kj} = \left\langle \chi_k^{\mathrm{A}}(r) \middle| \left[-\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}r^2} + l(l+1)/2r^2 + V_{00}^{\mathrm{A}}(r) \right] \middle| \chi_j^{\mathrm{A}}(r) \right\rangle \tag{7}$$

and the overlap matrix elements $(\boldsymbol{S}_l^{\mathrm{A}})_{kj}$ are

$$(\boldsymbol{S}_l^{\mathrm{A}})_{kj} = \langle \chi_k^{\mathrm{A}}(r) \,|\, \chi_j^{\mathrm{A}}(r) \rangle. \tag{8}$$

C_l^{A} is the combination coefficient matrix and E_l^{A} is the eigenvalue matrix. Solving Eq. (6) and substituting the obtained $P_{nl}^{\mathrm{A}}(r)$ into Eq. (3), the orbitals of atom A in the molecular environment can be obtained.

For computational convenience, $\{\chi_k^{\mathrm{A}}(r)\}$ can be taken as the same as the radial functions of the basis sets used in the related molecular calculation, though this is by no means necessary. Denote the basis set used in the molecular calculation as $\{\chi_{nlm}^X, X = \mathrm{A}, \mathrm{B}, \mathrm{C}, \ldots\}$ and the subset of the functions centered on atom A as $\{\chi_{nlm}^{\mathrm{A}}\}$. Suppose that $\boldsymbol{F}_{lm}^{\mathrm{A}}$ and $\boldsymbol{S}_{lm}^{\mathrm{A}}$ are the Fock and overlap matrix blocks, respectively, corresponding to the basis functions with quantum numbers l, m and belonging to atom A, it is easy to prove that [18]

$$\sum_m (\boldsymbol{F}_{lm}^{\mathrm{A}})/(2l+1) = \boldsymbol{F}_l^{\mathrm{A}}, \tag{9}$$

$$\sum_m (\boldsymbol{S}_{lm}^{\mathrm{A}})/(2l+1) = \boldsymbol{S}_l^{\mathrm{A}}. \tag{10}$$

So after the molecular calculation finishes, the matrices $\boldsymbol{F}_l^{\mathrm{A}}$ and $\boldsymbol{S}_l^{\mathrm{A}}$ can be evaluated by use of Eqs. (9) and (10), and then Eq. (6) is solved to obtain C_l^{A} and the AOIMs. When the basis set is complete for the molecular calculation and the subset is complete too for the atomic calculations, the obtained AOIMs will be exact and unique. The result of Mulliken analysis using the exact AOIMs as basis set should be unique, because the first-order density matrix with AOIMs as basis is uniquely determined. Even though $\{\chi_{nlm}^X(\mathrm{r})\}$ is not a complete basis set, so long as it is a balanced one, the obtained AOIMs would be the optimal approximate solution for the given basis set, because they are obtained from variational calculations. Then the arbitrariness of the Mulliken analysis results would be decreased to a certain extent. It has been found [19] that the molecular calculations using the extended basis sets of 3ζ plus polarization functions can give sufficiently accurate results. Therefore the approximate AOIMs should be good enough when the used basis sets are extended to 3ξ or higher quality.

By use of the matrix of transformation from basis functions to AOIMs, the overlap and density matrices can be transformed into those using the AOIMs $\{\phi_{nlm}^X(r)\}$ as basis set. Then the population and bonding analyses can be carried out in the ordinary way.

3 Results and discussion

The aforementioned method can be applied to any calculations based on orbital approximation, e.g. ab initio Hartree–Fock method or density functional theory (DFT) method. In recent years, the calculation practice has shown [20, 21] that the

DFT method based on Kohn–Sham equation usually gives more satisfactory results than the Hartree–Fock method does. In this work, a series of small molecules were calculated with different quality basis sets by the approximate density functional theory (LDA) method. The Kohn–Sham equation was solved, then the population and bonding analyses were carried out. The program for DFT calculations and that for population and bonding analyses with AOIMs as basis sets were developed in author's laboratory [22]. The reliability of the used DFT program has been confirmed by checking its outputs with those from the famous ADF program [23]. The minimal basis sets are numerical ones obtained by the approximate density functional theory calculations for free atoms. The double zeta basis sets were taken from Ref. [24]. The three and multiple zeta basis sets were taken from Ref. [25].

As an example, some calculation results for hydrogen fluoride are presented. The three ζ plus polarization function basis set was used. After the convergent result of the molecular calculation for hydrogen fluoride was obtained, the angular averaged Fock matrix $\boldsymbol{F}_l^{\mathrm{A}}$ and overlap matrix $\boldsymbol{S}_l^{\mathrm{A}}$ for the subshell l of atom A were calculated by means of Eqs. (9) and (10). Then Eq. (6) was solved to get the eigenvector matrix $\boldsymbol{C}_l^{\mathrm{A}}$ of AOIMs for the subshell l of atom A. In Table 1a presented are the Fock matrix $\boldsymbol{F}_0^{\mathrm{F}}$ and the overlap matrix $\boldsymbol{S}_0^{\mathrm{F}}$ for the s type orbitals of fluorine atoms in HF molecules. The AOIM vector matrix $\boldsymbol{C}_0^{\mathrm{F}}$ for s type orbitals of fluorine atoms in HF molecules is presented in Table 2a. Similarly, the matrices $\boldsymbol{F}_1^{\mathrm{F}}$ and $\boldsymbol{S}_1^{\mathrm{F}}$ as well as $\boldsymbol{C}_1^{\mathrm{F}}$ for p type orbitals of fluorine atoms in HF molecules are presented in Tables 1b and 2b, respectively. In Tables 1c and 2c, the corresponding data of $\boldsymbol{F}_0^{\mathrm{H}}$, $\boldsymbol{S}_0^{\mathrm{H}}$ and $\boldsymbol{C}_0^{\mathrm{H}}$ for hydrogen atoms in HF molecules are presented, respectively. Because single ζ bases were used for the polarization functions, the corresponding AOIMs are just the same as the bases. Therefore the matrix of transformation from bases to AOIMs could be constructed, with which the molecular density matrix and overlap matrix were transformed into those with AOIMs as the basis set. The transformed matrices were used to carry out the population and bonding analyses based on AOIMs. The results are presented in the following sections.

3.1. Population analysis

According to the definition [1], the electron population in the nl subshell of atom A is

$$q_{nl}^{\mathrm{A}} = \sum_{i \in (\mathrm{A}nl)} \sum_{j}^{\mathrm{all}} P_{ij} S_{ij}, \tag{11}$$

where P_{ij} and S_{ij} are the density and overlap matrix elements, respectively. For the minimal basis sets the basis functions are just the approximate AOIMs. For the extended basis sets, the population on the basis functions cannot be correlated to the electronic occupancy on subshells, because there is no simple correspondence between the basis functions and the atomic orbitals. When the population analysis is based on AOIMs, the results can be identified with the electronic configurations of atoms in molecular environments.

In Table 3 listed are the results of Mulliken population analysis based on AOIMs for some molecules calculated using different quality of basis sets. It can be seen from Table 3 that no matter which type of basis sets was used in the molecular calculations, the AOIMs can be divided into two clear-cut subsets. One subset

Table 1a. The angular-averaged Fock and overlap matrix blocks for *s* orbitals of F atoms in HF molecules

Overlap matrix (upper triangular matrix)

	1.0000	0.9841	0.9964	0.0000	0.3357	0.1589
− 24.074		1.0000	0.9655	0.0763	0.2406	0.1053
− 23.692	− 22.452		1.0000	− 0.0352	0.3762	0.1818
− 23.987	− 24.053	− 23.668		1.0000	− 0.9191	− 0.9479
0.0018	− 0.0714	0.0356	− 1.0434		1.0000	0.8682
− 8.0846	− 8.1416	− 7.9588	0.9434	− 3.4278		1.0000
− 3.8257	− 3.6379	− 3.8687	1.0114	− 2.3175	− 1.4228	

Fock matrix (lower triangular matrix)

Table 1b. The angular-averaged Fock and overlap matrix blocks for *p* orbitals of F atoms in HF molecules

Overlap matrix (upper triangular matrix)

	1.00000	0.83734	0.96422
− 0.35592		1.00000	0.66709
− 0.23488	0.76120		1.00000
− 0.37198	− 0.71016	− 0.14501	

Fock matrix (lower triangular matrix)

Table 1c. The angular-averaged Fock and overlap matrix blocks for *s* orbitals of F atoms in HF molecules

Overlap matrix (upper triangular matrix)

	1.00000	0.91016	0.98483
− 0.08417		1.00000	0.96471
− 0.15624	0.01610		1.00000
− 0.12870	− 0.12033	− 0.14217	

Fock matrix (lower triangular matrix)

is strongly occupied, denoted as $\{\mathrm{AOIM}\}_B$. It corresponds to the sum set of the occupied subshells of free atoms in ground states, in general, being a minimal compact set. The occupancy on $\{\mathrm{AOIM}\}_B$ describes the basic electronic configurations of an atom. Thus it is fundamental and indispensable for essentially correct description of the electron distribution on individual atomic orbitals. The population on the orbitals of the other subset is very small in general, called "Rydberg set" according to Ref. [7], and denoted by $\{\mathrm{AOIM}\}_R$. The electron population on "Rydberg" set is unessential for describing the charge distribution of a molecule. It reflects the fact that the LCAO-MO formulation based on minimal basis sets

Table 2a. The *s* type AOIMs of F atoms in HF molecules

1*s*	2*s*	3*s*	4*s*	5*s*	6*s*
1.0027	0.3448	− 22.211	45.014	− 128.93	487.22
− 0.0009	− 0.1283	8.2373	− 16.576	46.949	− 159.70
− 0.0008	− 0.0941	1.5196	4.8850	107.97	− 345.55
− 0.0039	− 1.3933	42.317	− 117.05	− 88.253	50.814
− 0.0024	− 0.3182	29.704	− 75.465	− 55.192	31.548
− 0.0018	− 0.1136	16.477	− 51.819	− 39.824	23.002

Table 2b. The *p* type AOIMs of F atoms in HF molecules

2*p*	3*p*	4*p*
1.4803	18.052	9.6451
− 0.2183	− 6.9496	− 2.2298
− 0.3096	− 12.727	− 8.0430

Table 2c. The *s* type AOIMs of H atoms in HF molecules

1*s*	2*s*	3*s*
− 1.4777	7.5082	− 9.0325
− 0.7720	1.4083	− 7.6339
3.1920	− 8.6572	16.177

cannot accurately describe the charge distribution in a molecule. The fact that two subsets, $\{\mathrm{AOIM}\}_B$ and $\{\mathrm{AOIM}\}_R$, can be clearly distinguished and the population on the subset $\{\mathrm{AOIM}\}_R$ is very small, shows that the LCAO-MO picture can be optimally realized by taking AOIMs as basis sets. In other words, by use of $\{\mathrm{AOIM}\}_B$, the charge distribution of a molecule can be described as accurately as possible within the limitation of minimal basis sets being used. For obtaining the simplified physical pictures of the charge distribution as accurately as possible, one can consider the population on subset $\{\mathrm{AOIM}\}_B$ only, while neglecting the subset $\{\mathrm{AOIM}\}_R$ completely. The above consideration may be helpful to explain the fact that in the DV-Xα-SCC method, although the molecular charge is approximately decomposed into the superposition of the spherical atomic charges by Mulliken population analysis method, which seems a very rough approximation, the obtained results approach to those from much more accurate treatment.

Two points should be noted. The first is that in some cases, some orbitals, e.g. the Li 2*p* orbitals in LiF and the S 3*d* orbitals in SF, possess significant population, although these orbitals are unoccupied in the related free atoms in ground states. These phenomena are usually ascribed to the change of atomic orbitals when the atoms transform to molecular environments from free states. It is well known that the chemical properties of some elements, such as Li, Be, P, S, some transition metals, lanthanides, etc., cannot be completely elucidated if only their occupied valent orbitals of free atoms in ground states are considered [15–17, 26]. This type of orbitals should be put in the subset $\{\mathrm{AOIM}\}_B$ instead of $\{\mathrm{AOIM}\}_R$. The other point is that small negative population appears on some orbitals in the subset $\{\mathrm{AOIM}\}_{\mathrm{R}}$, which is obviously unreasonable. This is correlated with the inherent flaw of the Mulliken analysis method [2]. Fortunately, the drawback is not so serious here. First, the negative population on the orbitals in $\{\mathrm{AOIM}\}_{\mathrm{R}}$ is very small, less than 0.01 electrons as usual, thus it can be neglected for obtaining a simplified physical picture of the electron distribution on individual atomic orbitals through population analysis. Second, as has been mentioned above, the population on $\{\mathrm{AOIM}\}_{\mathrm{R}}$ represents the deviation of the LCAO-MO formulation based on minimal basis sets under the optimal situation, its values may be positive or negative. If one wants to remove this unreasonable point, there are two approaches. The first approach is to perform a symmetrical orthogonalization of AOIMs before the population analysis, which has been used in the natural population analysis by Reed et al. [7]. This approach has the defect that the population in fact is not localized on individual atoms in view of its space distribution. Furthermore, the orthogonalized atomic orbitals cannot be conveniently used to the bonding analysis in terms of overlap charges. The other approach is to assign the small population on an orbital in subset $\{\mathrm{AOIM}\}_{\mathrm{R}}$ to the highest valence orbital with the same symmetry. This approach corresponds to construction of a valent orbital consisting of the AOIM in the highest valence shell and the orbitals in $\{\mathrm{AOIM}\}_{\mathrm{R}}$. But it seems usually unnecessary.

As an intermediate step of constructing the natural atomic orbitals (NAO), Reed et al. [7] obtained the pre-NAO through diagonalization of one center angular symmetry blocks of the density matrix and indicated that the PNAOs could be divided into two sets on the basis of occupancy, the "minimal" set corresponding to all atomic subshells of non-zero occupancy in the ground state of free atoms, and the "Rydberg" set consisting of the remainder orbitals. For comparison, the Mulliken population analysis based on PNAOs for N_2 molecules has been carried out and the result is listed in the row of N_2 in Table 3 (the data in the parentheses). It is obvious that, according to the orbital population, the PNAOs cannot be clearly divided into the two subsets, the strongly occupied minimal subset and the very weakly occupied "Rydberg" subset. Through the "occupancy-weighted symmetric orthogonalization" procedure, Reed et al. [7] got the natural atomic orbitals, then reached the aim to clearly divide the basis set into two subsets. Their procedure not only made the calculation more complicated, but also introduced additional arbitrariness. Thus it is not so simple and objective as the AOIM method proposed in this paper. It can be seen from Table 3 that the population based on AOIMs is fairly stable to changes of the basis sets used in the molecular calculations. Of course, there must be some variation of the population, because when the basis sets used in the molecular calculations are not large enough, not only the AOIMs are approximate, but also the result of the molecular calculation is approximate. Maybe because both the MOs and AOIMs are obtained through the variational procedure, the population based on AOIMs can be comparatively more stable.

3.2. Bonding potential

Wiberg defined the orbital bonding potential as [27]

$$\begin{aligned} b_i &= 2(PS)_{ii} - (PS)_{ii}^2 \\ &= 2q_i - q_i^2, \end{aligned} \tag{12}$$

Table 3. Population analysis based on AOIMs: The distribution of electrons on individual atomic subshells

Molecules	Basis sets		1*s*	2*s*	3*s*	4*s*	2*p*	3*p*	4*p*	5*p*	3*d*
N_2	MIN		2.001	1.756			3.243				
			(2.001)	(1.756)			(3.243)				
	DZ		1.999	1.872	0.027	0.000	3.099	0.004			
			(1.817)	(1.711)	(0.251)	(0.154)	(2.645)	(0.457)			
	DZP		2.000	1.772	0.030	0.000	3.123	0.011			0.065
			(1.701)	(1.621)	(0.328)	(0.152)	(2.651)	(0.481)			(0.065)
	TZ		1.999	1.860	0.047	−0.003	3.070	0.043	−0.015		
			(1.674)	(1.000)	(0.985)	(0.240)	(1.999)	(1.074)	(0.024)		
	TZP		2.000	1.749	0.044	−0.003	3.112	0.040	−0.008		0.066
			(1.700)	(1.173)	(0.720)	(0.195)	(2.459)	(0.667)	(0.020)		(0.066)
	MZ		1.999	1.836	0.057	−0.003	3.041	0.079	−0.009	0.000	
			(1.380)	(1.318)	(0.703)	(0.449)	(2.026)	(0.584)	(0.275)	(0.225)	
CO	MIN	O	2.000	1.865			4.461				
		C	2.001	1.676			1.994				
	DZ	O	1.999	1.920	−0.016	0.000	4.340	−0.009			
		C	2.001	1.802	0.080	0.001	1.824	0.058			
	TZ	O	2.000	1.894	−0.019	0.000	4.361	−0.002	−0.003		
		C	2.000	1.829	0.088	0.000	1.821	0.025	0.005		
	MZ	O	2.000	1.896	−0.028	0.000	4.202	0.031	−0.004	0.000	
		C	2.000	1.941	0.074	0.001	1.811	0.057	0.021	−0.012	
HF	MIN	F	2.000	1.971			5.554				
		H	0.476								
	DZP	F	2.000	1.996	−0.006		5.501	−0.005			0.022
		H	0.440	0.014			0.039				
	TZP	F	2.000	1.970	−0.005	0.000	5.553	−0.003	−0.003		0.022
		H	0.404	0.015	0.003		0.042				

LiF	MIN	F	2.000	1.994			5.730				
		Li	2.000	0.227							
	DZ	F	2.000	2.003	− 0.002	0.000	5.808	− 0.001			
		Li	2.001	0.192	− 0.002	0.000					
	DZP	F	2.000	1.948	0.000		5.774	0.008			0.006
		Li	1.996	0.024	− 0.002	0.001	0.228				0.019
	TZP	F	2.000	1.958	− 0.001	0.000	5.718	0.006	− 0.001		0.005
		Li	1.997	0.035	0.006	− 0.001	0.242	0.011			0.025
	MZP	F	2.000	1.914	0.000	0.000	5.785	0.007	0.000	0.000	0.005
		Li	1.999	0.035	0.007	− 0.004	0.222	0.010			0.019
SF	MIN	F	2.000	1.979			5.249				
		S	2.000	2.000	1.958		6.000	3.814			
	DZ2P	F	2.000	1.949	− 0.028		5.440	− 0.020			0.038
		S	2.000	2.000	1.962		6.000	3.385	0.022		0.255
	TZ2P	F	2.000	1.971	− 0.014	0.000	5.331	0.000	− 0.002		0.041
		S	2.000	2.000	1.945	0.002	6.000	3.398	0.056	− 0.010	0.290
CCl_4	MIN	C	2.000	1.332			3.003				
		Cl	2.000	2.000	1.950		6.000	4.967			
	DZP	C	2.000	1.181	0.068		2.793	0.045			0.070
		Cl	2.000	2.000	1.946	− 0.009	6.000	4.977	− 0.006		0.051
	TZP	C	2.000	1.094	0.097	0.000	2.883	0.051	− 0.060		0.070
		Cl	2.000	2.000	1.941	− 0.006	6.000	4.938	0.054	− 0.012	0.052

Note: MIN, DZ, TZ, MZ, denote minimal, double zeta, triple zeta and multiple zeta basis sets, respectively; **P or **2P is used to indicate that one or two polarization functions are added to the basis sets. In the row of N_2 molecules the data in the parentheses are the population obtained from Mulliken analysis based on prenatural atomic orbitals

Table 4. The Wiberg orbital bonding potentials based on AOIMs for some molecules

Molecules	Basis sets		1*s*	2*s*	3*s*	4*s*	2*p*	3*p*	4*p*	5*p*	3*d*
H_2	MIN		1.000								
	DZ		1.000	0.002							
	DZP		1.000	0.003			0.014				
	TZP		1.000	0.013	−0.004		0.014				
N_2	MIN		−0.001	0.428			2.941				
	DZ		0.003	0.240	0.053	0.000	2.994	0.010			
	DZP		−0.001	0.405	0.059	0.000	2.978	0.020			0.067
	TZ		0.001	0.261	0.092	0.006	2.998	0.084	−0.032		
	TZP		0.000	0.440	0.086	−0.007	2.980	0.075	−0.013		0.131
	MZ		0.001	0.302	0.111	−0.006	2.999	0.149	−0.018	0.001	
CO	MIN	O	−0.001	0.246			2.278				
		C	−0.001	0.543			2.591				
	DZ	O	0.003	0.153	−0.033	0.000	2.400	−0.016			
		C	−0.003	0.357	0.155	0.001	2.530	0.109			
	TZ	O	0.000	0.201	−0.038	0.000	2.380	−0.002	−0.006		
		C	0.000	0.313	0.169	0.000	2.510	0.051	0.011		
	TZP	O	0.000	0.286	−0.025	0.000	2.301	−0.005	−0.002		0.097
		C	0.000	0.452	0.143	0.000	2.456	0.047	0.015		0.101
	MZ	O	0.000	0.197	−0.057	0.001	2.501	0.060	−0.008	0.000	
		C	0.000	0.443	0.115	0.002	2.517	0.111	0.044	0.000	
HF	MIN	F	0.000	0.058			0.694				
		H	0.725								
	DZP	F	0.000	0.008	−0.012	0.000	0.782	−0.010			0.044
		H	0.687	0.029			0.078				
	TZP	F	0.000	0.058	−0.010	0.000	0.722	−0.004	−0.004		0.044
		H	0.645	0.030	0.007		0.082				

LiF	DZP	F	0.000	0.102	0.001	0.000	0.436	0.000			0.012
		Li	0.047	0.007	-0.004	0.000	0.437	0.018			0.038
	MZP	F	0.000	0.165	0.001	0.000	0.413	0.012	0.000	0.000	0.011
		Li	0.069	0.014	0.002	-0.007	0.425	0.019	0.000	0.000	0.037
SF	MIN	S	0.000	0.000	0.083		0.000	2.770			
		F	0.000	0.042			1.312				
	DZ2P	S	0.000	0.000	0.074	0.001	0.000	2.946	0.040		0.481
		F	0.000	0.099	-0.056		1.001	-0.039			0.081
	TZP	S	0.000	0.000	0.103	0.006	0.000	2.931	0.107	-0.011	0.524
		F	0.000	0.086	-0.036	0.000	1.201	-0.001	0.000		0.072
	TZ2P	S	0.000	0.000	0.107	0.005	0.000	2.945	0.105	-0.019	0.544
		F	0.000	0.058	-0.028	0.000	1.171	-0.001	0.000		0.080
	MZ2P	S	0.000	0.000	-0.015	0.038	0.000	2.912	0.129	-0.034	0.502
		F	0.000	-0.111	0.005	0.000	1.127	-0.085	-0.006	0.000	0.056
H_2O	MIN	O	0.000	0.165			1.491				
		H	0.823								
	DZP	O	0.001	0.542	-0.009	0.000	1.932	-0.042			0.063
		H	0.799	0.083			0.397				
	TZP	O	0.000	0.531	-0.021	-0.001	1.773	-0.008	-0.014		0.067
		H	0.762	0.109	-0.007		0.308				0.000
CCl_4	MIN	C	0.000	0.987			2.976				
		Cl	0.000	0.000	0.189		0.000	1.762			
	DZP	C	0.000	0.967	0.131		2.985	0.087			0.140
		Cl	0.000	0.000	0.105	-0.019	0.000	1.695	-0.099		0.102
	TZP	C	0.000	0.991	0.185	0.000	2.943	0.090	-0.123		0.140
		Cl	0.000	0.000	0.114	-0.012	0.000	1.749	0.108	-0.024	0.105

Note: See Table 3 for details

where q_i is the Mulliken population on orbital i. It is obvious that when the electron population on orbital i is zero or two, $b_i = 0$, which means that orbital i contributes nothing to bonding. This is in accordance with the general concept of chemists. Thus b_i can be used to measure the bonding potential of orbital i. It should be noted that $q_i = (PS)_{ii} = \sum_j^{\text{all}} P_{ij}S_{ij}$, includes the overlap charges between orbital i and all other orbitals and correlates to Mulliken bond orders. Usually, when the extended basis sets are used in the calculations, a basis function cannot represent an atomic orbital. Furthermore, the basis functions centered at one atom are not necessarily orthogonalized to each other. Therefore b_i will involve the unphysical "bond orders" within an atom. Thus it cannot be used to measure the bonding potential of basis function i. When AOIMs are taken as basis sets, there is no such problem. The bonding potentials calculated based on AOIMs for some molecules under different basis sets are listed in Table 4. It can be seen from the data in Table 4 that in general the Wiberg orbital bonding potential calculated based on AOIMs is fairly stable with respect to changes of the basis sets used in the molecular calculations. The results are reasonable. For example, the bonding potential for the core orbitals and for the orbitals in Rydberg subset is very small, while the bonding potential of the main valence orbitals is in general large. It is weaker for 2*s* than that of 2*p* orbitals, and so on. The Li 2*p* orbitals in LiF and S 3*d* orbitals in SF possess comparatively larger bonding potential, showing that they significantly participate in bonding, which is consistent with the experiments.

3.3. Bonding extent between atoms: the bond orders

The Bonding extent between atoms in a molecule is a topic discussed very often among chemists. Mulliken proposed in 1955 [1] to measure it by the overlap charge, that is, the Mulliken bond order M_{AB}:

$$M_{\mathrm{AB}} = \sum_{i \in \mathrm{A}} \sum_{j \in \mathrm{B}} 2P_{ij}S_{ij}. \tag{13}$$

Mulliken bond order has been distributed widely among chemists. Its flaw is that the values of the bond order are considerably different from those conventional to the chemists. For example, for the typical single, double and triple bonds the Mulliken bond order values often differ considerably from 1, 2 and 3, respectively. Furthermore, the values are not quite stable to changes of the basis sets. Wiberg [27] defined the bond order between atoms A and B, W_{AB}, as

$$W_{\mathrm{AB}} = \sum_{i \in \mathrm{A}} \sum_{j \in \mathrm{B}} (P_{ij})^2. \tag{14}$$

This definition, however, can only be adopted for orthogonalized basis sets. Mayer [28] extended it to non-orthogonalized basis sets and defined the bond order B_{AB} as

$$B_{\mathrm{AB}} = \sum_{i \in \mathrm{A}} \sum_{j \in \mathrm{B}} 2[(P^{\alpha}S)_{ij}(P^{\alpha}S)_{ji} + (P^{\beta}S)_{ij}(P^{\beta}S)_{ji}], \tag{15}$$

where P^{α} and P^{β} are the density matrices of α and β electrons, respectively. For the typical chemical bonds the values of Mayer bond order are very close to those conventional to chemists. For example, the Mayer bond order between N–N atoms in N_2 molecules is $B_{\mathrm{AB}} = 3.0$. Furthermore, B_{AB} is rather insensitive to changes of the basis sets.

Table 5. The Mulliken and Mayer bond orders under different basis sets

Molecules	Basis sets	Mulliken bond orders		Mayer bond orders	
		AO	$\{AOIM\}_B$	AO	$\{AOIM\}_B$
H_2	MIN	0.884	0.884	1.000	1.000
	DZ	0.814	0.844	1.000	1.000
	DZP	0.832	0.888	1.000	1.000
	TZP	0.832	0.890	1.000	1.000
	SD[a]	0.015	0.011	0.000	0.000
N_2	MIN	1.073	1.073	3.000	3.000
	DZ	0.681	0.936	2.750	2.780
	DZP	1.237	1.216	2.896	2.820
	TZ	0.789	0.972	2.744	2.732
	TZP	1.382	1.316	2.934	2.816
	MZ	1.004	1.146	2.857	2.766
	SD	0.108	0.059	0.042	0.039
CO	MIN	0.888	0.888	2.411	2.411
	DZ	0.603	0.728	2.307	2.270
	TZ	0.598	0.808	2.276	2.259
	TZP	0.894	0.900	2.293	2.215
	MZ	0.370	0.656	2.144	2.159
	MZP	0.705	0.804	2.197	2.288
	SD	0.081	0.038	0.038	0.034
HF	MIN	0.352	0.352	0.725	0.725
	DZP	0.422	0.362	0.741	0.700
	TZP	0.488	0.426	0.723	0.672
	SD	0.039	0.023	0.006	0.016
LiF	MIN	0.040	0.040	0.476	0.476
	DZ	0.029	0.038	0.342	0.345
	DZP	0.342	0.260	0.527	0.523 (0.430)[b]
	TZP	0.400	0.276	0.598	0.594 (0.455)
	MZP	0.372	0.270	0.545	0.544 (0.418)
	SD	0.017	0.005	0.021	0.021
SF	MIN	0.574	0.574	1.139	1.139
	DZ2P	0.431	0.608	0.946	0.970 (0.160)[b]
	TZP	0.117	0.490	1.126	1.149 (0.186)
	TZ2P	0.183	0.490	1.090	1.115 (0.181)
	MZ2P	0.520	0.454	0.937	1.026 (0.172)
	SD	0.091	0.029	0.044	0.035
H_2O (O–H)	MIN	0.509	0.509	0.798	0.798
	DZP	0.914	0.914	1.181	1.153
	TZP	0.796	0.810	1.075	1.037
	SD	0.120	0.121	0.114	0.105

Table 5. (Continued)

Molecules	Basis sets	Mulliken bond orders		Mayer bond orders	
		AO	$\{AOIM\}_B$	AO	$\{AOIM\}_B$
CCl_4	MIN	0.396	0.396	0.972	0.972
(C–Cl)	DZP	0.479	0.524	0.995	0.978
	TZ	0.360	0.442	0.975	0.954
	TZP	0.560	0.526	1.030	0.991
	SD	0.045	0.038	0.013	0.008
	TSD[c]	0.069	0.040	0.035	0.032

[a] Standard deviation

[b] The contribution of Li 2*p* or S 3*d* orbitals to the Mayer bond orders

[c] Weighted standard deviation averaged over all molecules

In Table 5 are listed the Mulliken bond orders and Mayer bond orders, and those calculated based only on subset $\{AOIM\}_B$. Different basis sets have been used and the standard deviation to the changes of the basis sets are listed.

It can be seen from Table 5 that the Mulliken bond orders calculated based on AOIMs are more stable with respect to the changes of basis sets than the conventional ones. In comparison to Mulliken bond orders, the Mayer bond orders are, in consistency with the previous conclusion [28], more stable to the changes of basis sets. The Mayer bond orders calculated based on $\{AOIM\}_B$ are more stable with respect to the changes of basis sets than the conventional ones. The Mayer bond order can be decomposed into the contributions of individual subshells of atoms, which are also stable to the changes of basis sets. The contribution of a subshell to Mayer bond orders calculated based on $\{AOIM\}_B$ can be used to objectively measure their contribution to bonding. For example, the contribution of Li 2*p* orbitals in LiF to Mayer bond orders is over 75%, showing that they are the main bonding orbitals. In SF the contribution of S 3*d* orbitals to Mayer bond orders is up to 17%, thus it cannot be neglected.

4. Concluding remarks

In this paper, it is proposed to perform the population and bonding analyses based on the AOIMs when the extended basis sets are used. The AOIMs are defined as the solution of the Schroedinger equation of an electron moving in the potential field which is the spherical average of the molecular potential field with the center at one atomic nucleus. The method to construct the AOIMs from the Fock matrix of molecular calculations is presented. The results of the practical calculations for a series of molecules show that AOIMs can be clearly divided into two subsets. According to the Mulliken population on individual orbitals, the strongly occupied minimal compact subset $\{AOIM\}_B$ and the very weakly occupied "Rydberg" subset $\{AOIM\}_R$ can be distinguished. The formulation of "the molecular orbitals consisting of minimal atomic orbital sets" can be optimally realized in terms of $\{AOIM\}_B$. Through addition of very weakly occupied subset $\{AOIM\}_R$ to construct the extended basis sets the molecular electronic structure can be described accurately. In terms of AOIMs the results of population analyses and the calculated Wiberg orbital bonding potentials are fairly stable to the changes of basis sets. The results of the bonding analysis based on AOIMs are reasonable and quite stable to the changes of basis sets. Mayer bond orders are more stable than Mulliken bond orders. The Mayer bond orders based on $\{\mathrm{AOIM}\}_B$ are even more stable. Thus it can be used to objectively measure the contribution of individual atomic orbitals to bonding.

Acknowledgements. This work has been supported by State Major Key Project for Basic Researches and National Natural Science Foundation of China, for which the authors gratefully acknowledge.

References

1. Mulliken RS (1955) J Chem Phys 23:1833, 1841, 2343; (1978) Ann Rev Phys Chem 29:1
2. Xu G, Li L, Wang D (1985) Quantum chemistry: Fundamental principle and ab initio calculations, Vol 2. Science Press, Beijing, p 798
3. Davidson ER (1967) J Chem Phys 46:3320
4. Roby KR (1974) Mol Phys 27:81
5. Heinzmann R, Ahlrichs R (1976) Theoret Chim Acta 42:33
6. Ehrhardt C, Ahlrichs R (1985) Theoret Chim Acta 68:231
7. Reed AE, Weinstock RB, Weinhold F (1985) J Chem Phys 83:735; Reed AE, Curtiss LA, Weinhold F (1988) Chem Rev 88:899
8. Cioslowski J (1989) J Am Chem Soc 111:8333
9. Huzinaga S, Sakai Y, Miyoshi E, Narita S (1990) J Chem Phys 93:3319
10. Craig DP, Maccoll A, Nyholm RS, Orgel LE, Sutton LE (1954) J Chem Soc p 332
11. Craig DP, Zauli C (1962) J Chem Phys 37:601
12. Mitchell KAR (1968) J Chem Soc A 2676, p 2683
13. Xiao L, Li L, Liang Z (1989) Acta Chim Sinica 4:295
14. Li L, Xiao L, Liang Z (1992) Chinese J Chem 10:97
15. Mitchell KAR (1969) Chem Rev 69:157
16. Huheey JE (1983) Inorganic chemistry, 3rd ed. Harper & Row, New York, p 824
17. Cotton FA, Wilkinson G (1988) Advanced inorganic chemistry, 5th ed. Wiley, New York
18. Li L, Xiao L (1990) Special collection of academic theses in celebration of Professor Tang Aoqing's teaching and researching for five decades. Jilin University Press, Changchun, p 187
19. Hehre WJ, Radom L, Schleyer PvR, Pople JA (1986) Ab initio molecular orbital theory. Wiley-Interscience, New York
20. Ziegler T (1991) Chem Rev 91:651
21. Laming GJ, Termath V, Handy NC (1993) J Chem Phys 99:8765
22. Liu Wenjian (1995) PhD Thesis, Peking University, China; Liu Wenjian, Hong Gongyi, Li Lemin (1996) Chem Res and Applic 3:369
23. Theoretical Chemistry, Vrije University (1994) Amsterdam Density Functional (ADF 1.1.3)
24. Clementi E, Roetti C (1974) At Data Nucl Data Tables 14:428
25. Bunge CF, Barrientos JA, Bunge AV (1993) At Data Nucl Data Tables 53:118
26. Pauling L (1960) The nature of the chemical bond, 3rd ed. Cornell University Press, Ithaca, New York
27. Wiberg KA (1968) Tetrahedron 24:1083
28. Mayer I (1983) Chem Phys Lett 97:270; (1983) Intern J Quantum Chem 23:341; (1986) Intern J Quantum Chem 29:73