Atomic partitioning schemes for wavefunction analysis and effective atomic orbitals

Pedro Salvador

Institut de Quimica Computacional i Catalisis i Departament de Quimica,
Universtitat de Girona, Facultat de Ciencies,
c/M Aurelia Capmany 69, 17003, Girona
email:pedro.salvador@udg.edu

CTTC2019 School, Quito, 29-30 June 2019

What is a molecule?

A system of electrons and nuclei (Physicist picture)

Physical Information:

- Hamiltonian, Wavefunction (density) Hermitian operators (functionals) associated to observables and their expectation values

A system of bonded (interacting) atoms (Chemist picture)

Chemical Information:

- Bonding interaction, steric repulsion, polarity, functional group, aromaticity...

We aim to interpret the results of ab initio calculations in classical chemical terms

The Atom in the Molecule

Atoms are the building blocks of chemistry

- Not observables in Quantum Mechanics.
- Is there a unique atom in molecule (AIM) definition?
- Any AIM is a conceptual construct but with an irrefutable utility.
- Taking advantage of AIM, quantum chemist have defined bond orders indexes, atomic populations, bond orders descriptors, energy partitioning, aromaticity indexes, among others.

Assigning physical quantities to the individual atoms

- Any physical quantity is written as expectation value of one- or two-electron operators
- Which is the physical quantity that can be related to a given chemical concept?

The Atom in the Molecule

- The nucleus is always considered as part of the atom, so the differences always arise in how the physical space (or the Hilbert-space) is subdivided into atomic shares.

- In the MO-LCAO approach (Molecular Orbitals as Linear Combination of Atomic Orbitals), the atom may be identified with the subspace of the basis functions attached to it. Such approaches are the so-called Hilbert-space analyses.

- A different strategy is to subdivide the physical 3D space into atomic regions or domains, which represent the atom. They may be defined disjoint, like in Bader's "quantum theory of atoms in molecules" (QTAIM), or may be allowed to overlap, like in the different variants of the "fuzzy" atoms

- In Hilbert-space analysis one deals with the AO matrix representation of operators. In 3D-space analysis one usually manipulates density functions

Hilbert-space analysis

In the LCAO framework, the MOs are expanded on a finite set of atomic-centered one-electron functions $\{\chi_\mu(\mathbf{r})\}$ as

$$\varphi_i(\mathbf{r}) = \sum_\mu c_{\mu i} \chi_\mu(\mathbf{r})$$

The set of functions centered on a given atom conform an atomic Hilbert subspace $\{\chi_\mu(\mathbf{r})\}_{\mu \in A}$

In this context, the atom may be defined by its nucleus and the subspace of one-electron basis set centered on it.

The molecular orbitals can be also rewritten as a sum of atomic contributions, coming from each atomic Hilbert's subspace.

$$\varphi_i(\mathbf{r}) = \sum_A \sum_{\mu \in A} c_{\mu i} \chi_\mu(\mathbf{r})$$

Hilbert-space analysis

The first-order electron density can be written as

$$\rho(\mathbf{r}) = \sum_{\mu\nu} D_{\mu\nu} \chi_\nu^*(\mathbf{r})\chi_\mu(\mathbf{r}),$$

where $\mathbf{D}$ is the density matrix in the atomic orbital (AO) representation. By integrating the density one obtains the total number of electrons

$$N = \int \rho(\mathbf{r})d\mathbf{r} = \sum_{\mu\nu} D_{\mu\nu} \int \chi_\nu^*(\mathbf{r})\chi_\mu(\mathbf{r})d\mathbf{r} = \sum_{\mu\nu} D_{\mu\nu} S_{\nu\mu}^{AO} = \sum_\mu (\mathbf{D}\mathbf{S}^{AO})_{\mu\mu},$$

where matrix $\mathbf{S}^{AO}$ is the atomic **overlap** matrix in AO basis, with elements

$$S_{\nu\mu}^{AO} = \int \chi_\nu^*(\mathbf{r})\chi_\mu(\mathbf{r})d\mathbf{r}$$

The basis functions that play the role of AOs in electronic structure calculations, usually Gaussian-type orbitals (GTO) or Slater-type orbitals (STO) do exhibit overlap.

Now, by systematizing the summation over the AOs according to which atom A they are assigned to, one can readily obtain the so-called **Mulliken's atomic gross populations**, N_A

$$N = \sum_A \sum_{\mu \in A} (\mathbf{DS}^{AO})_{\mu\mu} = \sum_A N_A.$$

Alternatively, one can also start from the previous expression in terms of two AO indices and systemtize them over each atom A and B, and write

$$N = \sum_A \sum_B \sum_{\mu \in A} \sum_{\nu \in B} D_{\mu\nu} S^{AO}_{\nu\mu} = \sum_A N_{AA} + \sum_{A \neq B} N_{AB}$$

where N_{AA} and $N_{AB} + N_{BA}$ are **Mulliken's net and overlap populations**, respectively.

Hilbert-space analysis

Pros

- The integrations are analytical for Gaussian functions.
- Straightforward implementation and low computational cost
- The errors associated with any additional numerical integration are avoided.

Cons

- **Basis set dependency**.
- It may show unphysical results for basis sets without marked atomic character e.g. diffuse basis functions.
- Ill-defined in the complete basis set limit (CBS).
- Inapplicable beyond the LCAO-MO approximation e.g. if plane waves are used to expand the MOs of the system.

Hilbert-space analysis

Beyond Mulliken Analyses

- Most of the problems associated with Mulliken analyses originate from the overlap of the underlying atomic basis.
- When basis set without marked atomic character are used:
  Overlap populations or bond orders can be negative
  Effective atomic orbitals with occupations not in the [0,1] range

Use of orthogonal basis

The AO basis can be transformed to an orthogonal basis, and the MO expanded into this orthonormalized basis set.

- Symmetric orthogonalization of Lowdin (and a number of variants, see e.g. Comput. Theor. Chem. 1008 (2013) 15)
- Natural orbitals and Natural Population analysis of Weinhold
- Quasi-atomic orbitals of Ruedenberg, etc..

Hilbert-space analysis in orthogonal basis

For instance, in the Lowdin basis one has (S matrix in AO basis)

$$\chi_i^L(\mathbf{r}) = \sum_\mu S_{\mu i}^{-1/2} \chi_\mu(\mathbf{r}) \quad \chi_\mu(\mathbf{r}) = \sum_i S_{i\mu}^{1/2} \chi_i^L(\mathbf{r})$$

Replacing the original AOs by the Lowdin orthonormalized ones

$$\rho(\mathbf{r}) = \sum_{\mu\nu} D_{\mu\nu} \chi_\nu^*(\mathbf{r})\chi_\mu(\mathbf{r}) = \sum_{\mu\nu} D_{\mu\nu} \left[\sum_i S_{i\nu}^{1/2\,*} \chi_i^{L\,*}(\mathbf{r}) \right] \left[\sum_j S_{j\mu}^{1/2} \chi_j^L(\mathbf{r}) \right] =$$

$$= \sum_{\mu\nu} \sum_{ij} S_{j\mu}^{1/2} D_{\mu\nu} S_{\nu i}^{1/2\,\dagger} \chi_i^{L*}(\mathbf{r})\chi_j^L(\mathbf{r}) = \sum_{ij} D_{ji}^L \chi_i^{L*}(\mathbf{r})\chi_j^L(\mathbf{r})$$

where $D_{ji}^L = \left[\mathbf{S}^{1/2}\mathbf{D}\mathbf{S}^{1/2\,\dagger} \right]_{ji}$.

In this new basis, the integration of the density leads to

$$N = \int \rho(\mathbf{r})d\mathbf{r} = \sum_{ij} D_{ji}^L \delta_{ij} = \sum_i D_{ii}^L$$

Lowdin's atomic populations are obtained by systematizing the indices of the orthogonal AO basis, whose elements are also associated to the atomic centers, i.e. $\left\{\chi_i^L(\mathbf{r})\right\}_{i \in A}$.

$$N = \sum_A \sum_{i \in A} (D^L)_{ii} = \sum_A N_A^L$$

Notice that in an orthogonal basis there is no overlap population.

Assigning orthogonalized AOs to atomic centers might be ambiguous in the case of extended basis sets.

Real-space analysis

Disjoint domains

- Voronoi cells (e.g. VDD charges)
- Quantum Theory of Atoms in Molecules (QTAIM)
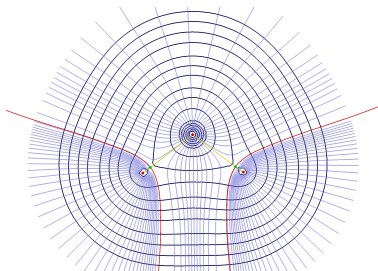- Topological analysis of vector fields other than $\rho(\mathbf{r})$

Overlapping domains ("fuzzy" atoms)

- Hirshfeld, Hirshfeld-Iterative, etc...(Ayers)
- Becke's constructs, Becke-ρ, TFVC
- Iterative stockholder Atoms (ISA), etc...

3D-space partitions make the results virtually basis set independent.

QTAIM

- Topological analysis of $\rho(\mathbf{r})$
- Atomic basins from the zero-flux condition

$$\nabla \rho(\mathbf{r}) \cdot \vec{n}(\mathbf{r}) = \vec{0} \quad \forall \mathbf{r} \in S(\mathbf{r})$$

$$\int f(\mathbf{r}) \, d\mathbf{r} = \sum_A \int_{\Omega_A} f(\mathbf{r}) \, d\mathbf{r}$$

$$\iint f(\mathbf{r}, \mathbf{r}') \, d\mathbf{r} d\mathbf{r}' = \sum_{AB} \int_{\Omega_A} \int_{\Omega_B} f(\mathbf{r}, \mathbf{r}') \, d\mathbf{r} d\mathbf{r}'$$

QTAIM

Pros

- Strong physical background (see notes from A.M. Pendás)
- Only information from $\rho(\mathbf{r})$ is required.

Cons

- Complex shape of the atomic domains leads to cumbersome numerical integrations (but more and more efficients codes are being developed, e.g. AIMAll)

Overlapping domains

One can define a weight function for each atom at each point of the space satisfying

$$w_A(\mathbf{r}) > 0 \quad \text{and} \quad \sum_A w_A(\mathbf{r}) = 1$$

- Atoms can share the 3D-space (overlap) in general.
- In QTAIM $w_A(\mathbf{r}) = 1$ if $\mathbf{r} \in \Omega_A$ and $w_A(\mathbf{r}) = 0$ otherwise.

$$\int f(\mathbf{r}) \, d\mathbf{r} = \sum_A \int w_A(\mathbf{r}) f(\mathbf{r}) \, d\mathbf{r}$$

$$\iint f(\mathbf{r}, \mathbf{r}') \, d\mathbf{r} d\mathbf{r}' = \sum_{AB} \iint w_A(\mathbf{r}) w_B(\mathbf{r}') f(\mathbf{r}, \mathbf{r}') \, d\mathbf{r} d\mathbf{r}'$$

Different "fuzzy" atom approaches

Hirshfeld approach, TCA 44 129 (1977).

- Makes use of reference (promolecular) spherically-averaged atomic densities, $\rho_A^0(\mathbf{r})$
- Superposition of promolecular atomic densities at the actual atomic positions of the molecule defines the promolecular total density , $\rho^0(\mathbf{r}) = \sum_B \rho_B^0(\mathbf{r})$
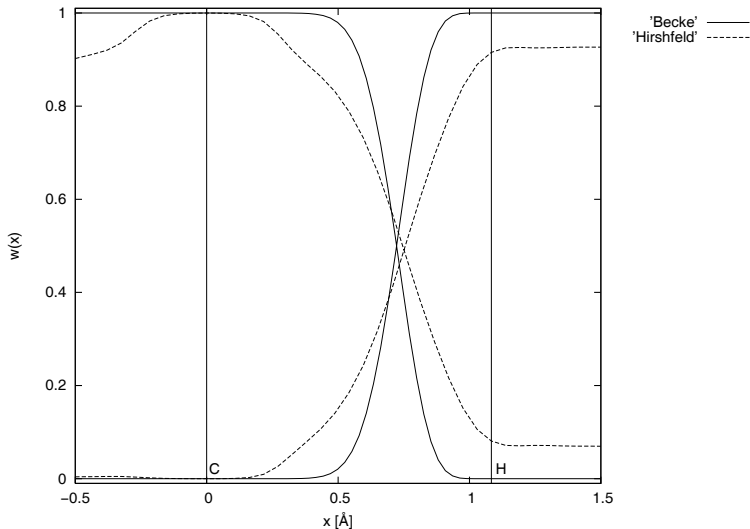
Then, in the classical Hirshfeld method

$$w_A(\mathbf{r}) = \frac{\rho_A^0(\mathbf{r})}{\sum_B \rho_B^0(\mathbf{r})}$$

Hirshfeld-Iterative schemes usually impose $\int \rho_A^0(\mathbf{r}) d\mathbf{r} = N_A \neq Z_A$, e.g. JCP 126 144111 (2007).

$$\rho_A^{0,N_A}(\mathbf{r}) = \alpha \rho_A^{0,int(N_A)+1}(\mathbf{r}) + (1-\alpha)\rho_A^{0,int(N_A)}(\mathbf{r}), \ \ \alpha = N_A - int(N_A)$$

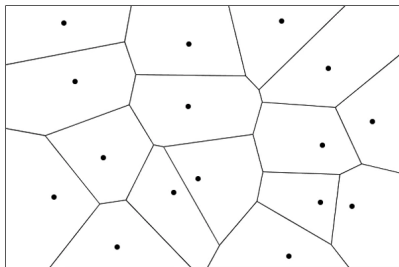**The value of the atomic weight functions on the nuclear position can deviate significantly from 1**

- Originally devised for effective numerical integration of tree-dimensional functions of marked atomic character.
- Makes use of the so-called Voronoi polyhedra

- The sharp boundaries of the Voronoi cells are substituted by soft functions that ultimately define the atomic weigths
- By construction, $w_A^{becke}(\mathbf{r}_A) = 1$ and $\partial w_A^{becke}(\mathbf{r})/\partial(r)|_{r=R_A} = 0$.

Becke's constructs

The scheme can be formulated as follows for the simplest case of a two nuclei system (A and B). For any point of the space one can define the following quantity

$$\mu_{AB} = \frac{r_B - r_A}{R_{AB}}$$

where R_{AB} is the internuclear distance and r_A and r_B represent the distance of that point to nucleus A and B.
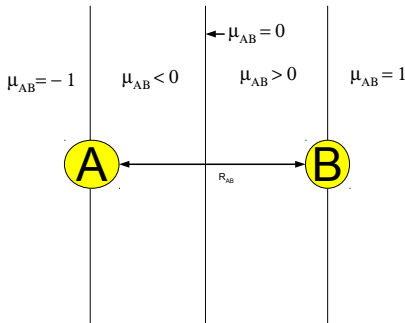
Becke's constructs

The simple step function

$$s_A(\mu_{AB}) = \begin{cases} 1 & -1 \leq \mu_{AB} \leq 0 \\ 0 & 0 < \mu_{AB} \leq 1 \end{cases}$$

can be used to define the sharp Voronoi cell of atom A in this case.

The step function can be replaced by a continuous, monotonically decreasing function in the range [-1,1], and fulfilling the requirements $s(-1)=1$ and $s(+1)=0$ in order to define **fuzzy Voronoi cells**

For that purpose Becke suggested the simple polynomial function

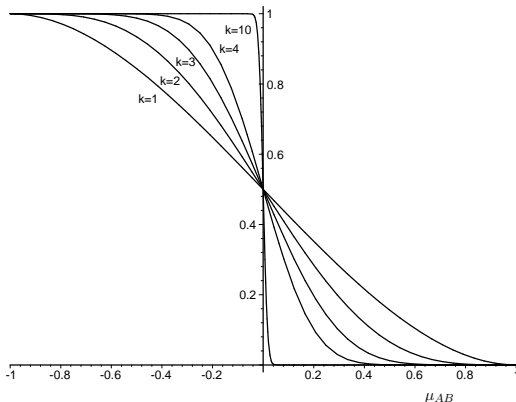$$s_A^k(\mu_{AB}) = \frac{1}{2}[1 - f_k(\mu_{AB})]$$

where

$$f_1(\mu) = \frac{3}{2}\mu - \frac{1}{2}\mu^3$$

Becke's constructs

And devised an iterative process to obtain shaper cutoff profiles

$$f_k(\mu) = f[f_{k-1}(\mu)]$$

The integer k is known as the **stiffness parameter**; the larger the steeper the cutoff profile is.

Becke constructs

In order to account for the different atomic sizes in heteronuclear systems, one can use the same cutoff function but introduce a shifted coordinate. Becke originally uses

$$\nu_{AB} = \mu_{AB} + a_{AB}(1 - \mu_{AB}^2) \text{ where } \quad a_{AB} = \frac{1 - \chi_{AB}^2}{4\chi_{AB}}$$

$-1/2 \leq a_{AB} \leq 1/2$ to ensure that $0 \leq w_A(\mathbf{r}) \leq 1$ and $\frac{R_A^0}{R_B^0} = \chi_{AB}$

- For overlapping atoms the position of the interatomic boundaries between all pairs of atoms must be specified.
- A distance criterion is used, namely if $R_{AB} > 2(R_A^0 + R_B^0)$ then $\chi_{AB} = 1$. Otherwise, the atoms are **neighbors** and $\chi_{AB} = \frac{R_A^0}{R_B^0}$

Becke constructs

- The set of atomic radii R_A^0 determines the relative size of the atomic Voronoi cells
- The parameter k controls the shape of the cutoff profile.
- Becke used Bragg-Slater radii and k=3 on the basis of more accurate integrations.

Pros

- Numerical integrations in the fuzzy domains are very efficient
- Does not required the use of promolecular atomic densities

Cons

- The use of a fixed atomic radii is a limitation of this AIM model because the same atoms are treated on equal footing in different chemical environments
- Highly arbitrary definition of the atom in the molecule

Becke and QTAIM

Topological fuzzy Voronoi cells (TFVC)

- Instead of using a set of fixed radii to define χ_{AB} one can use position of the minimum of the density along the internuclear axis connecting two neighbor atoms.

- The following alternative transformation is monotonic for any value of χ_{AB}.
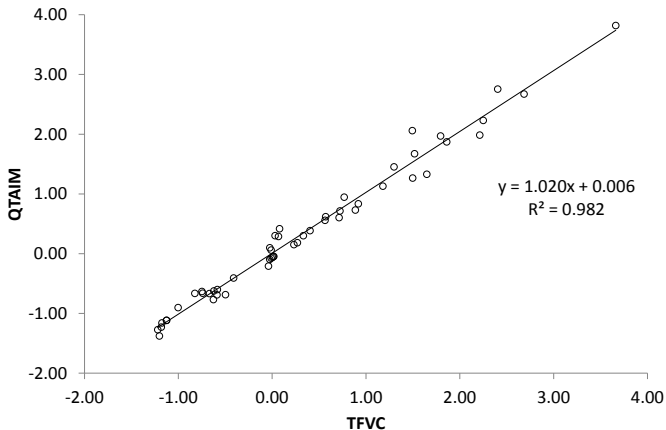
$$\nu'_{AB} = \frac{1 + \mu_{AB} - \chi_{AB}(1 - \mu_{AB})}{1 + \mu_{AB} + \chi_{AB}(1 - \mu_{AB})}$$

- Two atoms are not considered neighbors if their midpoint is closer to a third atom. Empirical set of radii no longer needed.
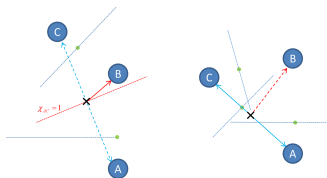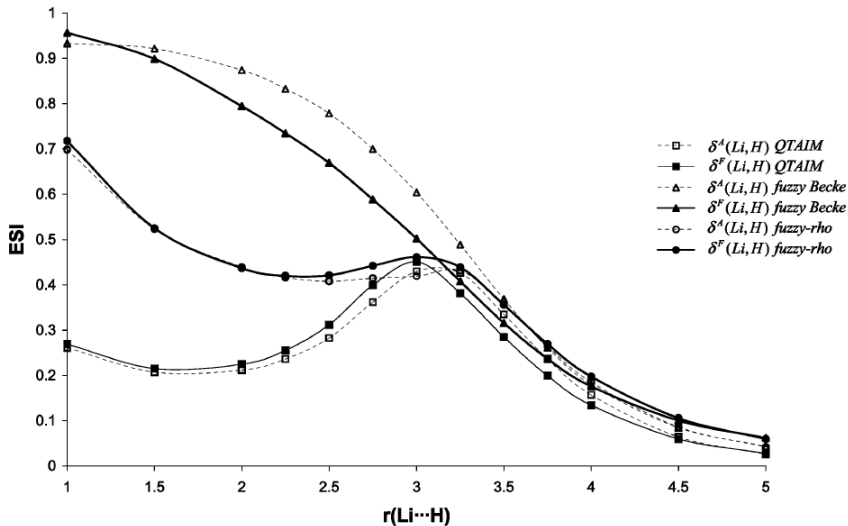
TFVC vs QTAIM

Partial atomic charges for a set of hydrides (LiH to HCl)

Conditions for proper AIM definitions

- An AIM should be able to account for complex bonding mechanisms (e.g. harpoon effect on LiH)

Conditions for proper AIM definitions

- In aromatic systems, the delocalization index (bond order) should be larger in para- position than in meta-.

Table 7. HF/6-311++G** ESI Values (Meta-ESI and PDI) for Benzene calculated Using Different Atomic Partitions (see text).

	meta-ESI	PDI
QTAIM	0.074	0.098
Becke $k = 3$	0.108	0.101
Becke $k = 4$	0.093	0.101
Becke $k = 5$	0.084	0.102
Becke-rho $k = 3$	0.093	0.092
Becke-rho $k = 4$	0.085	0.094
Becke-rho $k = 5$	0.079	0.095
ISA	0.191	0.122
Hirshfeld	0.192	0.114
Hirshfeld-I	0.203	0.120
Hirshfeld-ID	0.076	0.103

Identifying more conditions is desirable...

Instead of partitioning the physical space one can make use of the atomic weight functions $w_A(\mathbf{r})$ to define *effective atomic density functions*:

$$N_A = \int w_A(\mathbf{r})\, \rho(\mathbf{r})d\mathbf{r} \;\; \rightarrow \rho_A(\mathbf{r}) \equiv w_A(\mathbf{r})\rho(\mathbf{r})$$

such that integration over the whole space leads to the corresponding atomic contribution

$$N_A = \int \rho_A(\mathbf{r})d\mathbf{r}$$

In the case of two-electron functions such as the so-called exchange-correlation pair density, $\rho_{xc}(\mathbf{r}_1, \mathbf{r}_2)$, the same strategy naturally leads to atomic and di-atomic densities

$$\rho_{xc,AB}(\mathbf{r}_1, \mathbf{r}_2) \equiv w_A(\mathbf{r}_1)w_B(\mathbf{r}_2)\rho_{xc}(\mathbf{r}_1, \mathbf{r}_2) \;\; + \;\; A \leftrightarrow B$$

The delocalization index (or bond order) between atoms A and B is readily obtained by integrating the diatomic exchange-correlation pair density

$$DI_{AB} = \int \int \rho_{xc,AB}(\mathbf{r}_1, \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2$$

Similarly, diatomic exchange-correlation energy contributions in the context of the Interacting Quantum Atoms (IQA)

$$E_{xc,AB} = -\frac{1}{2} \int \int \frac{\rho_{xc,AB}(\mathbf{r}_1, \mathbf{r}_2)}{|r_1 - r_2|} d\mathbf{r}_1 d\mathbf{r}_2$$

Atomic and/or diatomic terms could also be easily grouped into molecular fragments or functional group contributions.

A different perspective for Mulliken analysis, too

Such effective density functions can also be defined in the context of Hilbert-space analysis, again when expressed in the AO basis. Thus, by defining the effective atomic density as

$$\rho_A(\mathbf{r}) \equiv \sum_{\mu \in A} \sum_{\nu} D_{\mu\nu} \chi_\nu^*(\mathbf{r}) \chi_\mu(\mathbf{r})$$

one can recover Mulliken's gross population of the atom upon integration

$$N_A = \int \rho_A(\mathbf{r}) d\mathbf{r}$$

But Mulliken-type effective atomic densities are not well behaved in general (e.g. can be negative), only the one corresponding to the **net** atomic density:

$$\rho_{AA}(\mathbf{r}) \equiv \sum_{\mu \in A} \sum_{\nu \in A} D_{\mu\nu} \chi_\nu^*(\mathbf{r}) \chi_\mu(\mathbf{r})$$

The realization of the quantities obtained in real-space analysis is usually carried out in a molecular orbital basis, $\{\phi_i(\mathbf{r})\}$.

Let us consider now the first-order density matrix in MO basis

$$\rho(\mathbf{r}; \mathbf{r}') = \sum_{ij} D_{ji} \phi_i^*(\mathbf{r}) \phi_j(\mathbf{r}')$$

It can also be written in diagonal form, leading to the so-called natural orbitals, $\{\Phi_i(\mathbf{r})\}$, and natural occupations, $\{n_i\}$.

$$\rho(\mathbf{r}; \mathbf{r}') = \sum_i n_i \Phi_i^*(\mathbf{r}) \Phi_i(\mathbf{r}')$$

The electron density is achieved by doing $\mathbf{r} = \mathbf{r}'$,

$$\rho(\mathbf{r}) \equiv \rho(\mathbf{r}; \mathbf{r})$$

Real-space analysis: working on MO basis

Just as we've seen before, we can make use of the atomic effective matrices to obtain real-space quantites upon integration, such as the atomic populations

$$N_A = \int \rho_A(\mathbf{r})d\mathbf{r} = \sum_{ij} D_{ji} \int w_A(\mathbf{r})\phi_i^*(\mathbf{r})\phi_j(\mathbf{r})d\mathbf{r} = \sum_{ij} D_{ji}S_{ij}^A = tr(\mathbf{DS}^A).$$

As the summation indices run over all molecular orbitals, the expressions can be conveniently written in matrix form, in terms of the **atomic overlap matrix** in MO basis, $\mathbf{S}^A$.

To further illustrate the formal use of the effective atomic densities, let us consider the particular case of a restricted single-determinant wavefunction in singlet state. In this case, the exchange density can be written in terms of the first-order density matrix

$$\rho_x(\mathbf{r}_1; \mathbf{r}_2) = -\frac{1}{2}\rho(\mathbf{r}_1; \mathbf{r}_2)\rho(\mathbf{r}_2; \mathbf{r}_1)$$

We construct the effective diatomic exchange density for centers A and B

$$\rho_{x,AB}(\mathbf{r}_1; \mathbf{r}_2) = -\frac{1}{2}\rho_A(\mathbf{r}_1; \mathbf{r}_2)\rho_B(\mathbf{r}_2; \mathbf{r}_1) \ + \ A \leftrightarrow B$$

and obtain again their Delocalization Index upon integration, expressed in matrix form in the MO basis.

$$DI_{AB} = \int\int \rho_{x,AB}(\mathbf{r}_1, \mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2 = \sum_{ijkl} D_{ji}S_{ik}^A D_{kl}S_{lj}^B = tr(\mathbf{D}\mathbf{S}^A\mathbf{D}\mathbf{S}^B).$$

It can be readily seen that in natural orbital (NO) basis it reads

$$DI_{AB} = \sum_{ij} n_i n_j S_{ij}^A S_{ji}^B.$$

Hilbert-space analysis in MO basis?

Let us introduce the atomic truncation matrix, $\boldsymbol{\eta}^A$, in AO basis, which has zero elements except if $i = j, i \in A$. With it, one can rewrite Mulliken's atomic population without restrictions in the summation indices

$$N_A = \sum_\nu \sum_{\mu \in A} D_{\nu\mu} S_{\mu\mu}^{AO} = \sum_{\nu\mu\lambda} D_{\nu\mu} S_{\mu\lambda}^{AO} \eta_{\lambda\nu}^A = tr(\mathbf{D} \mathbf{S}^{AO} \boldsymbol{\eta}^A).$$

Comparing this expression in matrix form with the previous real-space one, we can identify the atomic overlap matrix (still in AO basis) $\mathbf{S}^A \equiv \mathbf{S}^{AO} \boldsymbol{\eta}^A$.

Transforming the expressions to the MO basis, making use of the MO coefficients one can stablish direct comparison with the matrix representation of the real-space atomic overlaps in MOs

$$\mathbf{S}^A \Leftarrow \mathbf{C}^\dagger \mathbf{S}^{AO} \boldsymbol{\eta}^A \mathbf{C}.$$

Hilbert-space analysis in MO basis?

Such mapped Mulliken-type atomic overlap matrix is nonhermitian

$$\mathbf{S}^{A\,\dagger} = \mathbf{C}^{\dagger}\boldsymbol{\eta}^{A}\mathbf{S}^{AO}\mathbf{C} \neq \mathbf{S}^{A},$$

However, one can perform similar analogy in the Lowdin basis and obtain a hermitian atomic overlap matrix

$$\mathbf{S}^{A} \Leftarrow \mathbf{C}^{\dagger}\mathbf{S}^{1/2}\boldsymbol{\eta}^{A}\mathbf{S}^{1/2\,\dagger}\mathbf{C}.$$

This is one of the reasons that make Lowdin-type analysis more robust in terms of basis set dependence than the Mulliken-type approaches.

As a final message, we have seen that by using approoapriate mappings one can readily use the **same** real-space in MO/NO basis for Hilbert-space approaches.

Effective atomic orbitals:

the atomic orbitals of the topological atom

The effective atomic orbitals

- One very useful tool to characterize the state of the atom within the molecule are the so-called **effective atomic orbitals** (eff-AOs), Mayer JCP 6249 (1996).
- In this approach one obtains **for each atom** a set of orthogonal atomic hybrids and their respective occupation numbers, adding up to the **net** population of the atom.
- These atomic hybrids closely mimic the core and valence shells of the atom, as anticipated on the basis of classical notions of electron configuration of the atom/fragment within the molecule.
- The general formalism was introduced two decades ago in the framework of Hilbert-space analysis by Mayer, JCP 6249 (1996) and later generalized to QTAIM and real-space analysis in general.

Let us consider a single-determinant WF formed by n orthonormalized doubly occupied orbitals $\varphi_i(\mathbf{r})$, $i = 1, 2 \ldots, n$, and a "fuzzy" division of the 3D-space into N_{at} atomic domains Ω_A defined *e.g.*, by a continuous atomic weight function $w_A(\mathbf{r})$, fulfilling $w_A(\mathbf{r}) > 0$ and $\sum_A w_A(\mathbf{r}) = 1$.

Let us for each atom A $(A = 1, 2, \ldots, N_{at})$ form the $n \times n$ Hermitian matrix $\mathbf{Q}^A$ with the elements

$$Q_{ij}^A = \int w_A(\mathbf{r})\varphi_i^*(\mathbf{r})w_A(\mathbf{r})\varphi_j(\mathbf{r})d\mathbf{r}.$$

The matrix $\mathbf{Q}^A$ is essentially the "net atomic overlap matrix" in the basis of the MO-s φ_i. Furthermore, for each atom A we define the "intraatomic" part φ_i^A of every MO φ_i as $w_A(\mathbf{r})\varphi_j(\mathbf{r})$. Thus $Q_{ij}^A = \langle \varphi_i^A | \varphi_j^A \rangle$ *i.e.*, $\mathbf{Q}^A$ is the overlap matrix of the orbitals φ_i^A.

Theory

We diagonalize the Hermitian matrix $\mathbf{Q}^A$ by the unitary matrix $\mathbf{U}^A$:

$$\mathbf{U}^{A\dagger}\mathbf{Q}^A\mathbf{U}^A = \mathbf{\Lambda}^A = diag\{\lambda_i^A\} \quad .$$

It can be shown that every $\lambda_i^A \geq 0$, as is the case for an overlap matrix. The dimension of the matrix also indicates the max. number of non-zero eigenvalues that can be obtained (n), disregarding the size of the underlying set of AOs

For each atom A we can define n_A ($n_A \leq n$) "effective atomic orbitals $\chi_\mu^A(\mathbf{r})$ as linear combinations of the "intraatomic" parts $\varphi_i^A(\mathbf{r})$ of the MO-s as

$$\chi_\mu^A(\mathbf{r}) = \frac{1}{\sqrt{\lambda_\mu^A}} \sum_{i=1}^n U_{i\mu}^A \varphi_i^A(\mathbf{r}) ; \qquad \mu = 1, 2, \ldots, n_A \quad ,$$

where n_A is the number of non-zero eigenvalues λ_i^A.

Properties of the eff-AOs

- They are orthonormalized within the atomic domain:

$$
\begin{aligned}
\langle \chi_\mu^A | \chi_\nu^A \rangle &= \langle \frac{1}{\sqrt{\lambda_\mu^A}} \sum_{i=1}^n U_{i\mu}^A \varphi_i^A | \frac{1}{\sqrt{\lambda_\nu^A}} \sum_{j=1}^n U_{j\nu}^A \varphi_j^A \rangle = \frac{1}{\sqrt{\lambda_\mu^A \lambda_\nu^A}} \sum_{i,j=1}^n (\mathbf{U}^{A\dagger})_{\mu i} Q_{ij}^A U_{j\nu}^A \\
&= \frac{1}{\sqrt{\lambda_\mu^A \lambda_\nu^A}} \lambda_\mu^A \delta_{\mu\nu} = \delta_{\mu\nu}
\end{aligned}
$$

- The sum of occ. numbers equals the *net* atomic population of the atom
- In the case of disjoint domains (QTAIM), orbitals $\chi_\mu^A(\mathbf{r})$ differ from zero only in the atomic domain of atom A, thus

$$
\langle \chi_\mu^A | \chi_\nu^B \rangle = 0
$$

- In the case of disjoint domains, $\mathbf{Q}^A$ is also the atomic overlap matrix of the MOs, as $w_a(\mathbf{r})^2 \equiv w_a(\mathbf{r})$.

Properties of the eff-AOs

- The eff-AOs of atom A can also be obtained **even if no atom-centered basis functions where used in the original MO description**, see CPL 563 97 (2013).

- They can also be obtained in the AO basis from the diagonalization of the matrix $\mathbf{P}\mathbf{S}^A$, where $\mathbf{P}$ is the LCAO density matrix. This alternative permits the straightforward generalization to correlated WFs.

- Alternatively, in the natural orbital representation, the symmetrized form

$$\mathbf{U}'^{A\dagger}\eta^{1/2}\mathbf{Q}^A\eta^{1/2}\mathbf{U}'^A = \mathbf{\Lambda}^A = \textit{diag}\{\lambda_i^A\}$$

  that also produces the same set of eigenvalues, where η is the diagonal matrix of the natural occupations.

Properties

- In practice, the dimension of the eff-AO basis has nothing to do with the dimension of the LCAO basis (if any) used in the original calculations.
- The maximum number of eff-AOs that can be obtained is $N_{at} \times n$, which can be less that the total number of the LCAO basis functions.
- Experience shows that the number of eff-AOs with significant occupation numbers on each atom is limited, and typically much smaller than the number n of the doubly occupied orbitals in the whole molecule.
- The remaining eff-AOs with very small occupation numbers have a marginal significance.
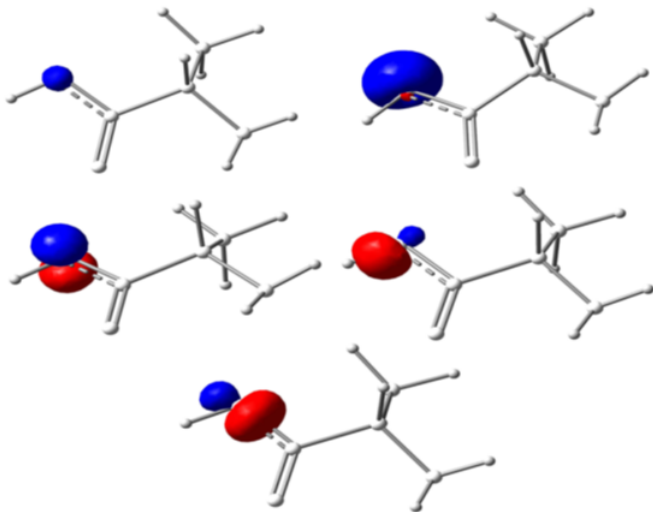
Figure 1: Occupied oxygen eff-AOs in alanine molecule (TFVC)
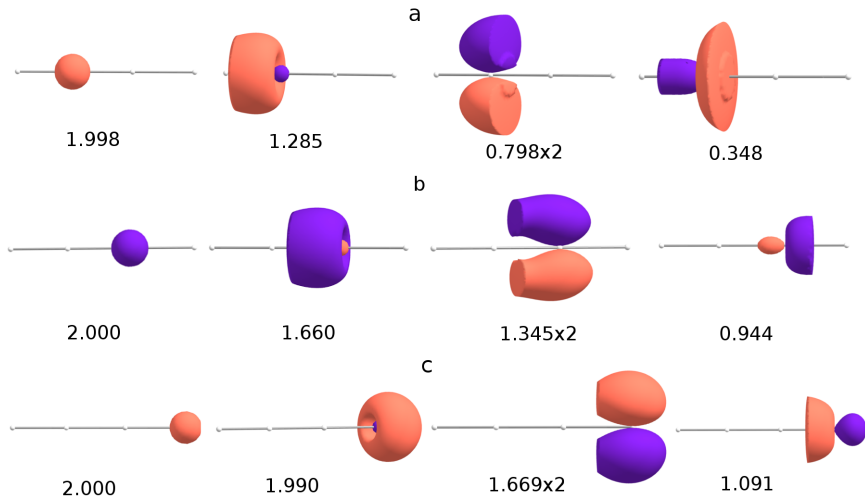
Visualization of the eff-AOs

Figure 2: Highly occupied carbon (a), nitrogen (b), and oxigen (c) orbitals in the HCNO molecule (QTAIM)

The eff-AOs as basis functions

The definition of the eff-AOs can be trivially inverted, and one gets

$$\varphi_j^A(\mathbf{r}) = \sum_{\mu=1}^{n_A} U_{j\mu}^{A*} \sqrt{\lambda_\mu^A} \chi_\mu^A(\mathbf{r}) = w_A(\mathbf{r}) \sum_{\mu=1}^{n_A} U_{j\mu}^{A*} \sqrt{\lambda_\mu^A} \chi(\mathbf{r})_\mu \quad .$$

Owing to this result, the MO-s can be written as linear combinations of the "effective AO-s" of different atoms:

$$\varphi_i(\mathbf{r}) = \sum_A^{N_{at}} w_A(\mathbf{r}) \varphi_i^A(\mathbf{r}) = \sum_A^{N_{at}} \sum_{\mu=1}^{n_A} U_{i\mu}^{A*} \sqrt{\lambda_\mu^A} \chi_\mu^A(\mathbf{r}) = \sum_A^{N_{at}} \sum_{\mu=1}^{n_A} c_{\mu i}^A \chi_\mu^A(\mathbf{r}) \quad .$$

The atomic population calculated by Mulliken analysis in the basis of the "effective AO-s" is **equal** to that obtained by the 3D QTAIM analysis.

$$Q_A^{AIM} = \int_{\Omega_A} \rho(\mathbf{r})dv = 2\int_{\Omega_A} \sum_{i=1}^{n} |\varphi_i(\mathbf{r})|^2 dv = 2\sum_{i=1}^{n} \int_{\Omega_A} |\varphi_i(\mathbf{r})|^2 dv = 2\sum_{i=1}^{n} Q_{ii}^A \ .$$

$$Q_A^{LCAO} = \sum_{\nu \in A} D_{\nu\nu} = 2\sum_{i=1}^{n}\sum_{\nu \in A} |C_{\nu i}|^2 = 2\sum_{i=1}^{n}\sum_{\mu=1}^{n_A} \left| U_{i\mu}^{A*} \sqrt{\lambda_\mu} \right|^2 = 2\sum_{i=1}^{n} Q_{ii}^A$$

Analogously, Wiberg bond orders calculated on the basis of eff-AOs are exactly equal to the Delocalization index, and so on.

Similar relationships are also found for fuzzy atomic domains such as TFVC, the only difference being the eff-AOs do not form and orthonormalized set.

Formal oxidation states from real-space analysis

- The concept of oxidation state (OS) is ubiquitous in transition metal (TM) chemistry and in the study of redox and catalytic reactions.
- Many properties such as reactivity, spin-state, spectroscopic and geometrical features of TM complexes are often rationalized on the basis of the oxidation state of the metal center.
- Formal OS are obtained by assigning integer number of electrons to the atoms/ligands according to some rules. In complicated bonding situations involving non-innocent ligands or in intermediates or transition states of catalytic reactions the formal OS assignment may be rather ambiguous.

- A revision of the concept of OS was requested to a number of experts. Pavel Karen's IUPACs Technical Report summarized their conclusions.
- New definition was provided, along with algorithms to determine the OS in molecular systems and solids.

The OS of a bonded atom equals its charge after **ionic approximation**.

P. Karen et al, Pure Appl. Chem. 2014 86 1017-1081.
P. Karen, Angew. Chem. Int. Ed. 2015, 54, 2-13

Lewis structure as a starting point

The Ionic Aproximation

- For practical applications, to the more electronegative atom, according to Allen's electronegativity scale.
- In bonds involving two atoms of the same element, divided equally.

Limitations/ambiguities of the new definition

- Comes with exceptions for the application of the IA (when the atom with higher Allen EN is a net acceptor of electrons).
- All electrons (σ or π) of a bond are necessarily assigned to the same atom after the IA.
- Different Lewis structures may lead to different OS assignation.
- EN is regarded as a genuine free atom property, hence same atoms are treated on equal footing irrespective of their chemical environment.

Thom *et al.* illustrated how neither Mulliken charges nor Mulliken spin populations match in general with the oxidation state.

Mulliken (left) and TFVC (right) charges for a set of transition metal complexes.

	Cl^-	H_2O^{HS}	H_2O^{LS}	CN^-	CO	Cl^-	H_2O^{HS}	H_2O^{LS}	CN^-	CO
V^{II}	0.98	1.12	–	0.05	0.64	1.60	1.76	–	1.60	1.64
Mn^{II}	1.10	1.24	1.18	0.10	0.64	1.36	1.64	1.70	1.53	1.53
Mn^{III}	0.93	1.58	1.52	0.35	0.80	1.46	2.05	2.09	1.58	1.67
Fe^{II}	0.86	1.22	1.15	0.01	0.51	1.27	1.63	1.77	1.46	1.44
Fe^{III}	0.99	1.64	1.48	0.24	0.66	1.44	2.06	1.98	1.49	1.58
Ni^{II}	0.99	–	1.08	-0.19	0.31	1.27	–	1.59	1.24	1.30
Zn^{II}	1.02	–	1.06	-0.03	0.52	1.25	–	1.45	1.15	1.19

ab initio determination of oxidation states

- A number or empirical approaches to *correlate* atomic distances or populations with transition metal (TM) oxidation states (bond-valence sum scheme).
- Sit et al.[1] used projection techniques to obtain d orbital populations of central metal in transition metal (TM) complexes
- Sit et al.[2] and Vidossich et al. [3] also used the positions of the centers of gravity of maximally-localized Wannier functions/localized orbitals.
- Localized Orbital Bond Analysis (LOBA) [4]: MO localization followed by population analysis
- ELIBON (electron-localizability-based oxidation number)[5], usually applied to solids.

1. Inorg. Chem. 2011, 50, 10259   2. Chem. Eur. J. 2011, 17, 12136   3. Dalton Trans., 2014, 43, 11145   4. Phys. Chem. Chem. Phys, 2009, 11, 11297   5. J. Solid State Chem. 2008, 181, 1983.

- Define atoms/fragments/ligands of the system

$$w_{\Gamma_i}(\vec{r}) \equiv \sum_{A \in \Gamma_i} w_A(\vec{r}) \qquad \mathbf{S}^{\Gamma_i} \neq \sum_{A \in \Gamma_i} \mathbf{S}^A$$

- Obtain spin resolved effOs for all Γ_i fragments
- Distribute the number of electrons among the fragments according to occupations of their effOs
- Get most appropriate atom/ligand Oxidation States and R% index.
- If necessary, the OS for the individual atoms of a given fragment can be subsequently obtained by getting the eff-AOs for every atom and distributing the number of alpha and beta electrons that were assigned to the ligand in the previous step.

EOS analysis: a very simple scheme
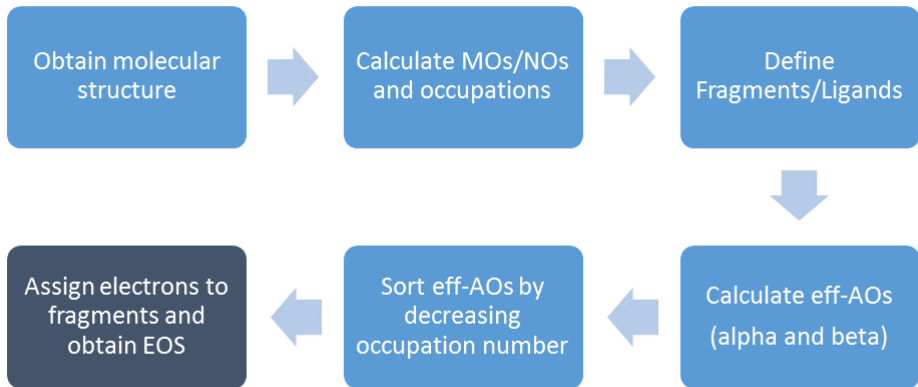
E. Ramos-Cordoba, V. Postils, P. Salvador, JCTC 2015, 11, 1501.

Alpha effective atomic orbitals for $Fe(CN)_6^{3-}$ at the B3LYP/6-31G(d) level of theory.

Total number of alpha electrons: 54

Fe		Fe		C		N	
eff-AO type	Occup	eff-AO type	Occup	eff-AO type	Occup	eff-AO type	Occup
1s	1.000	3d	0.876	1s	0.996	1s	1.000
2s	1.000	3d	0.848	2s	0.694	2s	0.997
2p	1.000	3d	0.848	2p	0.235	2p	0.728
2p	1.000	3d	0.344	2p	0.232	2p	0.703
2p	1.000	3d	0.314	2p	0.153	2p	0.701
3s	0.991	4s	0.103		0.014		0.020
3p	0.983		0.050		0.013		
3p	0.983		0.050				
3p	0.982		0.049				

Beta effective atomic orbitals for $Fe(CN)_6^{3-}$ at the B3LYP/6-31G(d) level of theory.

Total number of beta electrons: 53

Fe		Fe		C		N	
eff-AO type	Occup	eff-AO type	Occup	eff-AO type	Occup	eff-AO type	Occup
1s	1.000	3d	0.827	1s	0.996	1s	1.000
2s	1.000	3d	0.827	2s	0.708	2s	0.997
2p	1.000	3d	0.305	2p	0.246	2p	0.729
2p	1.000	3d	0.263	2p	0.239	2p	0.702
2p	1.000	3d	0.099	2p	0.153	2p	0.665
3s	0.990	4s	0.049		0.150		0.021
3p	0.982		0.048				
3p	0.982		0.048				
3p	0.982		0.032				

Oxidation states, last occupied eff-AO (LO) and first unoccupied eff-AO (FU) orbitals for $Fe(CN)_6^{3-}$.

Atom/ Fragment	O.S.	λ^{α}_{LOEFF}	λ^{α}_{HUEFF}	λ^{β}_{LOEFF}	λ^{β}_{HUEFF}
Fe	+3	0.850	0.344	0.827	0.305
C	+2	0.694	0.235	0.708	0.246
N	-3	0.701	0.020	0.665	0.021

From the "frontier" occupation numbers one can **derive** a simple index R(%) to **quantify** how close is the electron distribution provided by the actual wavefunction to the formal picture of the oxidation states.

$$R(\%) \equiv 100 \, min(1, max(0, \lambda^{\sigma}_{LO} - \lambda^{\sigma}_{HU} + 1/2), \text{ for } \sigma = \alpha, \beta.$$
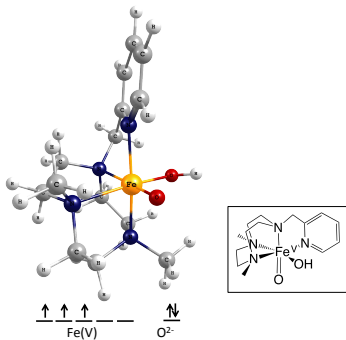
σ			λ^σ		
α	**0.934**	**0.932**	**0.930**	0.495	0.451
β	0.163	0.393	0.462	0.322	0.349
$\alpha - \beta$	0.771	0.539	0.468	0.173	0.102

d-type eff-AO occupations of the Fe atom in the [Fe(Pytacn)O(OH)]$^{2+}$ species.

High oxidation states
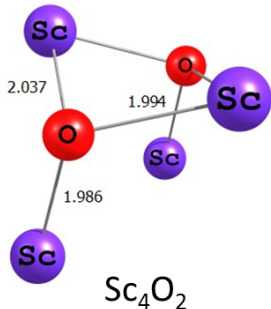
	Fragment	EOS	$\lambda_{LO}^\alpha / \lambda_{FU}^\alpha$	R(%)	N_{met}
FeO$_4$$^{2-}$	Fe	+6	0.919 / **0.452**	68.8	+1.52
	O	-2	**0.641** / 0.011		
ReO$_4$$^-$	Re	+7	0.954 / **0.295**	88.8	+2.71
	O	-2	**0.684** / 0.010		
ReH$_9$$^{2-}$	Re	+7	0.978 / **0.429**	60.4	+1.52
	H	-1	**0.533** / 0.024		
OsO$_4$	Os	+8	0.964 / **0.404**	72.9	+2.61
	O	-2	**0.633** / 0.009		
IrO$_4$$^+$	Ir	+9	0.970 / **0.509**	57.6	+2.54
	O	-2	**0.585** / 0.009		
PtO$_4$$^{2+}$	Pt	+10	0.542 / 0.197	40.6	+2.37
	O	-2	0.691 / 0.536		

$Sc_4O_2@C_{80}$ I_h

Sc_4O_2

2.037

1.994

1.986

Cage Charge	R(%)
-6	100

Atom/ Fragment	Oxidation state
Sc_{eq}	3+
Sc_{ax}	2+
O	2-

R(%) = 60%

Justification of the ionic model in EMFs

Cage	Isomer	Cluster	Spin state	EOS Cage	R(%)
C66	4059	Sc_2	singlet	-6	100
C66	4338	Sc_2	singlet	-4	88
C68	6073	Sc_2C_2	singlet	-4	92
C80	IPR 5	Sc_2C_2	singlet	-4	100
C80	ih	Sc_3C_2	doublet	-6	100
C78	d3h IPR	ScN_3	singlet	-6	100
C82	**C_2**	**U**	**triplet**	**-4**	**64**
C82	**C_{2v}**	**U**	**triplet**	**-3**	**67**

MoCp(C$_7$H$_7$), **37**

R(%) = 64

C$_5$H$_5$ $^{1-}$

Mo^{4+}

C$_7$H$_7$ $^{3-}$

Hapticity and aromaticity

Mn(C$_7$H$_7$)$_2$-1D, **54**

R(%) = 84

1$^-$ ↓↓

Mn^{2+} ↑↑↑↑↑

1$^-$ ↓↓
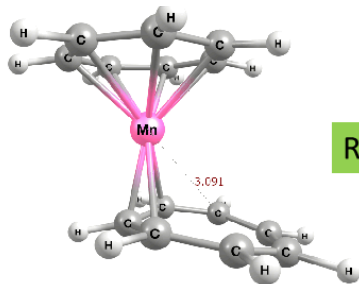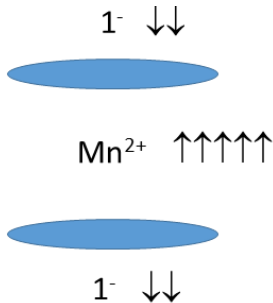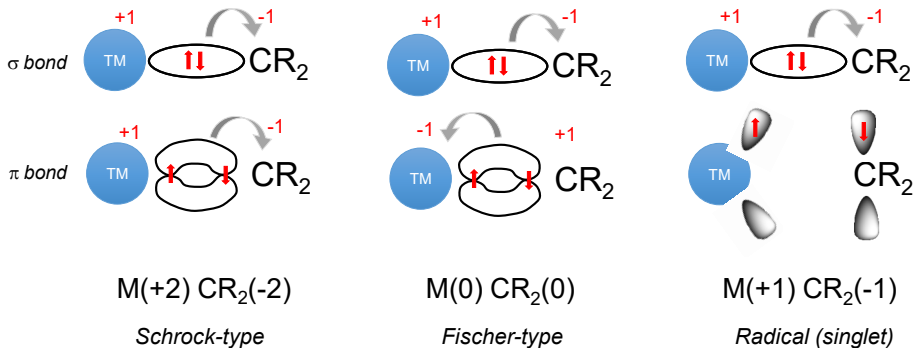
8 π electrons → **Baird aromatic**

- The formal charge on the ligands depend upon their local spin and can be rationalized in terms of Huckel and Baird aromaticities
- Partial aromaticity can be invoked for strong deviations from planarity (lower hapticities)

M(+2) CR$_2$(-2)

Schrock-type

M(0) CR$_2$(0)

Fischer-type

M(+1) CR$_2$(-1)

Radical (singlet)

- The IA can not afford for a formal neutral carbene, where the two electrons of the σ bond are assigned to the carbene and the π ones to the metal center.

2.53 Å

BO_{P-P} = 0.48
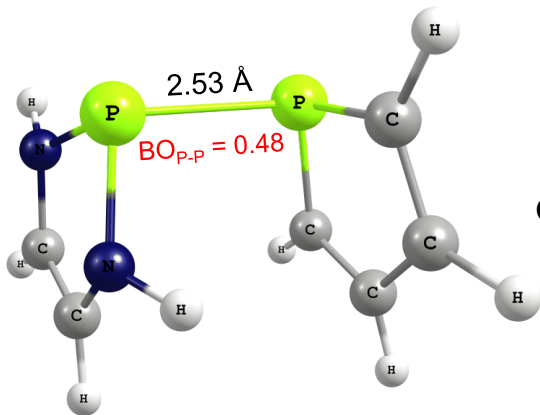
$P(NC_2H_4)-P(C_4H_4)$, **55**

EOS

$P(NC_2H_4)$ **+1**

$P(C_4H_4)$ **-1**

Global R(%) = 81.9%

	$\lambda_{LO} / \lambda_{FU}$
$P(NC_2H_4)$	0.948 / 0.308
$P(C_4H_4)$	0.627 / 0.030

Summary

- Oxidation states can be extracted from the analysis of the effective atomic/fragment orbitals and their occupancies with a strategy of general application (closed-shell systems, KS-DFT or correlated WFs)
- The EOS method is reconciled with the empirical assignation in most cases without any external guidance or exception.
- A better empirical approach to oxidation state assignation could start by identifying subsystems of enhanced stability in the molecular system in the first place, according to known rules (octet, 18e,...)