Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Electronegativity Equalization: Taming an old problem with new tools

Electronic Supplementary Information.

J. L. Casals-Sainz, E. Francisco, A. Martín Pendás*

September 29, 2020

Contents

1 Basic Chemistry in Real Space

The real space point of view in theoretical and computational chemistry tries to build proper quantum mechanical observables from orbital invariant descriptors with chemical meaning. Among them, all reduced densities and density matrices (RDs, RDMs). An identification of spatial regions with chemical concepts is also necessary. This can be done through spatial partitionings, normally induced by the topology of a scalar field. For instance, the topology of the electron density, ρ, induces an atomic partitioning known as the Quantum Theory of Atoms in Molecules (QTAIM). It was introduced and explored by R. F. W. Bader and coworkers.[1] Other fields, like the electron localization function of Becke and Edgecombe's,[2] provides a partition into cores, lone pairs and bonding domains, etc.

From atoms (or electron-pair domains), chemical bonding descriptors are built. Both the electron-counting perspective (leading to populations and bond orders) as well as the energetic view that provides bond strengths are needed. These are offered by, for instance, electron distribution functions (EDFs) and the interacting quantum atoms approach (IQA). In order to study Electronegativity Equalization (ENE) in molecular systems, both are needed to construct derivatives of energetic descriptors with respect to electron counts. We will thus provide a basic account of the theory of open systems in real space, IQA and EDFs.

1.1 Real space regions as open quantum systems

Let us consider two systems S^a, S^b with Hilbert spaces H^a, H^b, respectively. The composite system S lives in the tensor product spate $H^a \otimes H^b$. If $\{|\phi_i^a\rangle\}$ and $\{|\phi_j^b\rangle\}$ are orthonormal bases in H^a, H^b (countable bases have been assumed, but the results are general), then a state in S can be written as $\Psi = \sum_{ij} a_{ij} |\phi_i^a\rangle \otimes |\phi_j^b\rangle$. The subsystems are uncorrelated if $\hat{\rho} = \hat{\rho}^a \otimes \hat{\rho}^b$, where $\hat{\rho}$ is the density operator. Otherwise they are said to be entangled. In the complete system we may still be interested in the expectation value of an operator A^a that depends only on dynamic variables of subystem S^a. To obtain it, it is only necessary to known the so-called reduced density operator of the subsystem, defined by taking the partial trace of $\hat{\rho}$ over subsystem S^b: $\langle A \rangle = \mathrm{Tr}\ A\hat{\rho} = \mathrm{Tr} A\hat{\rho}^a$, where $\hat{\rho}^a = \mathrm{Tr}_b \hat{\rho}$.

We have shown[3] how to perform a partial trace in real space. For an N-electrons system in a pure state, $\hat{\rho} = \Psi^\star(\boldsymbol{x}'_1 \ldots, \boldsymbol{x}'_N)\Psi(\boldsymbol{x}_1 \ldots, \boldsymbol{x}_N)$. Con-

sidering a spatial region A and its complement, $\bar{A} = B$, $A \cup B = R^3$, we introduce the indicator function of a domain Ω, ω_Ω such that $\omega_\Omega(\boldsymbol{x}) = 0$ and $\omega_\Omega(\boldsymbol{x}) = 1$ for $\boldsymbol{x} \notin \Omega$ and $\boldsymbol{x} \in \Omega$. Equivalent definitions hold for the primed variables.

An N-electron spatial projection operator is obviously $1^N = \prod_{i=1}^{N}[\omega_A(\boldsymbol{x}_i)+ \omega_B(\boldsymbol{x}_i)]$. Applying this to the $\boldsymbol{x}$ and $\boldsymbol{x}'$ coordinates in the the $\hat{\rho}$ operator above, the full density operator becomes a sum of 2^{2N} terms in which the primed and unprimed electrons are separated into the A and B spatial domains. The reduced density operator of domain A, $\hat{\rho}^A$, is obtained from $\hat{\rho}$ by inegrating over B. After doing it, only 2^N terms survive, each corresponding to a given number of α and β electrons in domain A, what is called a spin sector. If spin is also summed over, we talk about spinless sectors. Using electron indistinguishability, the 2^N terms can be classified into $N+1$ spinless sectors, each contaning a different number of electrons in A, irrespectively of their spin:

$$\hat{\rho}^A = \bigoplus_{n=0}^{N} \hat{\rho}^A_n, \tag{1}$$

where $\hat{\rho}^A_0 = \int_B \Psi^\star(\boldsymbol{x}_1 \dots \boldsymbol{x}_N)\Psi(\boldsymbol{x}_1 \dots \boldsymbol{x}_N)d\boldsymbol{x}_1 \dots d\boldsymbol{x}_N$ and, for $n \geq 1$

$$\hat{\rho}^A_n(\boldsymbol{x}_{i\leq n}; \boldsymbol{x}'_{i\leq n}) = \prod_{i=1}^{n} \omega_A(\boldsymbol{x}'_i)\omega_A(\boldsymbol{x}_i) \times$$
$$\times \binom{N}{n} \int_B \Psi^\star(\boldsymbol{x}'_{i\leq n}, \boldsymbol{x}_{i>N})\Psi(\boldsymbol{x}_{i\leq n}, \boldsymbol{x}_{i>n} d\boldsymbol{x}_{i>n}, \tag{2}$$

with $\boldsymbol{x}_{i\leq n} = \boldsymbol{x}_1 \dots \boldsymbol{x}_n$ and $\boldsymbol{x}_{i>n} = \boldsymbol{x}_{n+1} \dots \boldsymbol{x}_N$, for instance. The trace of each sector density operator is equal to the probability that a given number of electrons reside in the spatial region, see below. To each of the sectors we can associate reduced densities of all orders up to the total number of electrons of the sector.

1.2 Energy partitioning: Interacting Quantum Atoms

Given an atomic spatial partition, the interacting quantum atoms (IQA) decomposition considers the one- and two-domain division of the non-relativistic Born-Oppenheimer electronic energy[4] described in the following equation,

$$E = \sum_A E^A_{self} + \sum_{A>B} E^{AB}_{int}$$
$$= \sum_A \left(T^A + V^{AA}_{ne} + V^{AA}_{ee}\right) + \sum_{A>B} \left(V^{AB}_{nn} + V^{AB}_{ne} + V^{BA}_{ne} + V^{AB}_{ee}\right), \quad (3)$$

wherein E^A_{self} and E^{AB}_{int} are the IQA self and interaction energies of domain A and pair AB, while T^A denotes the kinetic energy of domain A. Finally, the terms V^{AB}_{ne} and V^{AB}_{ee} stand for (i) the attraction between the nuclei of domain A and the electrons of domain B and (ii) the repulsion between the electrons in domain A with those in region B, respectively. The atomic or group self-energies are the expectation values of the atomic (or group) Hamiltonian in-the-molecule.

We can get further insight about the nature of the interaction between two atoms by separating the electronic repulsion into its Coulombic and exchange-correlation components. This splitting allows for a further separation of the IQA interaction energy of a pair AB into ionic and covalent contributions as[4]

$$E^{AB}_{int} = V^{AB}_{cl} + V^{AB}_{xc} = E^{AB}_{ion} + E^{AB}_{cov}. \quad (4)$$

Usually, binding is measured relative to appropriate references for the quantum fragments A, with $E^{A,0}$. Then $E^A_{self} - E^{A,0} = E^A_{def}$ is called the atomic or fragment deformation energy, which corresponds to a combination of the traditional promotion energy and other effects, like spin-recoupling, true electronic deformation, etc.[5] We have shown that the IQA interaction energies behave as *in situ* bond energies. IQA thus provides an invariant decomposition of the energy into group deformations and bond contributions in which covalent and ionic energies acquire rather pure forms.

1.3 Electron-counting: Electron Distribution Functions (EDFs)

Electron counting provides access to the more qualitative view of chemical bonding in which the number of electrons engaged in sharing or in pure transfer between atoms gives rise to bonding descriptors like bond orders. In real space, we examine how the total number of electrons distributes among the different atomic regions in which we divide the space.

EDFs are defined as follow. Given an N–electron molecule and an exhaustive partition of the real space ($\mathcal{R}^3$) into m arbitrary regions Ω_1, Ω_2, ..., Ω_m ($\Omega_1 \cup \Omega_2 \cup \cdots \cup \Omega_m = \mathcal{R}^3$), an EDF is the set of all the probabilities $p(n_1, n_2, \ldots, n_m)$ of finding exactly n_1 electrons in Ω_1, n_2 electrons in Ω_2, ..., and n_m electrons in Ω_m, $\{n_p\}$ being integers ($n_i \in \mathcal{N}$) satisfying $n_1+n_2+\cdots+n_m = N$. This view is in accord with considering subsystems as open quantum systems in which number operators do not commute with the subsystem Hamiltonian. In this way, Ψ is not an eigenstate of the operator defining the number of electrons in domain Ω_i, $\hat{N}_{\Omega_i}$. This means that the average number of electrons in Ω_i is not an eigenvalue of $\hat{N}_{\Omega_i}$, so that measuring the number of electrons in the domain will render values n_{Ω_i} ranging from 0 to N, the total number of electrons, with a defined set of probabilities, $p(n_{\Omega_1})$. This is the one-fragment EDF for domain Ω_i, and, in the general case, we are interested in the multivariate probabilities $p(n_1, n_2, \cdots, n_m)$. To obtain them one needs $\Psi(1, \ldots, N)$, Ψ being the complete wave function,

$$p(n_1, n_2, \ldots, n_m) = N!\Lambda \int_D \Psi^\star \Psi d\mathbf{x}_1 \cdots d\mathbf{x}_N, \tag{5}$$

where D is a multidimensional domain in which the first n_1 electrons are integrated over Ω_1, the second n_2 electrons over Ω_2, $\cdots$, and the last n_m electrons over Ω_m, and $N!\Lambda = N!/(n_1!n_2!\cdots n_m!)$ is a combinatorial factor that accounts for electron indistinguishability. The 3D domains of these integrations can be arbitrary, but when using QTAIM atomic basins, a partition of the N electrons of the molecule that assigns a given number of electrons (including possibly 0) to each of these regions will be called a *real space resonance structure* (RSRS)[6] and there are $N_S = (N+m-1)!/[N!(m-1)!]$ of these for a given N,m pair. With the notation $S(n_1, n_2, \ldots, n_m) \equiv S(\{n_p\})$, or simply $(n_1, n_2, \ldots, n_m) \equiv \{n_p\}$, we label the resonance structure having n_1 electrons in Ω_1, n_2 electrons in Ω_2, ..., and n_m electrons in Ω_m. If electrons are spin-segregated, then we come to spin-resolved EDFs, and a set of probabilities $p(n_1^\alpha, n_1^\beta, n_2^\alpha, n_2^\beta, \ldots, n_m^\alpha, n_m^\beta)$ which gives extremely fine-grained information about how electrons and their spins distribute.[7]

The computation of $p(n_1, n_2, \ldots, n_m)$ for all the RSRSs provides all the statistical moments of the electron populations, including the average number of electrons in a given region, or its fluctuation. The average population is obviously given by

$$N_i = \langle n_i \rangle = \sum_{\{n_p\}} n \times p(\{n_p\}) = \sum_{n=0}^{N} n \times p(n). \tag{6}$$

where $p(n)$ is the probability of having n electrons in Ω_i, which is obtained by adding the $p(\{n_p\})$'s for all possible values of n_1,...,n_{i-1},n_{i+1},..., and n_m such that $n_1 + \cdots + n_{i-1} + n_{i+1} + \cdots + n_m = N - n$.
It is not difficult to show that the number of shared pairs between two regions may be obtained directly by counting the number of intra- and interpairs.[8] This has given rise to the so-called localization and delocalization indices, $(\lambda^{ii}, \delta^{ij})$, which determine the number of *localized* and *delocalized* pairs. The latter, which is the covalent bond-order in real space can be obtained from the $p(\{n_p\})$ probabilities as

$$\delta^{ij} = -2\mathrm{cov}(i,j) = -2\left[\langle n_i n_j \rangle - \langle n_i \rangle \langle n_j \rangle\right] = \tag{7}$$

$$-2\left[\sum_{\{n_p\}} n_i n_j \times p(\{n_p\}) - \langle n_i \rangle \langle n_j \rangle\right] = \tag{8}$$

$$-2\sum_{n_i n_j} (n_i - N_i)(n_j - N_j) p(n_i, n_j) = 2N_{ij} \tag{9}$$

where the -2 factor has been included to comply with the usual definition of δ in terms of the exchange-correlation density and to ensure that the bond order for an ideal single bond is equal to 1,

$$\delta^{ij} = -2\int_{\Omega_i}\int_{\Omega_j} d\boldsymbol{x}_1 \; d\boldsymbol{x}_2 \rho_{xc}(1,2). \tag{10}$$

The localization index is given by

$$\lambda^{ii} = N_i - \mathrm{cov}(i,i) = N_i - \mathrm{var}(i) = N_i - \sum_{n_i} (n_i - N_i)^2 p(n_i) = N_{ii} \tag{11}$$

From equations 7-11 it is clear that $N_{ii} = N_i$ if the variance is zero and that $N_{ij} = 0$ if the covariance is cero. This is the starting point for a complete theory of chemical bonding based on the fluctuation of electron populations. There is chemical bonding between two regions if their electron populations

are not statistically independent. A sum rule, that classifies electrons into localized and delocalized sets appears:

$$N = \sum_{\Omega_i} N_i = \sum_{\Omega_i} \lambda^{ii} + \frac{1}{2} \sum_{\Omega_i \neq \Omega_j} \delta^{ij}. \tag{12}$$

Suitable generalizations in the case of multi-center bonding exist.[9]

It is easy to build energetic models using the open systems perspective and the IQA+EDF approaches.

1.4 The simple two-center, two-electron case

In the case of a system with two domains (A-left, B-right) and two chemically active electrons, the 2c-2e bond, we can easily map all possible resonance structures in real space. There are only three of them:$(2,0)$, $(1,1)$, $(0,2)$, where we label how many electrons lie in each of the A, B domains. The EDF space is two-dimensional, since $p(2,0) + p(1,1) + p(0,2) = 1$, and all bond indices become fully mapped in this 2D space. A convenient coordinate system can be built with the mean-field probability that any of the electrons lie in the left basin (for instance). We call this probability p and it provides a measure of heteropolarity. The second coordinate is a correlation factor $-1 \leq f \leq 1$ that determines how the electronic motion is correlated. $f = 1$ means that an electron is completely excluded from one domain if the other is already in it (positive correlation) and $f = -1$ implies that the two electrons are always found together within the same domain (negative correlation). The correlation factor here defined plays the same role as that used in density matrix theory, where $\rho^2(r_1, r_2) = \rho(r_1)\rho(r_2)(1-f)$. The (p, f) pair describes fully a 2c,2e link at this level:[10]

$$\begin{aligned} p(2,0) &= p^2 - p(1-p)f, \\ p(1,1) &= 2p(1-p)(1+f), \\ p(0,2) &= (1-p)^2 - p(1-p)f. \end{aligned} \tag{13}$$

It is clear that the average number of electrons in B is $N_B = 2 \times p(0,2) + 1 \times p(1,1) + 0 \times p(2,0) = 2(1-p)$. The charge in B is thus $Q_B = 2p - 1$. Homonuclear links have necessarily $p = 1/2$.

If we use these p, f parameters, the covalent bond order can be immediately obtained from $\delta = -2\mathrm{cov}(N_A, N_B)$ and becomes $\delta = 4p(1-p)(1-f)$.

An ionic bond order $\iota = -Q_a Q_b$ where Q is the net charge of a center has also been defined.[11] In standard weakly correlated bonds with positive $f \sim 0$, the EDF is close to binomial, and δ peaks at $\delta = 1$ for a purely covalent homopolar link with $p = 1/2$. As electron correlation, f, or polarity, p, increases δ decreases. Moreover, for non-correlated $f = 0$ links, $\iota = 1 - \delta$ so, in agreement with standard wisdom, the ionic and covalent bond orders are inversely correlated.

When electron correlation is important, f deviates considerably from zero, and the model describes positively or negatively correlated bonds. The latter case implies a bosonization of the link. Electrons try to delocalize together, giving rise to very large fluctuations. The most extreme 2c,2e case with $\delta = 2$ occurs when $p(0,2) = p(2,0) = 1/2$ and $p(1,1) = 0$. The real space bond orders have been shown to be electron count analogs of energetic quantities.[11] Under the IQA perspective a multipolar expansions shows that the first order ionic and covalent energies are immediately related to their corresponding bond orders. For an interaction between atoms A and B,

$$E_{ion}^{AB} \sim -\frac{\iota^{AB}}{R_{ij}} \quad E_{cov}^{AB} \sim -\frac{1}{2}\frac{\delta^{AB}}{R_{AB}}. \tag{14}$$

2 The Electronegativity Equalization (ENE) models

We will describe here the one- (1e) and two- (2e) electron models used in the manuscript. The second is immediately related to the 2c,2e bond that we have just described. In the following, A, B or left, right labels will be used indistinctly.

2.1 The 1e model

As shown in the main text, imposing a constrained charge in a real space atom in the large distance regime may be understood in terms of a one-way, one electron transfer. This is energetically indistinguishable from the behavior of a grand canonical atom with an average electron population equal to the constrained charge.

Let us then consider a 2c,2e system (like the H_2 molecule) at large internuclear distance for which we impose a constrained atomic charge $Q =$

$Q_A = -Q_B = -q$. At positive q values the A atom gains electrons ($N_A > 1$) while the B atom is depopulated ($N_B < 1$). At $q = 0$ we know that only one resonance structure is populated: $p(2,0) = p(0,2) = 0$, $p(1,1) = 1$. In the 1e model, electron transfer from B to A occurs only one-way, from right to left, maintaining $p(0,2) = 0$ during the process. This implies that a single coordinate (q for instance) describes the change, and that the EDF is $p(2,0) = q$, $p(1,1) = 1-q$, $p(0,2) = 0$. Notice that once the EDF is given, all descriptors can be computed: $N_B = 1 - q$ and $\delta = 2q(1-q)$ peaks at $q = 1/2$, $\delta = 1/2$.

In this long-range model it is also trivial to write analytical expressions for the atomic self-energies and the total molecular energy. We must simply take into account that since no short-range deformations exist, the sector density matrices are those of isolated systems with the appropriate number of electrons. In other words, the self-energy of the A, A^-, and A^+ species are the *in vacuo* energies of the neutral, anionic, and cationic moieties, respectively. The same can be said for B. This said,

$$\begin{aligned} E^A_{self} &= p(2,0) \times E(A^-) + p(1,1) \times E(A) + p(0,2) \times E(A^+), \\ E^B_{self} &= p(2,0) \times E(B^+) + p(1,1) \times E(B) + p(0,2) \times E(B^-), \\ E^{AB}_{int} &= p(2,0) \times Q_A Q_B / R_{AB} + p(1,1) \times 0 + p(0,2) \times Q_A Q_B / R_{AB}. \\ E &= E^A_{self} + E^B_{self} + E^{AB}_{int} \end{aligned} \tag{15}$$

Particularizing for the dihydrogen molecule, we get

$$\begin{aligned} E^A_{self} &= qE(\mathrm{H}^-) + (1-q)E(\mathrm{H}), \\ E^B_{self} &= 0 + (1-q)E(\mathrm{H}), \\ E^{AB}_{int} &= -q/R_{AB}, \\ E(Q) &= 2E(\mathrm{H}) + \eta|Q| - |Q|/R_{AB}. \end{aligned} \tag{16}$$

as described in the text. Notice that the behavior of each of the self-energies is exactly grand canonical, and that the energy model is that of Perdew and coworkers.[12] No other approximation has been made except that of long-range, so that the model is exact at dissociation. Interestingly, at constant internuclear distance the interaction energy is linear with the constrained charge, as shown in Fig. 1 in the text.

2.2 The 2e model

At variance with the model just examined, in which electron flow is accompanied by the change of only two out of the three possible resonance

structure probabilities of a 2c,2e interaction (being thus a one-parameter or 1e approximation), the three probablities can change simultaneously in general situations. This implies the existence of two independent parameters. For a homopolar case, e.g the H_2 molecule, at the stationary $Q = 0$ value $N_A = 1, N_B = 1$ always, and the electron distribution needs be symmetric: $p(0,2) = p(2,0)$. Following Eq. 13, the mean-field probability p as well as the correlation factor f suffice to offer a complete description from the electron counting point of view.

To simplify as much as possible, we can assume that not far from equilibrium geometries the correlation factor is sufficiently small so as to ignore it. With this new approximation, the 2e model depends again on only one parameter p. Notice that the distributions with $f = 0$ are binomial:

$$\begin{aligned} p(2,0) &= p^2, \\ p(1,1) &= 2p(1-p), \\ p(0,2) &= (1-p)^2. \end{aligned} \tag{17}$$

Now we use the sector reduced density matrices to write the energy as a weighted sum of resonance structure energies:

$$E = p(2,0) \times E(2,0) + p(1,1) \times p(1,1) + p(0,2) \times p(2,0). \tag{18}$$

Each of these resonance structure energies can be IQA partitioned,

$$E(n_A, n_B) = E^A_{self}(n_A, n_B) + E^B_{self}(n_A, n_B) + E^{AB}_{int}(n_A, n_B). \tag{19}$$

Notice that, in the general case, when a large number of resonance structures contribute, the list of atomic self-energies to include in the calculations can be extense. In the present toy example, the (2,0) and (0,2) structures include in-the-molecule H^+ cations and H^- anions, while the (1,1) possess two equivalent neutral in-the-molecule H atoms. Their self-energies are called E_+, E_-, and E_0, respectively. E_-, for instance, is the self-energy of the left atom in the (2,0) resonance structure. This is the self-energy of a "hydride" in a structure where, at a given distance, a domain with no electron but one proton is found. Similarly, E_0 is the self-energy of a domain with just one electron that has an equivalent neighboring region by its side.

The average self-energies of atoms A, B will thus be:

$$\begin{aligned} E^A_{self} &= p^2 E_- + 2p(1-p)E_0 + (1-p)^2 E^+, \\ E^B_{self} &= p^2 E_+ + 2p(1-p)E_0 + (1-p)^2 E_-. \end{aligned} \tag{20}$$

We can transform these results by writing them as functions of the constrained charge. Taking into account that $p = (1 - Q_A)/2 = (1 + Q_B)/2$,

$$E^A_{self}(Q_A) = E^A_{self}(0) + \chi^{self}_A Q_A + \eta^{self}_A Q^2_A/2, \tag{21}$$

with an equivalent relation for B. In this expression, $E^A_{self}(0)$ is the A self-energy at $Q = 0$, equal to $(E_+ + E_- + 2E_0)/4$, and χ^{self}_A and η^{self}_A the in-the-molecule electronegativity and hardness of A:

$$\begin{aligned} \chi^{self}_A &= (E_+ - E_-)/2 = (I^{self} + A^{self})/2, \\ \eta^{self}_A &= (E_+ + E_- - 2E_0)/2 = (I^{self} - A^{self})/2, \end{aligned} \tag{22}$$

where $I^{self} = E_+ - E_0$, $A^{self} = E_0 - E_-$ are the in-the-molecule atomic ionization potential and electron affinity, respectively. Differently from the 1e model valid at dissociation, now the atomic self-energies are differentiable at integer N from both sides:

$$\boxed{\textbf{At } \boldsymbol{Q = 0},\ \left(\frac{\partial \boldsymbol{E^A_{self}}}{\partial \boldsymbol{Q_A}}\right)_{\boldsymbol{v}} = \boldsymbol{\chi^{self}_A}}$$

and, of course, the atomic electronegativities of both atoms are equal. This is a very satisfying result. If we now further suppose that R_{AB} is large enough so that the atomic distribution is not far from spherical, then the interaction energy can be exactly computed from Gauss' theorem as purely Coulombic, so that a $-1/R_{AB}$ term will exist between the ions in the (0,2) and (2,0) structures. E_{int} will vanish for the neutral (1,1) distribution. With this,

$$E_{int} = -[p^2 + (1 - p)^2]/R_{AB} = -(1 + Q^2)/(2R_{AB}), \tag{23}$$

which is parabollic in Q and does not contribute to Q derivatives at $Q = 0$. For this 2e system, the open systems perspective transforms covalency into an electrostatic-like term. At $Q = 0$, the interaction energy between both atoms is $-1/(2R_{AB})$. This is nothing but the classic Coulombic attraction between a pure cation and a pure anion that are found with probabilities 1/4 in the (2,0) structure, and 1/4 in the (0,2) one, respectively, adding to a total ionic probability of 1/2.

Using $Q = Q_A = -Q_B$, and summing up,

$$E(Q) = 2E_{self}(0) + (\chi^{self}_A - \chi^{self}_B)Q + \frac{(\eta^{self}_A + \eta^{self}_B)}{2}Q^2 - \frac{(1 + Q^2)}{2R_{AB}} \tag{24}$$

2.3 The 1e to 2e transition

The open systems point of view allows for an easy rationalization of the 1e to 2e transition as the interatomic distance decreases. This is dependent on the constrained charge, and straightforward at $Q = 0$. Comparing Eq. 16 and Eq. 24, $E^1(0) = 2E(\mathrm{H})$, and $E^2(0) = 2E_{self}(0) - 1/(2R_{AB})$. Given that the in-the-molecule state is a deformed one, $E_{self} > E(\mathrm{H})$, so that if the deformation energy $E_{def} = E_{self} - E_{isolated}$ of each atom in a homodiatomic molecule is used, the two separated atoms will be more stable than a covalently held model at distances larger than $E_{AB,t} = 1/(4E_{def})$

When two different species A and B interact, a complete charge transfer may also occur within the 1e regime from a neutral to an ionic situation. This occurs when the energy cost needed to form an ion pair equals the mutual Coulombic attraction of the ions: $(I - A) = 1/R_{AB}$. This is Perdew et al. insight.[12] Notice that there are obviously two paths to form an ion pair, so that the observed neutral to ionic transition will be that in which the $(I - A)$ cost is smaller. According to this view, the formation of a covalent 2c,2e link can be understood as a resonant symmetric autoionization.

3 ENE in heterodiatomics: the LiH molecule

Insight into how ENE using open systems occurs is now shown in another toy example: The LiH molecule. We have thus performed FCI constrained calculations. Results are shown at two internuclear distances, $R = 5.0$ and $R_e = 3.02$ au. See the computational details section for more information. Fig. 1 gathers our main results. Some convergence problems where found when trying to obtain Li^- species, so that reliable results are only shown up to $Q(\mathrm{Li}) \approx -0.9$. At $R = 5$ au, the avoided crossing between the covalent and ionic states has almost been completed, and with no charge constraint the topological charge of Li is $Q(\mathrm{Li}) = 0.825$ au. This increases to $Q(\mathrm{Li}) = 0.909$ at equilibrium.

As it is clear from the data, at both distances we observe a behavior that can be classified as doubly 1e-like. Although the self-energy curves are differentiable, they show well-developed piecewise-linear segments that agree with a one-electron transfer processes. The goodness of the one-electron picture in LiH was pointed out years ago.[13]

Now the behavior of the two atomic species is clearly different. As the H

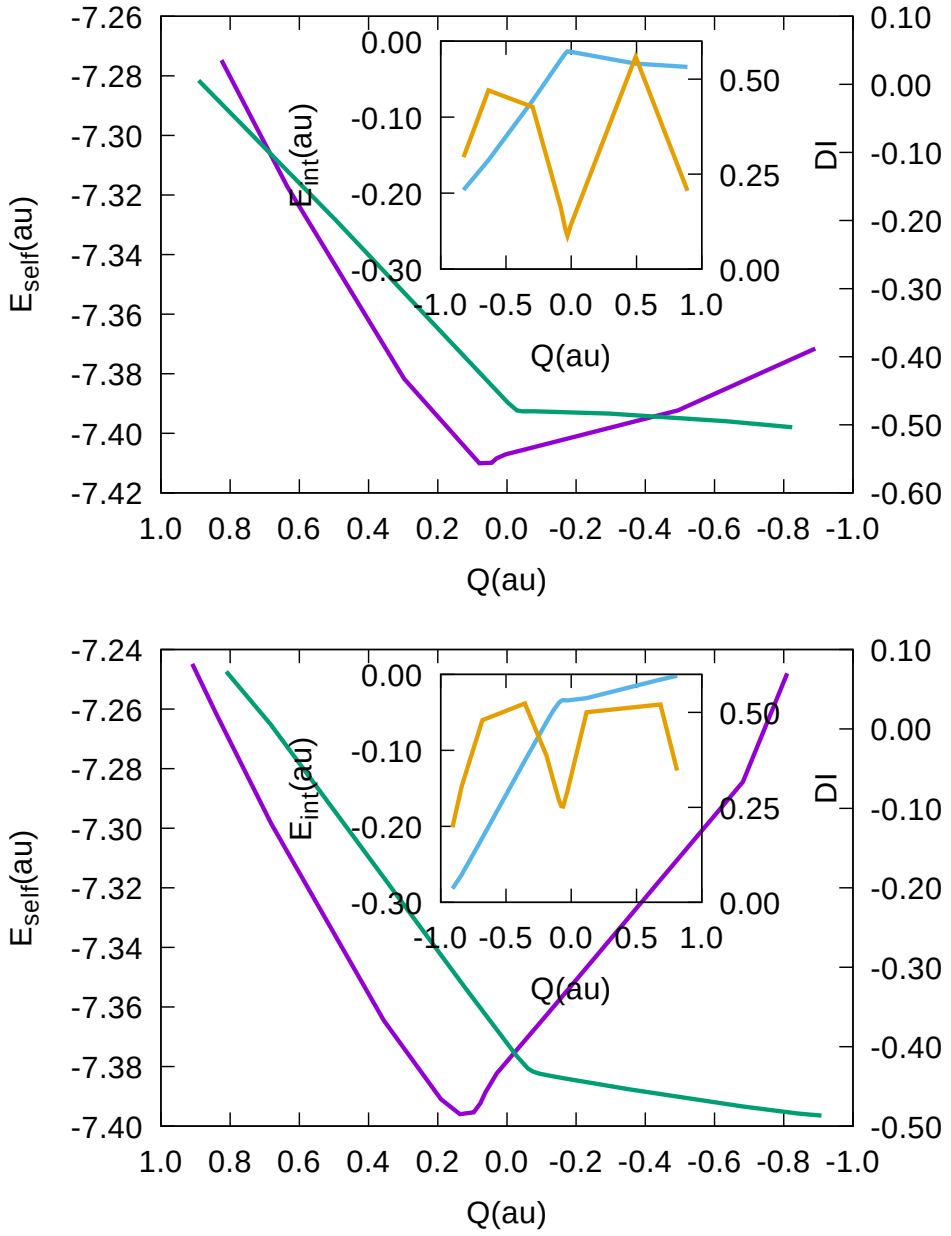

Figure 1: Evolution of E^{H}_{self} (green, right scale), and E^{Li}_{self} (blue, left scale), together with E^{AB}_{int} (cyan) and the delocalization index DI (orange) in the inset, as the CgC Q varies in LiH at $R_{AB} = 5.0$ (top) and 3.02 au (bottom), respectively. All data in au.

atom is concerned, we can see at both distances the different slope of the self-energy curves in the cationic (or normal) branch, in which Li gets positively charged and in its anionic (or inverted) counterpart. At $R = 5$ au, the H atom is well described as a grand canonical entity, with ionization potential and electron affinity matching rather well the free atomic value. Extracting electrons from the Li atom is relatively easy (notice the range difference between the right and left energy scales). However, adding electrons to the Li in-the-molecule does lead to a rather linear, although endothermic, self-energy variation, which is clearly more noticeable at the equilibrium geometry. Li has a low electron affinity (about 0.62 eV 10.1103/PhysRevA.53.4127), and our calculations show that it soon becomes an unstable entity in the field of its neighboring cation.

The doubly 1e character of the charging process is clearly evidenced when the behavior of the DI is examined. At both distances, the DI displays one peak with DI$\approx$ 0.5 when $|Q| \approx 0.5$ in each of the charging branches. This clearly means that we can apply a reasoning equivalent to that leading to Eq. 16 for each branch. To a good approximation, charging the quantum atoms in any of the two possible senses can be envisioned as a one-parameter process in which one of the three possible resonance structures remains unoccupied. Considering a frozen Li core, $p(2,0) \approx 0$ while we traverse the LiH to Li^+H^- branch, and $p(0.2) \approx 0$ when running over the LiH to LI^- H^+ branch. These two one-electron branches imply two parabollic $2p(1-p)$ DI regimes. Due to the considerable difference between the ionization potentials of Li and H, adding the atomic self-energies of the two atoms involved in each branch leads to considerably smaller s self-energies if we follow the Li^+H^- charging branch. Thus, the energy of the model can be written in terms of Q(Li) as $E(Q) \approx E_{self}(\mathrm{LiH}) + (I_{\mathrm{Li}} - A_{\mathrm{H}})Q - Q/R_{AB}$ which obviously leads to complete ionization. An exam of E_{int} shows that the normal branch shows the expected linear $-Q/R_{AB}$ behavior at fixed internuclear distance, but the inverted branch has a more complex behavior due to the anomalous structure of the Li^-H^+ moiety.

4 Computational implementation

We have implemented general atomic charge-constrained calculations in the PySCF suite.[14] To that end we have followed the ideas contained in the original Ciowsloski and Stefanov implementation.[15] We thus consider a diatomic

system (the generalization to several atoms is straightforward) and minimize its energy subjected to the fixed atomic charge constraint. For an AB system in which we constrain the number of electron for atom A, N_A, we thus minimize the Lagrange functional:

$$\hat{H} - \lambda \hat{N}_A, \tag{25}$$

where $\hat{N}_A$ is the operator whose expectation value provides the atomic population of atom A. Using topological atoms, this can be written as

$$\hat{N}_A = \int_A \delta(\boldsymbol{r} - \boldsymbol{r}') d\boldsymbol{r}', \tag{26}$$

so that its matrix elements in a given basis $\{i\chi_\mu\}$ are given by

$$N_{A,\mu\nu} = \int_A \chi^*_\nu(\boldsymbol{r}) \chi_\mu(\boldsymbol{r}) d\boldsymbol{r}, \tag{27}$$

i.e. the standard atomic overlap matrix (AOM) commonly used in the quantum theory of atoms in molecules. Notice that obtaining the AOMs implies choosing a given atomic topological partition for the system under study. We have used that of the unconstrained calculation throughout this work. Once this has been obtained at a selected level of theory, the core Hamiltonian is supplemented by the matrix elements of the $\lambda \hat{N}_A$ term. Calculations are then performed at a grid of values of the λ parameter. Since $\partial E / \partial Q_A = \lambda$ and the calculations are performed at constant external potential, λ gives the difference in electronegativity between the two atomic components. At $\lambda = 0$ the ENE follows directly. We have taken profit of the high modularity of the PySCF code, see below. At the time of writing, both SCF, MCSCF, Full CI, CASSCF, and CASCI implementations are available.

In a practical calculation, it is easier to impose λ and then read the constrained atomic charges from the calculation. This is what we have done.

The workflow for a constrained calculation for a given molecule at a selected level of theory at a fixed geometry and predefined basis set is as follows:

1. A first standard (unconstrained) calculation is performed: a .wfn or .wfx file is generated.

2. A QTAIM calculation is done so that atomic basins are identified and the AOM over all (occupied and virtual) orbitals is computed. This

AOM can be obtained through several different codes. We have used our in-home PROMOLDEN package,[16], although it is easy to (so-to-speak) cheat AIMALL[17] to perform the same task.

3. A back transformation obtains the AOM matrix over primitive functions.

4. A λ value is chosen and the *constrained* PySCF module is invoked to minimize the Lagrangian. This module reads the AOM matrix. A .wfn file with the final wavefunction is also generated.

5. Postprocessing of the wavefunction is performed, including its IQA or EDF analysis to provide self-energies, interaction energies, and probabilities of the relevant resonance structures. In this step the interatomic surfaces are kept constant at the unconstrained ones. This is easily done through the PROMOLDEN code.

The main PySCF constrained module is now described over a simple example.

1. PySCF gto.Mole specification

```
#!/usr/bin/env python

import numpy, h5py, os, sys
from pyscf import gto, scf, dft, lib, mcscf
from pyscf.tools import wfn_format

subname = 'h2'
name = 'h2_ct_0p35'
atm = 0
mult = 0.35

mol = gto.Mole()
mol.atom = '''
H      0.000000      0.000000      0.000000
H      0.000000      0.000000   2.645886245
'''
mol.basis = 'aug-cc-pvdz'
mol.verbose = 4
mol.spin = 0
mol.symmetry = 0
mol.charge = 0
mol.build()
```

A calculation on H_2 with $\lambda = 0.35$ au. λ is labelled as mult. The basis set is aug-cc-pVDZ, and a standard PySCF mol.build call is used.

2. Read & build AOM

```
# Read overlap matrix and transform to AO basis
nao = mol.nao_nr()
saom = numpy.zeros((nao,nao))
mol = lib.chkfile.load_mol(subname+'.chk')
mo_coeff = scf.chkfile.load(subname+'.chk', 'scf/mo_coeff')
coeff = numpy.linalg.inv(mo_coeff)
with h5py.File(subname+'.chk.h5') as f:
     idx = 'ovlp'+str(atm)
     saom = f[idx+'/aom'].value
     saom = coeff.T.dot(saom).dot(coeff)
```

The atomic overlap matrix over molecular orbitals is read from the chekpoint file obtained in an unconstrained calculation. The inverse of the matrix of MO coefficients is obtained and used to back-transform the AOM to the primitive basis. It is stored as saom

3. Define constraint

```
def get_pop(s, mult):
    nao = s.shape[1]
    fock = numpy.zeros((nao,nao))
    fock = mult*s
    return fock
```

A function to compute the λ part of the Fockian is constructed.

4. Build core Hamiltonian

```
# Calc
lig = get_pop(saom,mult)
hcore = mol.intor('int1e_kin') + \
        mol.intor('int1e_nuc') - lig
mf = scf.RHF(mol).newton()
mf.conv_tol = 1e-6
mf.max_cycle = 120
mf.get_hcore = lambda *args: hcore
mf.kernel()
```

The core Hamiltonian is redefined including the constraint ("lig").

5. CI

```
nelecas = 2
ncas = mf.mo_coeff.shape[1]
mc = mcscf.CASCI(mf, ncas, nelecas)
emc = mc.kernel()[0]

nmo = mc.ncore + mc.ncas
```

```
rdm1, rdm2 = mc.fcisolver.make_rdm12(mc.ci, mc.ncas, mc.nelecas)
rdm1, rdm2 = mcscf.addons._make_rdm12_on_mo(rdm1, rdm2, mc.ncore, mc.ncas,
nmo)
```

A CI calculation is performed and the 1- and 2-particle RDMs obtained

6. Write .wfn file with NatOrbs

```
natocc, natorb = numpy.linalg.eigh(-rdm1)
for i, k in enumerate(numpy.argmax(abs(natorb), axis=0)):
    if natorb[k,i] < 0:
        natorb[:,i] *= -1
natorb = numpy.dot(mc.mo_coeff[:,:nmo], natorb)
natocc = -natocc

wfn_file = name + '_fci.wfn'
with open(wfn_file, 'w') as f2:
    wfn_format.write_mo(f2, mol, natorb, mo_occ=natocc)
    wfn_format.write_coeff(f2, mol, mc.mo_coeff[:,:nmo])
    wfn_format.write_ci(f2, mc.ci, mc.ncas, mc.nelecas, ncore=mc.ncore)
```

A .wfn file is written. First the natural orbitals are obtained by diagonalizing the 1-RDM

7. Add constraint expectation value

```
dm = mc.make_rdm1()
elig = numpy.einsum('ij,ji->', dm, lig)
lib.logger.info(mol,'Energy due to ligadure : %f' % (elig))
lib.logger.info(mol,'Final energy : %f' % (elig+emc))
pop = numpy.einsum('ij,ji->',saom,dm)
lib.logger.info(mol,'Population atom %d : %f' % (atm,pop))
```

The expectation value of the constrained is obtained and added to the energy.

4.1 Computational details & Data Tables

Full configuration interaction (FCI) calculations have been performed both in the H_2 and LiH molecules with aug-cc-pVDZ basis sets at their corresponding equilibrium geometries (1.417 and 3.022 au, respectively), as well as at an elongated distance, $R_{AB} = 5$ au. The constrained PySCF just described was used. AOMs and IQA were obtained with PROMOLDEN[16] at several λ values.

Table 1: Representative results for H_2 at $R = R_e$. All data in au.

λ	Q_B	E^A_{self}	E^B_{self}	E_{int}	DI
0.0	-0.000055	-0.485683	-0.485683	-0.193473	0.847624
0.1	0.089175	-0.512471	-0.453220	-0.194691	0.843207
0.2	0.177192	-0.533479	-0.415590	-0.198126	0.829766
0.3	0.262652	-0.548888	-0.373804	-0.203164	0.806971
0.4	0.344077	-0.559083	-0.329355	-0.208959	0.774930
0.5	0.420010	-0.564607	-0.284044	-0.214626	0.734674
0.6	0.489258	-0.566126	-0.239684	-0.219440	0.688232
0.7	0.551096	-0.564377	-0.197802	-0.222938	0.638246
0.8	0.605336	-0.560103	-0.159454	-0.224941	0.587382
1.0	0.692474	-0.546613	-0.095126	-0.224779	0.491229
1.2	0.755903	-0.529831	-0.046732	-0.220523	0.409620
1.4	0.801871	-0.512222	-0.011339	-0.214008	0.344205
2.0	0.879893	-0.463573	0.046901	-0.191529	0.220921
3.0	0.930552	-0.402680	0.078111	-0.161327	0.132490
6.0	0.968739	-0.298531	0.083094	-0.114723	0.061310
8.0	0.976380	-0.255693	0.076680	-0.098498	0.046577

Table 2: Representative results for H_2 at $R = 5$ bohr. All data in au.

λ	Q_B	E^A_{self}	E^B_{self}	E_{int}	DI
0.00	-0.000024	-0.495825	-0.495825	-0.010481	0.075313
0.10	0.018404	-0.498372	-0.490671	-0.012097	0.088986
0.20	0.065736	-0.502039	-0.470668	-0.020821	0.162167
0.25	0.141603	-0.506588	-0.433965	-0.035602	0.275692
0.30	0.344486	-0.516165	-0.329550	-0.073892	0.460381
0.35	0.658154	-0.526522	-0.161143	-0.130151	0.474112
0.40	0.831605	-0.529051	-0.065148	-0.159257	0.284539
0.50	0.924492	-0.526785	-0.013111	-0.173015	0.141179
1.00	0.971434	-0.515487	0.009511	-0.176778	0.055834
4.00	0.991074	-0.481226	0.008998	-0.172852	0.017728

Table 3: Representative results for LiH at $R = R_e$ bohr. All data in au.

λ	Q_{Li}	E^{Li}_{self}	E^{H}_{self}	E_{int}	DI
0.00	0.909499	-7.244830	-0.486850	-0.282051	0.197816
0.10	0.839957	-7.261641	-0.484057	-0.263556	0.305345
0.15	0.680848	-7.298484	-0.475216	-0.214881	0.479713
0.20	0.355873	-7.364537	-0.453802	-0.113451	0.523344
0.25	0.190636	-7.390957	-0.441695	-0.062814	0.386366
0.30	0.135763	-7.395995	-0.437477	-0.047133	0.319630
0.40	0.095756	-7.395349	-0.434028	-0.037564	0.270135
0.50	0.076907	-7.392334	-0.431489	-0.034687	0.253480
0.60	0.061977	-7.388650	-0.427720	-0.033911	0.252193
0.70	0.029426	-7.382277	-0.412175	-0.034237	0.292665
0.75	-0.116895	-7.362386	-0.327733	-0.031311	0.500383
0.80	-0.682028	-7.284446	0.005722	-0.006761	0.520957
1.00	-0.810034	-7.247997	0.072217	-0.001863	0.347114
2.00	-0.848041	-7.202204	0.082503	-0.007336	0.282789
4.00	-0.856336	-7.177515	0.086280	-0.013904	0.268508

Table 4: Representative results for LiH at $R = 5$ bohr. All data in au.

λ	Q_{Li}	E^{Li}_{self}	E^{H}_{self}	E_{int}	DI
0.00	0.824923	-7.274712	-0.503719	-0.195830	0.294513
0.05	0.634559	-7.317259	-0.494826	-0.156564	0.470639
0.10	0.296571	-7.381745	-0.483582	-0.078180	0.426909
0.20	0.079778	-7.410033	-0.480067	-0.024500	0.161244
0.30	0.044139	-7.409910	-0.480059	-0.015952	0.104094
0.40	0.029274	-7.408360	-0.478847	-0.013487	0.087604
0.50	0.003164	-7.407057	-0.467399	-0.014070	0.115625
0.55	-0.493786	-7.392328	-0.199862	-0.029602	0.558696
0.60	-0.891156	-7.371562	0.005712	-0.033814	0.206358
0.80	-0.929852	-7.358537	0.015206	-0.030150	0.137029
1.20	-0.945284	-7.345705	0.016593	-0.028982	0.107626
2.00	-0.953958	-7.330976	0.017897	-0.031117	0.090729

References

[1] Bader, R. F. W. *Atoms in Molecules*; Oxford University Press: Oxford, 1990.

[2] Becke, A. D.; Edgecombe, K. E. A simple measure of electron localization in atomic and molecular systems. *J. Chem. Phys.* **1990**, *92*, 5397.

[3] Martín Pendás, A.; Francisco, E. Quantum Chemical Topology as a Theory of Open Quantum Systems. *J. Chem. Theory Comput.* **2018**, *15*, 1079–1088.

[4] Blanco, M. A.; Martín Pendás, A.; Francisco, E. Interacting Quantum Atoms: A Correlated Energy Decomposition Scheme Based on the Quantum Theory of Atoms in Molecules. *J. Chem. Theory Comput.* **2005**, *1*, 1096–1109.

[5] Menéndez-Crespo, D.; Costales, A.; Francisco, E.; Martín Pendás, A. Real Space *in situ* Bond Energies: Toward a Consistent Energetic Definition of Bond Strength. *Chem. Eur. J.* **2018**, http://dx.doi.org/10.1002/chem.201800979.

[6] Francisco, E.; Martín Pendás, A.; Blanco, M. A. Electron number probability distributions for correlated wave functions. *J. Chem. Phys.* **2007**, *126*, 094102.

[7] Martín Pendás, A.; Francisco, E.; Blanco, M. A. Spin resolved electron number distribution functions: How spins couple in real space. *J. Chem. Phys.* **2007**, *127*, 144103.

[8] Outeiral, C.; Vincent, M. A.; Martín Pendás, A.; Popelier, P. L. A. Revitalizing the concept of bond order through delocalization measures in real space. *Chem. Sci.* **2018**, *9*, 5517–5529.

[9] Francisco, E.; Martín Pendás, A.; García-Revilla, M.; Álvarez Boto, R. A Hierarchy of Chemical Bonding Indices in Real Space from Reduced Density Matrices and Cumulants. *Comput. Theor. Chem.* **2013**, *1003*, 71–78.

[10] Martín Pendás, A.; Francisco, E.; Blanco, M. A. An Electron Number Distribution View of Chemical Bonds in Real Space. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1087–1092.

[11] Martín Pendás, A.; Francisco, E. Real space bond orders are energetic descriptors. *Phys. Chem. Chem. Phys.* **2018**, *20*, 16231–16237.

[12] Perdew, J. P.; Parr, R. G.; Levy, M.; Balduz, J. L. Density-Functional Theory for Fractional Particle Number: Derivative Discontinuities of the Energy. *Phys. Rev. Lett.* **1982**, *49*, 1691–1694.

[13] Martín Pendás, A.; Francisco, E.; Blanco, M. A. Charge transfer, chemical potentials, and the nature of functional groups: answers from quantum chemical topology. *Faraday Discuss.* **2007**, *135*, 423.

[14] Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K.-L. PySCF: the Python-based simulations of chemistry framework. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018**, *8*, e1340.

[15] Cioslowski, J.; Stefanov, B. B. Electron flow and electronegativity equalization in the process of bond formation. *J. Chem. Phys.* **1993**, *99*, 5151–5162.

[16] Martín Pendás, A.; Francisco, E. A QTAIM/IQA code (Available from the authors upon request at ampendas@uniovi.es).

[17] Keith, T. A. The AIMAll program. The code is avalaible at `http://aim.tkgristmill.com`. 2015; The code is avalaible at `http://aim.tkgristmill.com`.