Contents

1 Introduction **2**

2 Electron distribution functions **5**

3 The two center bond **6**
3.1 The 2c,1e bond 6
3.2 The 2c,2e bond 8
3.3 2c,2e triplets. EDFs and the Pauli principle 10
3.4 The Aufbau principle in diatomics 12
3.5 Building polyatomics from 2c,2e bonds 13

4 The three center bond **15**
4.1 The 3c,1e bond 15
4.2 The 3c,2e bond 16
4.3 The Pimentel-Rundle 3c,4e bond 21

5 The equivalent primitives manifold **22**
5.1 All the states from an electron configuration 25

6 From EDFs to chemical bonding **27**
6.1 Checking statistical (in)dependence 28
6.2 Grasping the graph structure 30
6.3 Partial EDFs 33

7 Fitting EDFs **35**

8 Conclusions **37**

Bibliography **40**

Chemical bonding from the statistics of the electron distribution

A. Martín Pendás* and E. Francisco*

Abstract

An introduction to the theory of chemical bonding from the point of view of the statistics of the electron distribution is presented. When atoms bind to form a molecule, their originally fixed number of electrons ceases to be a well-defined observable, and this implies that their in-the-molecule electron populations fluctuate. If a chemically meaningful definition of an atom in a molecule is assumed, the probabilities of finding a given number of electrons in each of the atoms comprising the molecule can be computed. We show in this review how the complete electron distribution function (EDF) can be used to reconstruct the basic concepts and quantities used in chemical bonding without recourse to the orbital paradigm. From the statistical point of view, which inherits Born's probabilistic interpretation of quantum mechanics, a set of atoms are bonded when their electron populations are mutually dependent. We quantify this statistical dependence by the cumulant moments of the EDF, which provide a consistent description of both two- and multi-center bonding. Particular attention is paid to building EDFs from model wavefunctions. With this, a simple bridge with orbital thinking is built. The statistical interpretation allows to easily classify all possible bonds of a given kind. We show that there are vast unexplored territories that should receive due consideration. Although building EDFs from models is easy and very instructive, the contrary is considerably more difficult. Recipes to extract chemical information from computed EDFs are also reviewed and, in all the cases, simple toy systems are used to show how the methodology works, allowing non-experts to follow easily the presentation.

1 Introduction

The theory of the chemical bond is inextricably linked to the development of quantum mechanics. After Heitler and London's paper [1] on the dihydrogen

*Departamento de Química Física y Analítica. Facultad de Química. Universidad de Oviedo. 33006-Oviedo. Spain. e-mail: ampendas@uniovi.es & evelio@uniovi.es

molecule was published, it turned out clear that the basic framework to understand chemical phenomena within physics had been discovered. In the years to come, the work of Lewis (who had introduced the electron pair well before the advent of the modern quantum theory), Pauling, Slater, Mulliken, and many others shaped what we now know as theoretical chemistry. Given the dependence of this new discipline on calculation, the computational efficiency of the molecular orbital (MO) scheme advocated by Mulliken over the more cumbersome valence bond (VB) alternative of Heitler and London spread by Pauling led the MO paradigm to overwhelmingly dominate the field. Within the MO theory, the simplest mean-field or single-determinant (SD) approximation leads to the one-electron states that will have soon been used during a century to rationalize the chemical bond. In this picture, atoms lose their identity as a result of molecular orbitals being generally delocalized over the molecular space. As a result, different methods [2, 3] have been proposed over the years to reintroduce the atom into theoretical chemistry. Once the atom is back, a molecule may be seen as a set of interacting atoms, much like in the phenomenological atomistic simulations popular in physics [4].

Be that as it may, the deep roots of chemical bonding theory are still a source of bitter disputes among different schools. Some of them focus on the role of quantum mechanical interference [5], others on the buildup of electron density in internuclear regions [6], or on the decrease in the kinetic energy density accompanying bond formation [7], to cite just a few. Although every chemist learns soon about Born's [8] statistical interpretation of quantum mechanics, and the *cloud* image of the square of an orbital as a probability density has become ubiquitous, it is surprising that not much effort has been devoted to understand the chemical bond in statistical (or probabilistic) terms.

There are several simple clues that support the existence of such a probabilistic interpretation. It is for instance well known that the bonding glue is exchange [9], and that no bonding can exist in an orbital picture without overlap. Since to define a *traditional* chemical bond we need at least two interacting atoms A and B, it is not hard to understand that exchange, or overlap, among the electrons located originally on the non-interacting atoms leads to *delocalization*: electrons originally seen on A may now be located on B. For instance, binding of a H atom to a proton to form the $H_2{}^+$ cation, possibly the simplest chemical species, implies that the electron has delocalized over the two nuclei. The probability of finding the electron in each of the atoms has changed from $(1, 0)$ to $(0.5, 0.5)$.

In a more mathematical language, delocalization is a consequence of the expansion of the Hilbert space following the interaction of A and B, e.g. the isolated B system characterized by N_B electrons lives now in an expanded space with up to $N_A + N_B$ electrons when considered as a subsystem of the AB molecule. Our standpoint in this review is that this delocalization can be monitored as a signature of bonding. If two atoms are bonded, their electron populations must necessarily *fluctuate* as a result of the delocalization of their electrons. As we intend to show, the statistics of electron populations provides a refreshing new perspective of chemical bonding in which Born's probabilistic

interpretation is directly introduced in the chemical language.

As we will see, the probabilistic view is able to encode equally two- and multicenter bonding, providing descriptors for the latter which are intimately linked with indicators successfully used over the years [10–15]. Unlike in other frameworks, the statistical image unifies all types of bonding in terms of two- or many-center population fluctuations. Moreover, it provides hierarchical classification schemes. All possible fluctuation types can be categorized, something that is not easy with other methodologies. This expands considerably the chemical bonding landscape, allowing to understand exotic bonding situations on equal footing with standard ones. For instance, the statistical treatment allows to examine the connection between seemingly independent phenomena like bonding and entanglement. Moreover, statistical descriptors of bonding (i.e. statistical bond orders) are directly related to the system's energetics. This already known link [16, 17] is extremely important, for it provides a rigorous tool to estimate the energetic strength of a given *bond*.

Finally, the statistical perspective presented here provides a valuable toolbox for constructing models. These are much easier to build than in MO theory, and provide a nice playground in which to examine when this or that situation may be found in a real system. Models are also important for instructional purposes.

The aim of this review is to present in a unified manner the statistical description of chemical bonding. Although all the material that will be presented can be already found dispersed in several works [18–23], no single logically consistent account exists. We have decided to focus on models and to provide an example-assisted presentation, since formal accounts severely limit the potential audience of our conclusions. The basis of the statistical theory of the chemical bond (STCB in the following) is the construction of the electron distribution function (EDF) [19, 23]. The EDF for a system of N electrons and m nuclei is the probability distribution of a partition $S = (n_1, n_2, \ldots, n_m)$, $n_1 + n_2 + \cdots + n_m = N$ of the electrons in the available nuclei, $p(S)$. The spin of each electron can also be specified. We talk then of a spin-resolved EDF. It is clear that EDFs imply that atoms-in-the-molecule (AIMs) have been introduced by whatever means. Most of the available definitions of AIMs are easily implemented [24], from that provided by Mulliken's prescription to the quantum theory of atoms in molecules (QTAIM) of Bader and coworkers [3]. Although we prefer the latter to show production results, all the insights are independent of the particular AIM implementation, so that we will turn to Mulliken's for many models and examples.

We start by briefly considering how an EDF is rigorously obtained and how it can be approximated by back of the envelope calculations. Then we will characterize two-, three– and general multi-center bonds from simple models and switch to the inverse problem, characterizing chemical bonds from computed EDFs. We will end the presentation with some conclusions. A number of more detailed mathematical derivations as well as a few specific points that may be difficult to grasp for the non-specialist can be found in the supporting information.

2 Electron distribution functions

The probability of finding a partition $S = (n_1, n_2, \ldots, n_m)$ of an N electron system described by the wavefunction $\Psi(1, N)$ into m atoms can be found as [19]

$$p(S) = \mathcal{P} \int_D \Psi^* \Psi d\boldsymbol{x}_1 \boldsymbol{x}_2 \cdots d\boldsymbol{x}_N, \tag{1}$$

where $\boldsymbol{x}_i$ is a spin-spatial coordinate for electron i, $\mathcal{P} = N!/(n_1! \ldots n_m!)$ is an indistinguishability combinatorial factor, and D is a domain in which the first n_1 electrons are projected (integrated) onto atom one, the second n_2 electrons onto atom two, etc. If the spin coordinates are not summed up, then we arrive at a spin-resolved partitition [20], in which we specify also the m_s value of each electron and $S \equiv \{n_1^\alpha, n_2^\alpha, \ldots, n_m^\alpha; n_1^\beta, n_2^\beta, \ldots, n_m^\beta\}$. The projection (integration) procedure for different definitions of AIMs can be unified through the use of atomic weight functions [24]. We will consider here only Mulliken's projection, which is convenient in modelling, and QTAIM integration, in which the above integration is simply done over the real space atomic domains Ω_i.

Although algorithms to compute EDFs in the case of correlated wavefunctions are known [19], it is in the SD case when back of the envelope calculations can be performed with simple models to gain chemical insight. If $\Psi(1, N) = (N!)^{-1/2} \det |\chi_1(1) \ldots \chi_N(N)|$, where χ_1, $\ldots$, χ_N are the N occupied molecular spinorbitals (MSO), the probability of a partition S is simply given by [25]

$$p(S) = \mathcal{N} \sum_{\{k_j\} \in \mathcal{S}_N} \det \left[S_{ij}^{b(k_j)} \right]. \tag{2}$$

where $\mathcal{N} = \mathcal{P}/N!$, $\mathcal{S}_N$ is the set of $N!$ permutations of the $1 \ldots N$ set, $\{k_j\} \equiv \{k_1, \ldots, k_N\}$ is one of these permutations, and the overlap integral between primitives χ_i and χ_j are projected over the atoms in the order that leads to the partition S subjected to permutation k_j. The procedure will become clear in the following.

The projected overlap integrals are defined as $S_{ij}^b = \int_b \chi_i(\boldsymbol{x}) \chi_j(\boldsymbol{x}) d\boldsymbol{x}$. If we use QTAIM atoms, $\int_b \equiv \int_{\Omega_b}$. If we perform a Mulliken projection and spinorbitals are expanded in a minimal basis, $\chi_i = \sum_j c_j^i \phi_j$, then

$$S_{ij}^b = \sum_k \sum_l c_k^i c_l^j \langle \phi_k | \phi_l \rangle^b, \tag{3}$$

where the last scalar product is the share of the primitive overlap in atom b, which may be defined in multiple ways. In the zero differential overlap (ZDO) approximation that we will use in modelling this can just be written as

$$S_{ij}^b = \sum_{k \in b} \sum_{l \in b} c_k^i c_l^j \langle \phi_k | \phi_l \rangle. \tag{4}$$

These sums can be evaluated immediately from model orbitals, as we will show.

Notice that if spinorbitals in $\Psi(1,N)$ are ordered by spin, so that all α χ's come before the β ones, the overlap determinants in Eq. 2 are spin-block diagonal, so that the probability distribution can be written as the product of independent distributions for the α and β sets: $\boldsymbol{p} = \boldsymbol{p}^\alpha \otimes \boldsymbol{p}^\beta$. At the one-determinant level, alpha electrons are statistically independent from beta electrons. As we will see, this is the statistical basis for the Lewis pair. Moreover, in closed shells $\boldsymbol{p}^\alpha = \boldsymbol{p}^\beta$. This property is very valuable in reducing the algebraic work when working with models.

Once the EDF is known, all kind of statistical measures can be obtained from it. We can, for instance, construct marginal distributions from the full joint one. As an example, the one atom distribution $p(n_1)$ gives the probability of finding a given number of electrons in atom 1, $p(n_1) = \sum_{n_2,\ldots,n_m} p(n_1, n_2, \ldots, n_m)$. This is a measure of its entanglement with the rest of the system, and its Shannon entropy $\sum_i p(i)\log_2(p(i))$ does only vanish if the atom is isolated. Moreover, its first moment $\langle n_1 \rangle = \sum_i i \times p(i)$ gives the standard atomic population. As we will show, two-, and in general, n-atom marginals provide information about multicenter bonding through the several central or cumulant (see below) moments of the distributions. Since the number of elements in a given distribution is limited, the several marginal distributions that we will be using can be exhaustively classified. For instance, a two-atom (two-center), two-electron distribution has only three different electron partitions: two electrons in the first atom, none in the second, one and one, none and two. Given that the sum of the three probabilities add to one, any 2c,2e distribution can be classified using two parameters. This is the basis for the classification of links that will be presented.

3 The two center bond

We sample here the most important type of chemical bond, in which only two atomic centers are involved. We intend to work out several examples from scratch, laying the grounds for more interesting situations that will be delayed to subsequent sections. We start by examining the two-center one-electron bond.

3.1 The 2c,1e bond

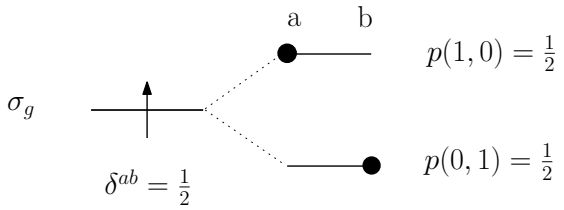
Although the consideration of this type of link is not usually found in freshman chemistry courses (the Lewis pair is to blame), there is no doubt that much can be learned from the molecular hydrogen cation, $H_2{}^+$. Let us set up our model: with one electron, the EDF has two components: $p(1,0)$, the probability of finding the electron on the left H, H_a, and $p(1,0)$, that of finding it on the right H, H_b. Symmetry dictates the solution, so that each of them needs be equal to 1/2. Let us, however, build the solution from the bottom up through Eq. 2.

To do that, we need a model wavefunction determinant, equal in this case to the electron's orbital. Using a minimal basis set formed up from two $1s$ functions, $\sigma_g = (1s_a + 1s_b)/\sqrt{2(1+s)}$, where s is $\langle 1s_a | 1s_b \rangle$. In the ZDO approximation, $\sigma_g \approx (1s_a + 1s_b)/\sqrt{2}$, and $S^a = c_a^2 = 1/2$, $S^b = c_b^2 = 1/2$. This holds by

symmetry whatever atomic partition we use. Then Eq. 2 reads $p(n_a = 1, n_b = 0) = S^a = 1/2$, $p(n_a = 0, n_b = 1) = S^a = 1/2$, as expected. It is clear that the average number of electrons in each H atom equals $1/2 = 0 \times p(0,1) + 1 \times p(1,0)$, and that the electron population fluctuates, the electron is delocalized between the two atoms. Taking a large number of snapshots, half the frames will show the electron in atom a, half in b. A measure of this fluctuation is the covariance of the distribution, $\mathrm{cov}(n_a, n_b) = \langle (n_a - \bar{n}_a)(n_b - \bar{n}_b) \rangle$, where $\bar{n}_a = \langle n_a \rangle$. Using the EDF,

$$\mathrm{cov}(n_a, n_b) = \sum_{n_a, n_b} (n_a - \bar{n}_a)(n_b - \bar{n}_b) \times p(n_a, n_b). \tag{5}$$

For a bivariate distribution, a vanishing covariance implies that the two electron populations are independent, i.e. that $p(n_a, n_b) = p(n_a)p(n_b)$. In the present case, the covariance is immediately computed as $\mathrm{cov}(n_a, n_b) = \frac{1}{2}(1 - \frac{1}{2})(-\frac{1}{2}) + \frac{1}{2}(-\frac{1}{2})(1 - \frac{1}{2}) = -\frac{1}{4}$. Let us build for convenience the quantity $\delta^{ab} \equiv \delta(a,b) = -2\mathrm{cov}(n_a, n_b)$. Its value equals $1/2$ in this case, which is the conventional bond order associated to a 2c,1e bond. Scheme 1 illustrates these ideas.



Scheme 1: The 2c,1e bond in H_2^+

The scaled covariance just defined is known in the literature as the delocalization index (DI) [10, 13, 14], and can be obtained from the exchange-correlation density without recourse to the EDF concept. It has been used as a measure of the covalent bond order, and is intimately related to the covalent component of the interaction energy between two atoms [26]. As we can see, δ will remain equal to $1/2$ in $H_2{}^+$ up to dissociation inasmuch the electron remains described by the symmetric σ_g function. The DI is thus a measure of the spatial entanglement of the electron. *The quantity that is associated to bond order uncovers spatial quantum mechanical entanglement.* The DI in a two center system is semipositive definite, since an increase in the electron population of one atom is necessarily followed by a decrease in the other. We will see that this needs not be the case in more general situations.

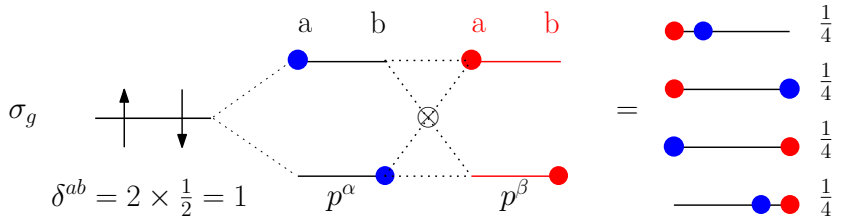
Let us now introduce an electronegativity difference in the two atoms A, B that are bonded via the one-electron bond (an example would be the HHe^{2+} system). We can now use $\sigma_g = \lambda 1s_a + \mu 1s_b$, with $\lambda^2 + \mu^2 = 1$, and play with the results. Now, $p(1,0) = \lambda^2 = \bar{n}_a$ and $p(0,1) = \mu^2 = \bar{n}_b$, as expected from basic quantum mechanics. More interestingly, $\mathrm{cov}(n_a, n_b) = -\lambda^2\mu^2$, so that $\delta = 2\lambda^2\mu^2$. An electronegativity difference leads to a decrease in the DI, that is

confined: $\delta \in [0, 1/2]$. Bond polarity thus decreases the bond order, maximum if the electron is perfectly *shared*, and falling to zero if the electron is completely *localized* in one of the atoms so that only one of the two *resonance structures* appears. Simple as it is, this naïve example introduces a number of standard concepts in chemical bonding theory in an effortless manner. We define a *pure* or *perfect* 2c,1e bond as that with maximum delocalization. It is achieved when $\lambda = \mu$.

3.2 The 2c,2e bond

Adding a second electron to the above picture brings the Lewis pair in. If we do not leave the SD approximation, the simplest model changes us from the $H_2{}^+$ to the H_2 molecule. Now $\Psi = |\sigma_g \bar{\sigma_g}|$. Although we might use Eq. 2, it is easier to take profit of the independence of the p^α and p^β distributions for a single determinant. Since each spin block is equivalent to the $H_2{}^+$ case, the EDF is the product of those found in Scheme 1. This leads to a binomial distribution: $p(2,0) = p^\alpha(1,0) \times p^\beta(1,0) = 1/4$, $p(0,2) = p(2,0)$, $p(1,1) = p^\alpha(1,0) \times p^\beta(0,1) + p^\alpha(0,1) \times p^\beta(1,0) = 1/2$. The two electrons delocalize independently, forming a Lewis pair. A 2c,2e bond is thus better envisioned as two quasi-independent opposite spin 2c,1e bonds. If we are interested in the spin resolved EDF, the direct product shown in Scheme 2 shows that each of the four spin structures is equally populated. We define these distributions as those of a *perfect or pure* 2c,2e covalent bond.

Since the covariance of independent contributions is additive, $\delta = 1$, in agreement with its bond order interpretation. In general, $\mathrm{cov}(n_a, n_b) = \mathrm{cov}(n_a^\alpha, n_b^\alpha) + \mathrm{cov}(n_a^\beta, n_b^\alpha) + \mathrm{cov}(n_a^\alpha, n_b^\beta) + \mathrm{cov}(n_a^\beta, n_b^\beta)$. This implies that the same spin ($\sigma\sigma$ in the following) covariances are measuring the spatial entanglement or delocalization of the two independent electrons, while the opposite spin contributions ($\sigma\sigma'$) (zero in this case) refer to opposite spin electrons dependency, i.e. to Coulomb correlation. Although we will not engage here in a full discussion of the effects of electron correlation on EDFs [19, 23], it is rather easy to understand the behavior of the probability distribution from simple symmetry properties.



Scheme 2: The ideal 2c,2e bond in H_2 (blue-red $\equiv \alpha - \beta$).

At dissociation, where the SD description does not hold, and we need at least a two-determinant wavefunction $\Psi = c_1|\sigma_g\bar{\sigma}_g| + c_2|\sigma_g\bar{\sigma}_g|$ (complete active space (CAS) calculation with two electrons in two spinorbitals) to correctly

dissociate dihydrogen, the spinless EDF collapses onto $p(1,1) = 1$, $p(0,2) = p(2,0) = 0$. Full configuration interaction (FCI) results are known [27]. Thus, the electrons are perfectly localized (one in each atom), and $\delta = 0$. However, $^1\Sigma_g$-dissociated dihydrogen is still a singlet, so the spin resolved EDF contains two equiprobable components: $p(\uparrow,\downarrow) = p(\downarrow,\uparrow) = 1/2$. As it can be readily calculated, $p \neq p^\alpha \otimes p^\beta$. Actually, the p^α and p^β distributions are equal to those in $H_2{}^+$. Straighforward calculation leads to $\mathrm{cov}(n_a^\alpha, n_b^\alpha) = \mathrm{cov}(n_a^\beta, n_b^\beta) = +1/2$, so that the total covariance vanishes. The two electrons are spin-entangled, describing a non-bonded singlet diradical.

As it has already been fully discussed [27], at equilibrium the SD approximation is reasonable, and Coulomb correlation slightly localizes the electrons in the atoms, increasing $p(1,1) \approx 0.58$ at the expense of the $p(2,0), p(0,2)$ contributions. This decreases the DI to about 0.85, so that correlated bond orders are usually smaller than non-correlated ones [28]. For our purposes here, although a perfect 2c,1e bond may exist, *electron correlation prevents the existence of a pure Lewis pair or a pure covalent bond.*

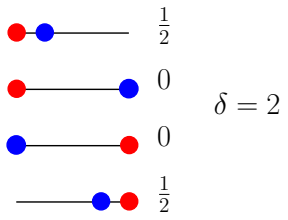
Polarity is obviously introduced with our λ, μ model as before. Due to statistical independence, the polar $|\sigma_g \bar{\sigma}_g|$ determinant EDF is immediately constructed as a product of the one electron distributions, so that $\delta = 4\lambda^2\mu^2$. For a 2c,2e non-correlated bond, *polarity decreases covalency and viceversa.*

The introduction of electron correlation (thus abandoning statistical independence) opens many new possibilities. The complete (spinless) EDF landscape for a 2c,2e bond depends on two variables, and since the three probabilities add to one, any EDF can be represented on a triangular diagram in which each corner is a pure $(1,1)$, $(0,2)$, or $(2,0)$ distribution. We have shown [23] that any point can be described by the formal charge of one of the atoms $q \in [-1,1]$, measuring polarity, and a correlation factor $f \in [-1,1]$ that determines how the two electrons are correlated.

From our present perspective it is better to start from the two non-interacting electrons, each with a probability distribution $p(1,0) = p = 1 - p(0,1)$, where p plays the same role as λ^2. This leads to the binomial $p(2,0) = p^2$, $p(1,1) = 2p(1-p)$, $p(0,2) = (1-p)^2$. We now introduce the correlation factor f so that $p(1,1) = 2p(1-p)(1+f)$, that subtracts $p(1-p)f$ from each of the $(2,0)$ and $(0,2)$ components. This symmetric subtraction stems from considering correlation as a phenomenon affecting the interaction between the electrons in a pair, not the intercenter charge transfer. As it can be easily proven, the electron population of each center is not affected when the correlation factor f is introduced and the symmetric partitioning is employed. When $f > 0$ the electrons try to avoid staying in the same center (in a physicist-like language, they have a repulsive on-site energy), and $p(1,1)$ increases over its uncorrelated value. This is the situation expected in chemistry. When $f < 0$ (attractive on-site energy) $p(1,1)$ decreases, so that the electrons prefer to pair and move together. This is equivalent to a bosonization of their behavior. It is easily shown that the average population of each atom in this model is not affected by f, $\langle n_a \rangle = 2p$, but the covariance senses strongly correlation, $\delta = 4p(1-p)(1-f) \in [0,2]$. This extremely simple expression provides a very useful general model of the 2c,2e

bond.

Since polarity is easy to introduce in the models, we will basically restrict in the following to non-polar interactions (except in exceptionally interesting cases) without loss of generality. With $p = 1/2$, then $\delta = (1 - f)$ also runs over its possible spectrum, and $p(1,1) = (1 + f)/2$, $p(2,0) = p(0,2) = (1 - f)/4$. The journey from $f = 0$ to $f = 1$ describes the dissociation of H_2 and is no more interesting. However, the bosonized limit $f = -1$ leads to the resonance between two electrons located at each center, as described in Scheme 3.



Scheme 3: The bosonized 2c,2e bond in H_2.

This shows that standard bond orders may be difficult to interpret in many interesting situations. DIs greater than one for 2c,2e bonds have actually been found, for instance in the E,F-$^1\Sigma_g^+$ excited state of H_2, which correlates with a $1\sigma_u^2$ [29]. This is, grossly speaking, the traditional $H^+{-}H^- \longleftrightarrow H{-}H^+$ VB resonance. This type of bonding pattern can also be found via electron-phonon coupling, being thus the simplest possible chemical analogue of a Cooper pair.

A straightforward back of the envelope spin analysis on the $(p = 1/2, f)$ non-polar bonds shows that $DI^{\sigma\sigma} = 1$, independently of f, so that $DI^{\sigma\sigma'} = -f$. All non-polar 2c,2e bonds, correlated or not, have a $DI^{\sigma\sigma} = 1$ bond order. This is why some authors [30] prefer to use same-spin component as a more chemical bond order definition. In our opinion, this result should be better read in the following terms: if we ignore the correlated electron movement, each of the two indistinguishable electrons in between two equivalent nuclei are fully spatially entangled (like in $H_2{}^+$) with $DI^{\sigma\sigma} = 1$. For a non-polar 2c,2e system, $DI^{\sigma\sigma'} = -f$ measures the degree of pair repulsion/attraction.

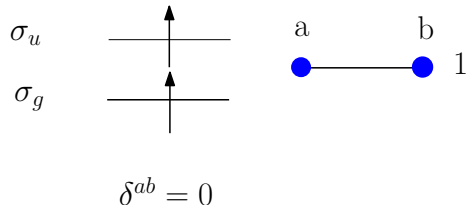
3.3 2c,2e triplets. EDFs and the Pauli principle

In our bottom-up approach, much is learnt from modelling the first $^3\Sigma_u^+$ non-bonded state in H_2. The simplest MO description of this state implies populating the σ_g and σ_u states: $\Psi = |\sigma_g\sigma_u|$ for the $M_S = 1$ spin component. With two populated orbitals we need to use Eq. 2 for the first time. In the ZDO approximation, $\chi_{1,2} = \sigma_{g,u} = (1s_a \pm 1s_b)/\sqrt{2}$, and our Mulliken strategy leads to $S_{ij}^a = (1/\sqrt{2})^2 = 1/2\ \forall i,j$. The elements of S_{ij}^b are also equal to 1/2 except

$S^b_{12} = S^b_{21} = -1/2$. If we now use Eq. 2,

$$p(2,0) = \begin{vmatrix} \overbrace{1/2}^{a} & \overbrace{1/2}^{a} \\ 1/2 & 1/2 \end{vmatrix} = 0, \quad p(1,1) = 2 \begin{vmatrix} \overbrace{1/2}^{a} & \overbrace{-1/2}^{b} \\ 1/2 & 1/2 \end{vmatrix} = 1.$$

Obviously, $p(0,2) = p(2,0) = 0$. There is only one permutation leading to two electrons on center a, while two equivalent ones, (a,b) and (b,a), lead to one electron on each center. The result implies that the two-same spin electrons are perfectly localized so that δ vanishes (Scheme 4).



Scheme 4: The triplet distribution in model H_2.

The calculation shows that it is the *interference* between σ_g and σ_u components that cancels out the delocalization. In chemical terms, if we only allow the $1s$ orbitals to interact, the delocalization channel is blocked once an electron is present. In $H_2{}^+$ we may think of one occupied and one empty $1s$ orbital. The channel is open, and the electron delocalizes. Adding a second electron forces the first to stay in its original center, and so does the second. This is Pauli principle acting in real space.

From the present perspective, *there is no reason to call σ_u an antibonding orbital*. As expressed above, a $|\sigma_u\bar{\sigma}_u|$ determinant delocalizes electrons as well as its $|\sigma_g\bar{\sigma}_g|$ counterpart, and has $\delta = 1$. It is only through the g, u interference that the open delocalization channels in both determinants become blocked. Of course, if we now couple the p^α distribution to an equivalent p^β one to form a closed shell system like He_2 with a $\sigma_g^2\sigma_u^2$ configuration, the product of the different spin configurations lead (like in going from $H_2{}^+$ to triplet H_2) to an additive δ which is still zero, and to a pair of completely localized electrons in each center. In this way it is easy to understand the Aufbau principle in diatomics that we will briefly examine in the next subsection. The "bonding-antibonding" cancellation of the bond order appears in the EDF formalism as a Pauli induced localization of electrons. The *no bonding without delocalization* maxim thus acquires a clear meaning.

Unlike the triplet H_2 case, the triplet model is sensitive to the ZDO approximation. It is easy to show that as overlap between the $1s$ functions is switched on, the S_{12} matrix elements are no longer constrained to be equal to $\pm 1/2$, and the $p(2,0)$ probability ceases to be zero, although it remains much smaller than in the singlet. If a true real space analysis is done [27], then the evolution of delocalization in the triplet can be followed.

3.4 The Aufbau principle in diatomics

The bonding pattern in simple diatomics appears effortlessly as new delocalization channels (orbitals) open and close on adding new electrons. Let us briefly summarize how this is done with first and second row diatomics.

Starting with the $1s$ block, in $1s^2$ H_2 the independent opposite spin $1s^\alpha$ and $1s^\beta$ channels are open and, as already explained, each p^α or p^β distribution provides an additive $\delta = 1/2$ contribution. The total bond order is 1, and the two electrons are perfectly delocalized in the single determinant approximation. Passing to He_2 adds one electron to each of the $1s$ spin channels, blocking both of them. The spinless δ vanishes. Two localized $1s^2$ cores are formed.

In Li_2 a $2s$ block is added, so that at the SD level and within the ZDO approximation, $\Psi = |1\sigma_g 1\bar{\sigma}_g 1\sigma_u 1\bar{\sigma}_u 2\sigma_g 2\bar{\sigma}_g|$, with $1\sigma_{g,u} = (1s_a \pm 1s_b)/\sqrt{2}$, $2\sigma_g = (2s_a + 2s_b)/\sqrt{2}$. Following Eq. 2, and considering the independent spin channels separately we have 3 alpha electrons and only two non-equivalent distributions, $(3,0) \equiv (0,3)$ and $(2,1) \equiv (1,2)$:

$$p^\alpha(3,0) = \left|\begin{array}{ccc} a & a & a \\ \hline 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1/2 \\ \hline 1\sigma_g & 1\sigma_u & 2\sigma_g \end{array}\right| = 0,$$

$$p^\alpha(2,1) = 2\left|\begin{array}{ccc} a & b & a \\ \hline 1/2 & -1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1/2 \\ \hline 1\sigma_g & 1\sigma_u & 2\sigma_g \end{array}\right| = 1/2$$

Notice that the overlap matrices are block-diagonal due to the $1s, 2s$ orthogonality. The 2×2 blocks are easily recognized as those in triplet H_2 or He_2. Given that any new orbital block we add is orthogonal to all previous ones, once the block closes, i.e. once the subshell closes, its contribution to the overlap matrices becomes isolated, remaining block-diagonal in the following. In other words, in this case the $1s^2$ localized core is preserved, and the $2s$ channel behaves independently of the $1s$ core. With only $1s$ and $2s$ channels, the probability of finding three same spin electrons in the same center vanishes, and the only remaining EDF components are $p^\alpha(2,1) = p^\alpha(1,2) = 1/2$. This obviously provides $\delta^{\alpha\alpha} = 1/2$, and we can safely ignore the $(1,1)$ constant contribution so that there is effectively only one alpha electron that delocalizes: $p^\alpha(2,1) \equiv p^\alpha(1,0)$. Adding the independent beta terms, an image of two localized two-electron cores and a perfect covalent bond with $\delta = 1$ formed by two opposite spin perfectly delocalized electrons emerges. In this approximation, Li_2 behaves like H_2.

The following rationale emerges. Two extra electrons close the $2s$ channels, and Be_2 has $\delta = 0$. In traditional MO terms, two full $1, 2\sigma_{g,u}$ bonding-antibonding pairs. If we add the $2p$ block, we open six independent delocalization channels (from spin independency coupled to $m_l = -1, 0, 1$ or x, y, z

orthogonality). On traversing B_2, C_2 and N_2 the bond order increases from one to three, as expected. An interesting situation is, for instance, dioxygen, where two extra alpha electrons are added on top of dinitrogen that occupy the (antibonding) $1\pi_{g,x}$ and $1\pi_{g,y}$ states as a triplet. The 1π alpha channels are thus blocked with localized $p_{x,y}$ orbitals on both centers such that the bond order of the alpha set of electrons is 1. Contrarily, the beta $1\pi_u$ set is still fully open, with bond order equal to 3. On average, $\delta = 2$. This simple picture, which has been described in detail elsewhere [19, 20], shows how, with the same back of the envelope effort, the EDF image provides a far more detailed picture of the electron distribution than the MO one. In Ne_2, all the L channels are full, δ vanishes again, and in the SD approximation, two Ne atoms cannot delocalize any electron and do not bind to each other.

3.5 Building polyatomics from 2c,2e bonds

The electronic structure of simple polyatomic molecules can be easily modeled from several 2c,2e distributions that are considered independent. We take the water molecule as an example. Table 1 shows the dominant EDF component at the HF//6-311G(d,p) level obtained from QTAIM atoms. We can check that to about 1% in probability, each H atom is only involved in the exchange of one electron, so that in terms of 2c,2e links there are only two electrons in the oxygen atom that are also engaged in bonding. From the Table we compute easily that the net charge of the oxygen atom is -1.203 electrons, and that $\delta(\mathrm{O},\mathrm{H}) = 0.648$, $\delta(\mathrm{Ha},\mathrm{Hc}) = 0.008$. If we ignore the last covariance, a clear model of two 2c,2e links emerges. Summing up, six electrons in oxygen are always localized and thus inactive (i.e. in chemical terms the 1s core and two lone pairs). We can safely ignore them.

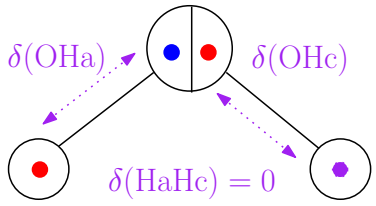
Table 1: QTAIM EDF for a HF//6-311G(d,p) calculation in H_2O. Only structures contributing to the EDF with a probability greater than 0.01 are shown.

n(Ha) a	n(O) b	n(Hc) c	p
0	10	0	0.4133
0	9	1	0.2032
1	9	0	0.2032
1	8	1	0.0971
0	8	2	0.0274
2	8	0	0.0274
1	7	2	0.0126
2	7	1	0.0126

To build an EDF model, we construct a determinant with the three inactive functions of the oxygen atom and two bonding orbitals, ϕ_1 and ϕ_2, built from

the $1s_a \equiv a$ and $1s_c \equiv c$ primitives of both hydrogen atoms, together with two equivalent orthogonal oxygen hybrids that bind to Ha and Hc, respectively, say b and b'. With this, $\phi_1 = \lambda a + \mu b$, $\phi_2 = \lambda c + \mu b'$. It is a simple exercise to obtain the alpha or beta block distributions obtained by the product of two equivalent polar 2c,2e bonds already examined: $p^\alpha(a,b,c) = p^\alpha(a,b) \otimes p^\alpha(b,c)$, so that $p^\alpha(1,1,0) = p^\alpha(0,1,1) = \lambda^2\mu^2$, $p^\alpha(0,2,0) = \mu^4$, $p^\alpha(1,0,1) = \lambda^4$. In this model, $\delta(\mathrm{Ha},\mathrm{Hc})$ strictly vanishes, not far from our assumption. We will return to this result, since *the covariance between two atoms not directly linked in a model of independent 2c,2e bonds is necessarily zero*, in agreement with the chemical intuition that atoms not linked by a "dash" should have independent electron populations. The situation is sketched in Scheme 5.

We can now fit the parameters of the model to the data. This can be done in many ways, see below. For our present purposes, a simple procedure is to recognize that the two bonds are statistically independent and compute the average population of any of the H atoms, that turns out to be $\langle n(\mathrm{Ha})\rangle = 2\lambda^2$. This leads to $\lambda^2 = 0.199$. From this result, $\delta(\mathrm{H},\mathrm{O}) = 4p(1-p) = 4\lambda^2\mu^2$ is equal to 0.638, to be compared with the actual Hartree-Fock value, 0.648. The fact that a single parameter (λ) model can very reasonably reconstruct the dominant Hartree-Fock QTAIM EDF of H_2O shows that the Pauli principle decreases enourmosly the complexity of the electron distribution that, in many cases, may be very accurately understood from very simple principles.



Scheme 5: The H_2O model.

The general procedure of extracting chemical information from a general correlated EDF is considerably more cumbersome than that shown in the previous toy example. An introduction to how this can be done will be delayed to Section 7.

In the general case, the molecular EDF coming from a set of n independent 2c,2e bonds is the product of the n 2c,2e probability distributions: $\boldsymbol{p} = \bigotimes_i \boldsymbol{p}^i$. The i-th component $\boldsymbol{p}^i$ specifies the $(2,0),(1,1),(0.2)$ probabilities betweeen the centers i_1 and i_2 associated to the bond. If the links are supposed to be uncorrelated ($f = 0$), then each $\boldsymbol{p}^i$ is specified by one polarity parameter (λ or p, as in the previous H_2O example). If the links are correlated then we need two parameters (p, f) per bond.

4 The three center bond

The need for multicenter bonding did not consolidate in Chemistry until it became clear that new rules were necessary to understand the chemical bond in boranes [31–33]. At present, three-center bonds are customarily classified within the MO paradigm into two- and four-electron categories. Here, we will show how multicenter bonding is easy to characterize from the statistical point of view. Much as two-center bonding senses two-center electron delocalization (i.e. a mutual fluctuation of the electron population of the two centers that can be quantified through the covariance), we define a multicenter bonding interaction by a mutual n-center fluctuation of the electron populations. In other words, much as when the populations of two atoms are statistically independent we say that they are not bonded, whenever we can partition the $n(>2)$ atoms that are being considered into two (or more) subsets with independent electron populations we say that there is no n-center bonding.

Measures to quantify multivariate influence are well studied in statistics. An i-th order cumulant [34], $\kappa_i(n_1,\ldots,n_i)$ is *a combination of central moments that measures how far a set of random variables are from mutual independence.* If κ_i vanishes the set of i variables can be partitioned into two or more subsets that are statistically independent. We take a non-vanishing κ_n as a definition of n-center bonding. Among the many properties of cumulants, it is particularly important for our purpose to notice that cumulants are additive. The cumulant of a distribution which can be separated into a set of independent components is just the sum of its contributions. Expressions for the n-th order cumulants can be found elsewhere [15]. Second and third order κ's coincide with the multivariate covariances: $\kappa_2(n_1,n_2)=\mathrm{cov}(n_1,n_2)\equiv \mathrm{cov}_{12}$, $\kappa_3(n_1,n_2,n_3)=\mathrm{cov}_3(n_1,n_2,n_3)$ $=\langle(n_1-\bar{n}_1)(n_2-\bar{n}_2)(n_3-\bar{n}_3)\rangle \equiv \mathrm{cov}_{123}$. However, this is not the case for further order cumulants. For instance, $\kappa_4=\mathrm{cov}_4-\mathrm{cov}_{12}\mathrm{cov}_{34}-\mathrm{cov}_{13}\mathrm{cov}_{24}-\mathrm{cov}_{14}\mathrm{cov}_{23}$. It can be shown [15] that atom-projected cumulants can be obtained from cumulant densities. The latter depend on general-order reduced density matrices.

4.1 The 3c,1e bond

Let us proceed analogously to what was done in the two-center case. A simple example would be the H_3^{2+} system, which we consider in an equilateral triangle configuration. Using the $1s$ manifold, the totally symmetric orbital in the ZDO approximation is $\chi=(1s_a+1s_b+1s_c)/\sqrt{3}$. The distribution function provides $p(1,0,0)=p(0,1,0)=p(0,0,1)=1/3$, and $\bar{n}_a=\bar{n}_b=\bar{n}_c=1/3$. It is easily verified that there is two-center delocalization between each of the three pairs of atoms: $\delta(a,b)=-2\mathrm{cov}(n_a,n_b)=2/9$.

From the standard MO point of view, an electron is delocalized over the three centers, the orbital describing it having equal contributions on each of them. From the statistical point of view, the electron can be equiprobably located in each center, and there is a mutual fluctuation in the electron population. The triple covariance ($\mathrm{cov}_{abc}=\kappa_3$) is immediately computed from $\kappa_3=$

$\sum_{n_a,n_b,n_c} p(n_a,n_b,n_c)(n_a-\bar{n}_a)(n_b-\bar{n}_b)(n_c-\bar{n}_c)$ as $\kappa_3 = 3\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)\left(\frac{-1}{3}\right)\left(\frac{-1}{3}\right) = 2/27$. As before, we define this EDF as that *defining a pure or perfect 3c,1e bond*.

It is customary to define an $n-$center bond order δ_n as $\delta_n \equiv \delta(a,b,c) = (-1)^{n+1} n\kappa_n$, that coincides with the overall many center projection of the n-th order cumulant density [15]. In this case, $\delta(a,b,c) = 2/9$. Since the specialized literature is plagued with different normalization conventions for these multicenter delocalization indices, we will try to stick to the cumulants or clearly specify to which quantities we refer to.

Electronegativity differences among the three centers are introduced as above, with $\chi = \lambda 1s_a + \mu 1s_b + \nu 1s_c$ and $\lambda^2+\mu^2+\nu^2 = 1$ at the ZDO level. The average electron populations in each center are given by the squares of the coefficients, and $p(1,0,0) = \lambda^2$, $p(0,1,0) = \mu^2$, $p(0,0,1) = \nu^2$. From this, it can be easily obtained that $\delta(a,b) = 2\lambda^2\mu^2$, just like in the $H_2{}^+$case (with equivalent values for the other two pairs), and that $\kappa_3 = \mathrm{cov}(n_a,n_b,n_c) = 2\lambda^2\mu^2\nu^2$. This shows that the maximum value for κ_3 is again obtained in the symmetric, pure or perfect bond ($\lambda^2 = \mu^2 = \nu^2 = 1/3$). Exactly as before, polarity decreases the three-center electronic delocalization, $\kappa_3 \in [0, 2/27]$ for a 3c,1e bond. Notice that κ_3 vanishes whenever any of λ, μ, or ν is zero. Within this model, for a three center bond to exist, it is necessary that the electron can be found in any of the three centers, as expected.

4.2 The 3c,2e bond

At the SD level, adding an opposite spin electron to form the $|\chi\bar{\chi}|$ determinant of $H_3{}^+$ is equivalent to adding an independent spin channel, so that $\delta(a,b) = 4/9$ and $\delta(a,b,c) = 4/9$. The EDF is easily obtained by direct product, and, if polarity is allowed, $\delta(a,b) = 4\lambda^2\mu^2$ and $\kappa_3 = 4\lambda^2\mu^2\nu^2$. It is important to stress that, with two independent opposite spin electrons, κ_3 and all the three two-center δ's are positive, but that this needs not be the case if correlation is allowed.

The existence of a negative three-center index was put forward early in Quantum Chemical Topology (QCT) [35]. It was proposed that 3c,2e bonds, like that in $H_3{}^+$, were characterized by positive κ_3 or δ_3 values, while 3c,4e bonds, like the standard Pimentel description of $F_3{}^-$, provides negative δ_3's. In the present context, the sign of the several multicenter indices reflects simply the possible types of fluctuation with respect to the average population of the centers. Since the total number of electrons is fixed, in the 2c case $\delta_2 = -2\mathrm{cov}_2 > 0$, although adding a third center that may act as an electron reservoir allows for negative δ_2 values. We will return to this. Similarly, three centers may host $+--$ or $++-$ fluctuations with respect to the average, i.e. fluctuations in which one of the centers gains population and the two others lose it ($+--$), or in which two centers gain at the expense of the third ($++-$). These two classes are associated to positive/negative $\kappa_3 = \mathrm{cov}_3$, respectively. At the SD ZDO level with one three-center delocalized orbital, only the $\kappa_3 > 0$ case is allowed, in agreement with the 3c,2e classification ($2\lambda^2\mu^2\nu^2$ can not be negative). At

least two orbitals that interfere are needed so that the possibility $\kappa_3 < 0$ can occur, again in agreement with the 3c,4e class. However, as soon as correlation is allowed, the classification smoothes out, as we will see.

Abandoning the ZDO approximation and introducing electron correlation opens a vast new territory in this case. To model it, we can notice that the number of components of the (spinless) EDF in the nc,2e case is made up from the n $p(2,0,\dots), p(0,2,\dots), \dots$ probabilities and the $n(n-1)/2$ $p(1,1,0,\dots), p(1,0,1,\dots), \dots$ components in which the two electrons lie in different centers. The total number of different probabilities is thus $n(n+1)/2$, so we need $n(n+1)/2-1$ independent parameters to span the full 3c,2e space.

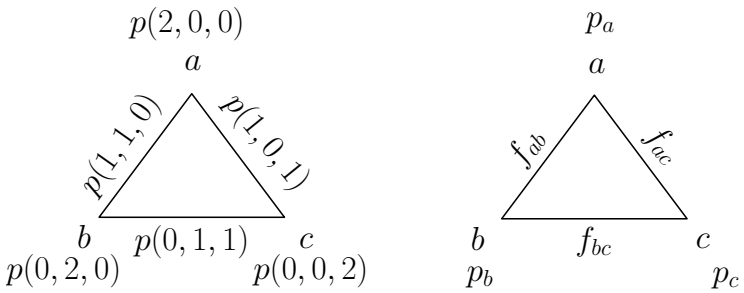
We can easily build a general model along the same lines used in the 2c,2e discussion. We first consider two independent indistinguishable electrons, letting $p_a \equiv p(1,0,0)$, $p_b \equiv p(0,1,0,$ and $p_c \equiv p(0,0,1)$ (with $p_a + p_b + p_c = 1$) be the probabilities that any of the them be found in each of the a, b, c centers. This provides the following two-electron EDF:

$$\begin{array}{ll} p(2,0,0) = p_a^2, & p(1,1,0) = 2p_a p_b, \\ p(0,2,0) = p_b^2, & p(1,0,1) = 2p_a p_c, \\ p(0,0,2) = p_c^2, & p(0,1,1) = 2p_b p_c. \end{array}$$

We now add a correlation factor for each pair of centers, f_{ab}, f_{ac}, f_{bc}, much in the light of what we did in the two center case. Any two of the set (p_a,p_b,p_c) plus the three correlation factors forms a set of five independent parameters. This may be immediately generalized to the nc,2e case. The correlated probabilities are,

$$\begin{array}{ll} p(2,0,0) = p_a^2 - p_a p_b f_{ab} - p_a p_c f_{ac}, & p(1,1,0) = 2p_a p_b(1 + f_{ab}), \\ p(0,2,0) = p_b^2 - p_b p_a f_{ab} - p_b p_c f_{bc}, & p(1,0,1) = 2p_a p_c(1 + f_{ac}), \\ p(0,0,2) = p_c^2 - p_c p_a f_{ac} - p_c p_b f_{bc}, & p(0,1,1) = 2p_b p_c(1 + f_{bc}). \end{array}$$

Whatever the values of the correlation factors, these add correctly to one. A pictorial representation of the 3c,2e EDFs and their modelling is sketched in Scheme 6



Scheme 6: The EDF and model parameters of the 3c,2e bond.

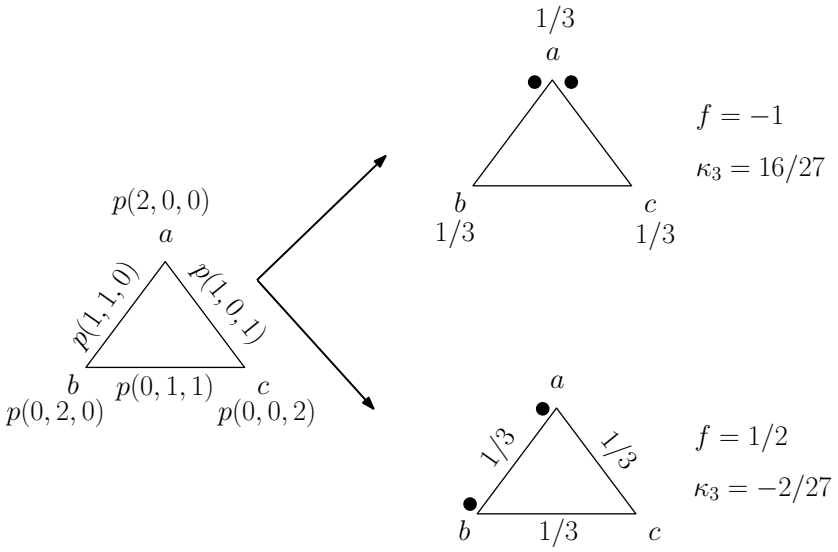
It can be checked that in this model the average center populations are not altered by the correlation factors, staying the same as in the independent electron case: $\bar{n}_a = 2p_a$, $\bar{n}_b = 2p_b$, $\bar{n}_c = 2p_c$. This is convenient to show the possible effects of several types of correlation. One can also find simple expressions for the delocalization measures:

$$\delta(a,b) = -2\mathrm{cov}(n_a, n_b) = 4p_a p_b(1 - f_{ab}) \text{ etc,}$$
$$\delta(a,b,c) = 3\mathrm{cov}(n_a, n_b, n_c) = 12p_a p_b p_c(1 - f_{ab} - f_{bc} - f_{ac}).$$

The correlation factors are bounded by below, $f_{ij} \geq -1$, but several situations allow for very large positive f's, as we will show.

Particularly interesting is the consideration of the symmetric non-polar $p_a = p_b = p_c = 1/3$, $f_{ab} = f_{bc} = f_{ab} = f$ situations that might describe H_3^+ and similar systems. The independent electron model (all f's equal to zero) gives $p(2,0,0) = 1/9$ and $p(1,1,0) = 2/9$ plus permutations. The general solution is immediate: $p(2,0,0) = \frac{1}{9}(1-2f)$, $p(1,1,0) = \frac{2}{9}(1+f)$. To guarantee that the probabilities stay between zero and one, it is necessary that $-1 \leq f \leq 1/2$. The positive limit is easily seen to be $1/(n-1)$ for an n-center case. Similarly, all δ_2's are equal to $\delta_2 = \frac{4}{9}(1-f)$, and $3\kappa_3 = \delta_3 = \frac{4}{9}(1-3f)$.

While δ_2 is always positive, meaning that the two-center fluctuation in a symmetric 3c,2e interaction is always like that in standard 2c,2e bonds, κ_3 is positive if $f < 1/3$ (which includes the non-correlated case) but changes sign if $1/3 < f \leq 1/2$. As commented above, we do not need a 3c,4e interaction to observe negative three-center indices. Electron correlation is enough.



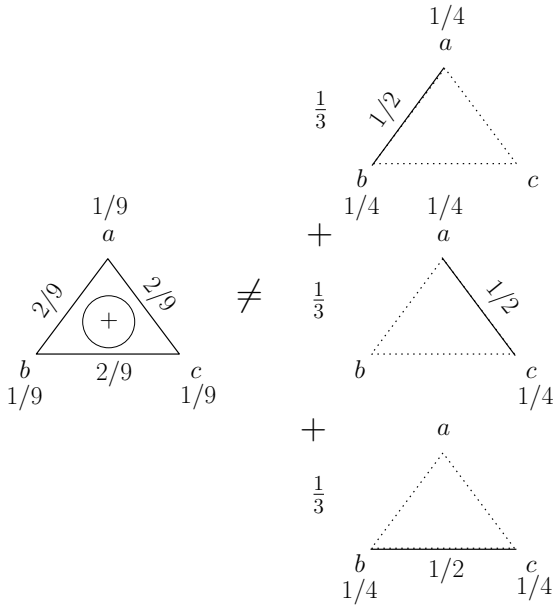
Scheme 7: Extreme cases of the symmetric 3c,2e bond

Scheme 7 shows the EDFs of the two extreme cases. When $f = -1$, δ_3

peaks at the maximum positive value (much like in the ionic resonance in H_2). Only the $(2,0,0)$-like structures are equally populated. It is rather clear that we only have $+--$ fluctuations. In the $f = 1/2$ case it is the three $(1,1,0)$-like structures that resonate, and again, it is clear that all fluctuations are of the $++-$ type: *it is the fluctuation type that determines the sign of the three-center delocalization index*.

The $f = 1/2$ case is interesting, since it describes the triangular entangled dissociation singlet of H_3^+. Its δ_2 does not vanish, becoming equal to $2/9$. This is a similar situation to that examined in H_2^+. At infinite distance we have a frustrated system with unavoidable spatial delocalization and three equiprobable distributions.

If we do not consider the spin structure, the delocalization pattern of the $f = 1/2$ case can be understood not as two particles that delocalize, but as a hole that can occupy the three centers equiprobably. Notice that 3c,1e model displayed $\delta_2 = 2/9$, $\kappa_3 = 2/27$. This particle-hole symmetry, very familiar in the physical literature, will be analyzed below. Sometimes the picture of delocalized holes is easier to grasp than that of delocalized particles. In the present case, the delocalization pattern of two heavily correlated electrons matches that of an independent hole.

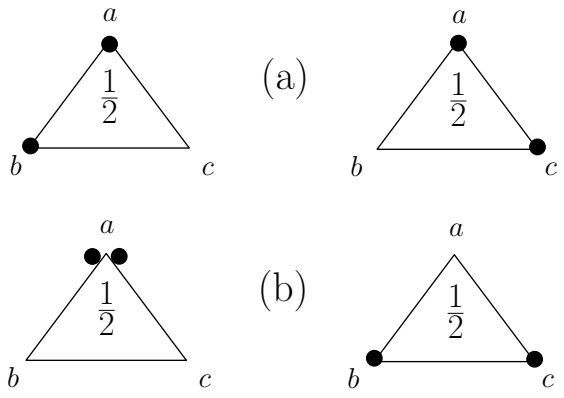


Scheme 8: Multicenter bonding is more than resonance. The H_3^+ case.

Another point that deserves comment regards the inadequacy of some traditional chemical images to understand multicenter bonding. It is customary to understand the 3c,2e bond in H_3^+ as a resonance of three normal cova-

lent bonds, as shown in Scheme 8. In the simple $f = 0$ case, it is immediate to show that the 3c,2e EDF is not the weighted sum of three 2c,2e bonds. *There is more in multicenter bonding than in plain resonance.* For instance, using Scheme 8 as a guide, the weighted resonance would provide $p(2,0,0) = 1/3 \times 1/4 + 1/3 \times 1/4 + 1/3 \times 0 = 1/6 \neq 1/9$. Not imposing an independent particle behavior, one can indeed write the EDF of H_3^+ as a weighted average of a correlated 2c,2e link with $f = 1/3$. Again, this tells about how chemistry allows for different, equivalent interpretations from a single dataset.

We end this subsection by considering the role of different correlation types on the final EDFs of asymmetric systems. As shown above, only through asymmetry may a $\delta_2 < 0$ be observed in this model. We take the very simplest isosceles asymmetry in which $p_a = 1/2$, $p_b = p_c = 1/4$. This leads to $\bar{n}_a = 1$, $\bar{n}_b = \bar{n}_c = 1/2$. We impose also $f_{ab} = f_{ac} = f$, $f_{bc} = f'$. The general EDF depends on f, f' and is easily worked through. We will restrict here to two simple possibilities: (a) $f = 1, f' = -1$; (b) $f = -1, f' = 3$. The EDFs are sketched in Scheme 9.



Scheme 9: Asymmetric 3c,2e models described in the text.

The correlation between the two electrons is clearly different. In (a) the two distributions contribute to κ_3 with a zero value, since the population of center a does not fluctuate. It can readily be checked that not only the triple covariance vanishes, but that $\delta(a,b) = \delta(a,c) = 0$. What is left is a non-vanishing $\delta(b,c) = 1/2$, typical of the delocalization of a H_2^+ system. In fact, the EDF is the product of a 1e EDF with a fully localized electron in a and an entangled bc H_2^+ system. Being formed by two independent subsystems, κ_3 vanishes.

The situation in (b) is more interesting. The three centers have non-vanishing population fluctuations, but two of them (b,c) are synchronous. This causes the first fluctuation $(+--)$ to be equal and of opposite sign to the second one $(-++)$ so that $\kappa_3 = 0$ at the end. All the 2c delocalizations are non-vanishing: $\delta(a,b) = \delta(a,c) = 1$, and $\delta(b,c) = -1/2$. This is rather simply interpreted. The

a center acts as an electron buffer. As the b, c pair is regarded, when the population of the first center increases, so does that of the second, and viceversa. This, of course, at the expense of center a. Weird at a first look, this behavior cannot be ignored and, in our opinion, has not received any kind of attention in the theory of chemical bonding. Might real chemical systems show this kind of synchronous population oscillations and negative δ_2's? As we will show below, the answer is yes.

4.3 The Pimentel-Rundle 3c,4e bond

We will now comment on the 3c,4e bond, particularized to the Pimentel-Rundle [36, 37] model. Although this is not the general case that can be worked out using the strategies defined above, it paves the way to the contents of the next Section. We build a three-center configuration with open b, c ends, that may be understood as linear, if necessary. We use the orbitals spanned by three $1s$ functions (or three $2p_z$ primitives in the F_3^- system) under the ZDO approximation so that

$$\begin{array}{l} \chi_1 = \frac{1}{2}(1s_b + \sqrt{2}1s_a + 1s_c), \\ \chi_2 = \frac{1}{\sqrt{2}}(1s_b - 1sc), \\ \chi_3 = \frac{1}{2}(1s_b - \sqrt{2}1s_a + 1s_c). \end{array}$$

Building a four electron determinant from the lowest lying levels, $\Psi = |\chi_1\bar{\chi}_1\chi_2\bar{\chi}_2|$, the probabilities can be easily obtained as the direct product of two electron same spin p^α or p^β components. The latter are obtained from 2×2 determinants using the prescriptions used above. For instance,

$$p^\alpha(0,2,0) = \left| \begin{array}{cc} a & a \\ \hline 1/4 & 1/2\times 1/\sqrt{2} \\ 1/2\times 1/\sqrt{2} & 1/2 \\ \hline \chi_1 & \chi_2 \end{array} \right| = 0.$$

The only non-vanishing contributions to the alpha spin EDF are $p^\alpha(1,1,0) = p^\alpha(1,0,1) = 1/4$, $p^\alpha(0,1,1) = 1/2$. From our present point of view, the alpha distribution is a correlated 3c,2e EDF, that we know how to systematize by using the expressions derived in Subsection. 4.2. For instance, since $\bar{n}_a^\alpha = 1/2$, $\bar{n}_b^\alpha = \bar{n}_c^\alpha = 3/4$, then $p_a^\alpha = \bar{n}_a^\alpha/2 = 1/4$. Similarly, $p_b^\alpha = p_c^\alpha = 3/8$. This provides $f_{ab}^\alpha = f_{ac}^\alpha = 1/3$, $f_{bc}^\alpha = 7/9$. As it can be checked, all fluctuations are of the $++-$ type, for in all the $(1,1,0)$, $(1,0,1)$, and $(0,1,1)$ structures two centers bear larger electron population than the average, and the remaining one hosts no electrons. This leads to $\delta^{\alpha\alpha}(a,b,c) = -1/4$. We do not need four electrons to obtain a negative DI. Adding the beta spin block, the total $\delta(a,b,c) = -1/2$. It is thus the correlation between the two $\sigma\sigma$ electrons in each spin channel that generates a negative three-centered index, and not the four electron nature of the links. In agreement with all our previous discussions, it is not possible to get a negative κ_3 with a SD formed from one orbital, but it is certainly possible in the case that the SD (in a triplet state) be constructed from two independent functions.

More insight about the particle-hole symmetry that was briefly introduced above is now at hand. Forming a fully unpaired quartet determinant with three alpha electrons $\Psi = |\chi_1\chi_2\chi_3|$, i.e. closing the orbital manifold expanded by the three $1s$ orbitals, provides only one surviving probability, $p^\alpha(1,1,1) = 1$. All delocalization is blocked. The original Pimentel's alpha $\Psi = |\chi_1\chi_2|$ is then equivalent to opening a hole in the "alpha closed-shell". The 3c,1e state χ_3 is a hole state from Pimentel's Ψ. Its electron populations are $\bar{n}_a^\alpha = 1/2$, $\bar{n}_b^\alpha = \bar{n}_c^\alpha = 1/4$. If interpreted as a positronic function, these are the positive charges induced by χ_3 on the filled sea of three alpha electrons. Being a one-particle state, its $\kappa_3 > 0$. It is immediately computed from the $2\lambda^2\mu^2\nu^2$ rule as exactly minus Pimentel's κ_3. As we will see, the sign change depends on the odd/even character of the order of the cumulant. The behavior of a three-electron state, that cannot be rationalized without Fermi correlation, is again immediately uncovered from that of an independent hole if we change from the particle to the hole description.

5 The equivalent primitives manifold

In this Section we will show how to simplify the multicenter indices calculations in a simple but representative case: a single determinant built from orbitals which are linear combinations of a manifold of equivalent primitives in the ZDO approximation (e.g., the n orthonormal orbitals that can be built from n equivalent $1s$, $2p_z$, etc, functions centered at n different centers). Most simple examples of conjugation in unsaturated hydrocarbons (including aromatic systems) are studied within these constraints: a set of equivalent p_z functions that are used to build the π set of delocalized one-electron states.

To continue, we need to borrow a result from the framework that links the fluctuation of the electron populations with the cumulant densities of wavefunction theory [15]. It is known that at the SD level the n-center delocalization index can be written as a symmetrized trace of products of atom-projected overlap integrals. For a closed-shell with $N/2$ molecular orbitals

$$\delta(a,b,\ldots,n) = 2\sum_{\mathcal{S}_n} \mathrm{Tr}(S^a S^b \ldots S^n), \tag{6}$$

where the sum runs over the $n!$ permutations of the set $a,b,\ldots,n$ and all the S's are $(N/2)\times(N/2)$ matrices.

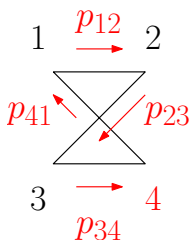
Now we impose that each of the $N/2$ populated orbitals χ_i is a ZDO linear combination of our manifold primitives: $\chi_i = \sum_j c_i^j \phi^j$, where superindices refer to the atomic centers $a,b,\ldots,n$. It is now easy to rearrange the sum in Eq. 6 given that $S_{ij}^a = c_i^a c_j^a$. As an example, in the case of $n = 2$, one of the components of $\delta(a,b)$ is

$$\mathrm{Tr}(S^a S^b) = \sum_{ij} S_{ij}^a S_{ji}^b = \sum_{ij} c_i^a c_j^a c_j^b c_i^b = (\sum_i c_i^a c_i^b)(\sum_i c_i^b c_i^a).$$

Defining $p_{ab} = \sum_i c_i^a c_i^b$, which may be interpreted as a kind of ab transition amplitude, the result is easy to generalize, so that

$$\delta(a, b, \ldots, n) = 2 \sum_{\mathcal{S}_n} p_{ab} p_{bc} \ldots p_{na}.$$

The sum in this expression runs over all the possible permutations of the $a, b, \ldots n$ set of centers for which a closed circuit of amplitudes $p_{ab} \ldots p_{na}$ (where each ab edge is visited only once) is constructed. In this sense, the formalism binds to graph theory, where those closed paths are known as Hamiltonian cycles. Multicenter delocalization is seen within this approach as the result of at least one closed circuit for which the product of transition amplitudes does not vanish. An example with the $p_{12}p_{23}p_{34}p_{41}$ cycle is shown in Scheme 10.



Scheme 10: A Hamiltonian circuit in a four center system

Either in terms of S^a's or p_{ab}'s, the multicenter index results from orbital interference. Returning to the triplet state of H_2 or to He_2, $\delta_{ab} = 2p_{ab}p_{ba}$, and its vanishing is due to the destructive interference of the σ_g and σ_u components: $p_{ab} = c_g^a c_g^b + c_u^a c_u^b = 1/2 - 1/2 = 0$. Notice that, again, if only the g or the u channel is populated the delocalization is different from zero.

This can be immediately generalized. Since the manifold of n orbitals are orthonormal, one has $\langle\chi|\chi\rangle = C\langle\phi|\phi\rangle C^\dagger = I$, where I is the unit matrix, $|\chi\rangle \equiv (|\chi_1\rangle \ldots |\chi_n\rangle)$, and $|\phi\rangle \equiv (|\phi_1\rangle \ldots |\phi_n\rangle)$. However, in the ZDO approximation, the identity $\langle\phi|\phi\rangle = I$ holds. Hence, it follows that C is unitary. If all the n orbitals of one of the manifold are occupied, then p_{ab} is the scalar product of two columns of C, and vanishes for any pair ab. A full manifold implies zero delocalization indices of all orders.

This is also the basis for the hole-particle symmetry already presented. If all but m orbitals $(n-m)$ are fully occupied, then $p_{ab}(n-m) = \sum_i^{(n-m)} c_i^a c_i^b = -\sum_{i=(n-m+1)}^{m} c_i^a c_i^b = -p_{ab}(m)$. All delocalization measures of the $2(n-m)$ electron system can be obtained from a negative transition amplitude computed from the $2m$ unoccupied holes. Notice that the multicenter index will change sign if the number of centers is odd. This is exactly the behavior that was found on examining the 3c,4e Pimentel-Rundle model.

Now we consider an nc,2e cyclic system formed by one totally symmetric orbital $\chi = 1\sqrt{n}\sum_a \phi_a$. There are $n!$ different Hamiltonian cycles to consider

with all amplitudes $p_{ab} = 1/n$. Then $\delta(a, b, \ldots, n) = 2n!/n^n$. For large n we can use Stirling's approximation $\ln(n!) \approx n\ln(n) - n$ to show that $\ln(\delta_n) \approx -n$, and that the multicenter index decays exponentially. This is the result of the product rule of probability theory, and has usually been put forward as a drawback of these and similar formalisms. If this is a concern, the n-th root of δ_n or κ_n, $\delta_n^{1/n}$ or $\kappa_n^{1/n}$, scales like $\mathcal{O}(n^0)$, at the expense of losing, for instance, the extensivity property.

It is also relevant to point out that for a general cyclic system with equivalent nodes, p_{ab} will not depend on the particular ab pair that is chosen. All the multicenter indices are determined in this case by a single parameter p_{ab}.

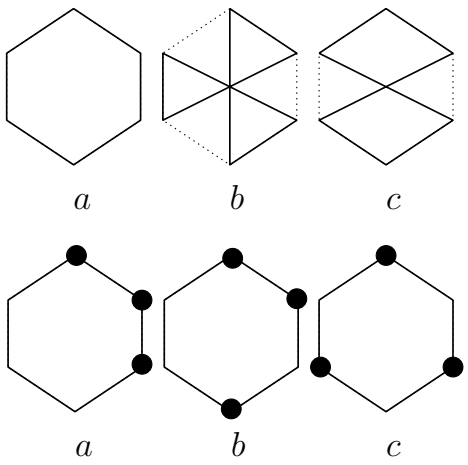
Let us work out these ideas on the π skeleton of benzene, modeled from six equivalent $2p_z$ orbitals. A possible description of the manifold is

$$\begin{array}{ll}
\chi_1 = \frac{1}{\sqrt{6}}(\phi_1 + \phi_2 + \phi_3 + \phi_4 + \phi_5 + \phi_6), & \\
\chi_2 = \frac{1}{\sqrt{12}}(\phi_1 + 2\phi_2 + \phi_3 - \phi_4 - 2\phi_5 - \phi_6), & \chi_3 = \frac{1}{2}(\phi_1 - \phi_3 - \phi_4 + \phi_6), \\
\chi_4 = \frac{1}{\sqrt{12}}(\phi_1 - 2\phi_2 + \phi_3 + \phi_4 - 2\phi_5 - \phi_6), & \chi_5 = \frac{1}{2}(\phi_1 - \phi_3 + \phi_4 - \phi_6), \\
\chi_6 = \frac{1}{\sqrt{6}}(\phi_1 - \phi_2 + \phi_3 - \phi_4 + \phi_5 - \phi_6). &
\end{array}$$

Here, χ_2, χ_3, and χ_4, χ_5 are double degenerate e orbital pairs, and the ground state is the determinant formed by doubly populating the χ_1, χ_2, χ_3 orbitals. It is immediate to show that the ortho, meta, and para transition amplitudes are $p_{12} = 1/3$, $p_{13} = 0$, $p_{14} = -1/6$. The null meta amplitude justifies the vanishing of the meta delocalization index, which is a well known property behind the experimental substitution pattern in benzene. The patterns of two-center delocalization in alternant hydrocarbons have been studied before [38, 39], and the decay rate of these has been related to electrical conductivity by recourse to the modern theory of insulators [39, 40]. For the time being, it is clear that only Hamiltonian cycles containing exclusively ortho and para steps will contribute to the six-center delocalization index, δ_6. There are only three types of these that we label as a, b, anc c in the top part of Scheme 11, which contain zero, three, or two para links, respectively.

There are 12, 24, and 36 equivalent permutations of type a, b, and c, respectively (we can start each cycle on any of the six centers and traverse it clock- or anticlock-wise). Each of the cycles of type a contributes with $p_{12}^6 = 1/729$, the b type of channels contributes with $p_{12}^3 p_{14}^3 = -1/5832$ each, and each cycle of type c with $p_{12}^4 p_{14}^2 = 1/2916$. Adding everything together, $\delta_6 = 4/81$. In standard chemical language, the a circuits are to be associated to Kekulé resonance structures, while b, c to the Dewar ones. As it can be checked, the first contribute double than the second to the 6-center index. Notice, incidentally, that the ortho two-center DI is immediately obtained to be $4 \times (1/3)^2 = 4/9$, very close to the standard $1/2$ value usually accepted for the π contribution to the bond order in benzene, and that the para DI, which has been successfully used as a measure of aromaticity [14], is equal to $1/9$.

With the help of the orbitals χ_1, χ_2, and χ_3 sketched above it is also easy to write down the six determinants involved in Eq. 2 that are needed to obtain the three non-equivalent alpha probability distributions shown in the bottom



Scheme 11: Non-vanishing Hamiltonian circuits (top), and non-equivalent alpha electron distributions (bottom) in benzene.

part of Scheme 11. The a type has $p(1,1,1,0,0,0) = 1/72$, while the b and c types display probabilities equal to $1/18$ and $1/8$, respectively. There are $6, 12$, and 2 equivalent resonance structures belonging to each type. On constructing the direct product with the beta distribution to build the full EDF it is clear that the largest probability spin-resolved resonance structure is that with the fully paired arrangement $(\uparrow,\downarrow,\uparrow,\downarrow,\uparrow,\downarrow)$, that shares the same probability as the $(0,2,0,2,0,2)$ and $(2,0,2,0,2,0)$ ones.

5.1 All the states from an electron configuration

We close this section by analyzing how the present approach may shed new light on the role played by bonding indices in chemical bonding theory. We choose the square planar D_4 configuration of H_4, and build a set of four orbitals from the $1s$ manifold. Our presentation is also valid for the π skeleton of cyclobutadiene. In the ZDO approximation,

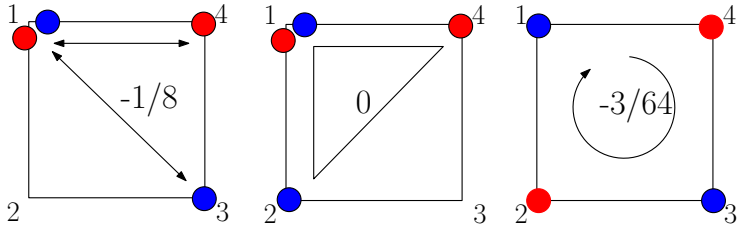
$$\begin{array}{l} \chi_{a_1} = \frac{1}{2}(\phi_1 + \phi_2 + \phi_3 + \phi_4), \\ \chi_{e_1} = \frac{1}{2}(\phi_1 + \phi_2 - \phi_3 - \phi_4), \quad \chi_{e_2} = \frac{1}{2}(\phi_1 - \phi_3 - \phi_4 + \phi_6), \\ \chi_{b_1} = \frac{1}{2}(\phi_1 - \phi_2 + \phi_3 - \phi_4). \end{array}$$

We examine the states coming from the $a_1^2e^2$ configuration. These include a triplet and three singlets. The $M_S = 0$ components are,

$$\begin{array}{l} {}^3A_2 = 1/\sqrt{2}(|\chi_{a_1}\bar{\chi}_{a_1}\chi_{e_1}\bar{\chi}_{e_2}| - |\chi_{a_1}\bar{\chi}_{a_1}\chi_{e_2}\bar{\chi}_{e_1}|), \\ {}^1A_1 = 1/\sqrt{2}(|\chi_{a_1}\bar{\chi}_{a_1}\chi_{e_1}\bar{\chi}_{e_1}| + |\chi_{a_1}\bar{\chi}_{a_1}\chi_{e_2}\bar{\chi}_{e_2}|), \\ {}^1B_1 = 1/\sqrt{2}(|\chi_{a_1}\bar{\chi}_{a_1}\chi_{e_1}\bar{\chi}_{e_2}| + |\chi_{a_1}\bar{\chi}_{a_1}\chi_{e_2}\bar{\chi}_{e_1}|), \\ {}^1B_2 = 1/\sqrt{2}(|\chi_{a_1}\bar{\chi}_{a_1}\chi_{e_1}\bar{\chi}_{e_1}| - |\chi_{a_1}\bar{\chi}_{a_1}\chi_{e_2}\bar{\chi}_{e_2}|), \end{array}$$

and differ in the open-shell e^2 structure. None of them are single determinants, but it can be shown [20] that the EDF in these simple multideterminant cases can also be obtained from back of the envelope calculations. If $\Psi = \sum_r c_r D_r$, where D_r is a Slater determinant, then $p(S) = \mathcal{N} \sum_{rs} c_r c_s \sum_{\{k_j\} \in \mathcal{S}_N} \det \left[S^{b(k_j)}_{i(r)j(s)} \right]$, where now the overlap integrals are between orbital i in determinant r, and orbital j in determinant s. This is a slightly more cumbersome to work out than in our previous examples, but it can still be done by hand.

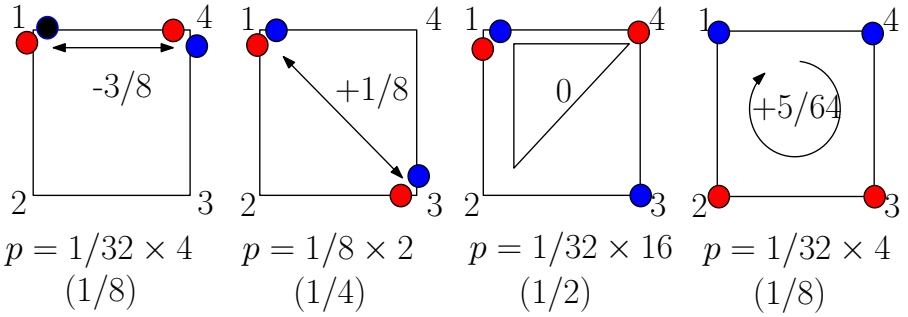
It is not difficult to show that the four states display the same p^α or p^β distribution. Ordering the centers anticlockwise, $p^\alpha(1,1,0,0) = 1/8$, $p^\alpha = (1,0,1,0) = 1/4$. There are four and two equivalent structures of the first and second kind, respectively. Since we do not have a single determinant description, the full EDF $p \neq p^\alpha \otimes p^\beta$.



Scheme 12: ${}^3A_2$ full EDF for a planar D_4 system. The κ_2, κ_3, and κ_4 values are shown as numbers within the squares. The probability of each arrangement as well as the multiplicity of each structure are shown below, together with the total contribution of all structures of each kind (in parenthesis).

Scheme 12 shows the EDF for the triplet state. Notice that the central and right structures can be reached from the leftmost one by a vicinal shift (jump) of one electron. All bonding indices in this case can also be obtained from the $M_S = +1$ component of the triplet, which is the single determinant ${}^3A_2 = |\chi_{a_1} \bar{\chi}_{a_1} \chi_{e_1} \chi_{e_2}|$. Its delocalization pattern can be separated into an alpha part and a beta part. The first is equivalent to a hole delocalized in χ_{b_1}, and the second to an electron delocalized in χ_{a_1}. We can thus understand why κ_3 vanishes, since the hole and the electron κ_3 annihilate each other in the three center case. Both, however, add to the two- and four-center indices. Using our transition amplitudes rule, δ_2 and δ_4 are immediately shown to be equal to 1/4.

The ${}^1A_1$ EDF is shown in Scheme 13. In this state the four-center index has changed sign. There are no completely paired $(\uparrow,\downarrow,\uparrow,\downarrow)$ spin structures, and the value of the diagonal two-center delocalization index $\delta(1,3)$ is negative, $-1/4$. *This provides the first confirmation that negative delocalization indices may be found in real chemical systems.* We stress that the consequences of these and



Scheme 13: $^1A_1$ full EDF for a planar D_4 system. All notation as in Scheme 12.

other ideas contained in this work should be carefully checked in the future. The third structure counting from the right in Scheme 13 is equivalent to the first in Scheme 12, and again, we can reach all the others from it through one electron shifts. The probability of two unoccupied centers, vanishing in the triplet, is here different from zero.

Table 2: Classification of the $a_1^2e^2$ states by means of the set of κ_n values.

	κ_{12}	κ_{13}	κ_{123}	κ_{1234}
$^3A_{2g}$	-1/8	-1/8	0	-3/64
$^1B_{2g}$	-3/8	+1/8	0	-3/64
$^1A_{1g}$	-3/8	+1/8	0	+5/64
$^1B_{1g}$	-1/8	-1/8	0	+5/64

Another interesting point regards the classificatory power of the set of κ_n values. Table 2 shows that a specification of the different values of κ_2 and κ_4 fully identifies the state of the H_4 system. The generality of this statement should also be investigated. Finally, a real CAS[4,4] D_{4h} calculation with the GAMESS [41] code employing its TZV basis set in the $^1B_{2g}$ state provides an equilibrium H-H distance of 1.121 Å and the following cumulant values: $\kappa_{12} = -0.326(-0.375)$, $\kappa_{13} = +0.013(+0.125)$, $\kappa_{123} = -0.0084(0.000)$, $\kappa_{1234} = -0.018(-0.047)$. The figures in parenthesis are those provided by our ZDO modelling. This demonstrates that negative δ_2 are here to stay.

6 From EDFs to chemical bonding

We now move to consider the inverse problem. Given an EDF, computed from a (possibly general) wavefunction, we want to extract chemical information from it. Namely, from the discrete probability distribution providing the (spinless or spin-resolved) probabilities of finding a partition of the N electrons into the m

centers of a molecule, we would like to obtain the set of n-center independent bonds that *best* describe the EDF.

In a sense, this problem is similar to that of approximating optimally a discrete probability distribution by a dependence relationship among the variables upon which it depends. This is an important problem in statistics. In 1968, Chow and Liu [42, 43] used dependence trees and mutual information measures to offer a possible solution. Since then, it has become clear that the general problem is NP-hard (i.e. harder to solve than those that can be solved by a nondeterministic Turing machine in polynomial time), and many techniques, including Bayesian networks, Markov chains and recently, machine learning algorithms have been used to tackle it [44, 45]. It is far from our aim to discuss how these techniques can be applied here. We will simply show that in many cases the chemical information encoded in the EDFs can be easily extracted.

Accepting some constraints, the problem can be easily formalized. Imagine that we only contemplate a description in terms of two-electron bonds, which is usually a valid approximation in most cases. Let us also restrict to the spinless case for simplification. Then, as we already showed, we look for partitioning the EDF $\boldsymbol{p}$ as a product of $n = N/2$ nc,2e components: $\boldsymbol{p} = \bigotimes_i \boldsymbol{p}_n^i$, where $\boldsymbol{p}_n^i$ is an n-center probability distribution in which we must specify the set of n centers in which the nc,2e bond is delocalized, as well as their $n(n+1)/2$ probability components. In many systems only 2c,2e links will be relevant, so we will just need to find which pairs of centers hold 2c,2e links and what type of 2c,2e interaction each of these links will be (as specified by its $\boldsymbol{p}_2$).

Without recourse to advanced statistical techniques we can identify a number of independent steps in proposing a solution. We need to find, on the one hand, whether 2c,2e bonds are enough or whether we need multicenter (i.e. $n > 2$) bonding. On the other, we must identify the centers involved in each multi- (including two-) center link. We will first show that, indeed, the locality of interactions can be probed easily. Then we will examine how to grasp some information on the other two problems.

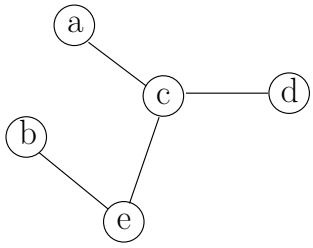
6.1 Checking statistical (in)dependence

The full spinless EDF must fulfill Bayes' chain rule. Let us, as usual, label the centers of a system $a, b, \ldots, m$. Bayes' rule allows us to write, in general, the joint probability distribution in terms of marginals and conditional probabilities,

$$p(n_a, \ldots, n_m) = p(n_a)p(n_b|n_a)p(n_c|n_a n_b)\ldots p(n_m|n_a \ldots n_{(m-1)}),$$

where $p(n_a)$ is the one-center marginal probability of finding a given number of electrons in center a, and $p(n_i|n_j n_k..)$ is the conditional probability of finding n_i electrons in center i provided that $n_j, n_k, \ldots$ electrons have been found in center $j, k, \ldots$. The conditional probability is computed by $p(n_i|n_j n_k..) = p(n_i, n_j, n_k, \ldots)/(p(n_j)p(n_k)\ldots)$.

If we now use our chemical intuition to assume a dependence structure for the center (or group) electron populations, as exemplified in Scheme 14, where



$$p = p(n_a)p(n_b)p(n_c|n_a)p(n_d|n_c)$$

Scheme 14: An example of Bayes' rule from chemical assumptions.

dashes are to be interpreted both as causal relations between populations and as *chemical bonds*, then we expect that, for instance, the populations of centers a and b (or those of d and c) be statistically independent. Thus, the chain

$$p = p(n_a)p(n_b|n_a)p(n_c|n_a n_b)p(n_d|n_a n_b n_c)p(n_e|n_a n_b n_c n_d)$$

can be significantly simplified since $p(n_c|n_a n_b) = p(n_c|n_a)$, $p(n_d|n_a n_b n_c) = p(n_d|n_c)$, and $p(n_e|n_a n_b n_c n_d) = 1$ due to the constraint that the total electron population is equal to N, so that if we know that n_a, n_b, n_c, n_d electrons are found in the a, b, c, d centers, then the population of e is fixed and the conditional propability is equal to one.

We can apply these ideas to a simple example. We will take the Hartree-Fock data in Table 1 regarding the H_2O (Ha-Ob-Hc) molecule. Using Bayes' rule, we approximate $p = p(n_a)p(n_c|n_a)p(n_b|n_a n_c) \approx p(n_a)p(n_c)$ if the two H atoms are not bonded and show independent populations. The H atom marginals (Ha is equivalent to Hc) can be easily obtained by summation. There are only three non-negligible populations with $p(0) = 0.6444$, $p(1) = 0.3130$, and $p(2) = 0.0414$. If our statistical dependence tree assumption is correct, then the full EDF contained in the Table can be reconstructed from just two parameters (the sum of the marginals should be one). The comparison can be found in Table 3.

As expected, the approximation works rather well, and allows us to draw a plausible chemical graph from plain statistics. The two H atoms are not bonded in a first approximation. We stress that in this approach there is no hypothesis whatsoever as to what kind of statistical dependence exists between the H and the O atoms. If two independent O-H bonds are superimposed, then we already showed that the number of parameters can be reduced to one λ asymmetry term.

A more systematic approach on how to draw the chemical graph can by devised through the use of the cumulant moments κ_n or mutual information measures lile I_2. The latter are entropy-like quantities that vanish in the case of independence of two random variables:

$$I_2(n_1, n_2) = \sum_{n_1, n_2} p(n_1, n_2) \ln \left(\frac{p(n_1, n_2)}{p(n_1)p(n_2)} \right).$$

Table 3: Exact QTAIM EDF for a HF//6-311G(d,p) calculation in H_2O together with that approximated by assuming independence of the Ha and Hb populations. Only inequivalent probabilities are shown.

n(Ha) a	n(O) b	n(Hc) c	p(exact)	p(approx)
0	10	0	0.4133	0.4153
1	9	0	0.2032	0.2017
1	8	1	0.0971	0.0980
2	8	0	0.0274	0.0267
1	7	2	0.0126	0.0130
2	6	2	0.0002	0.0017

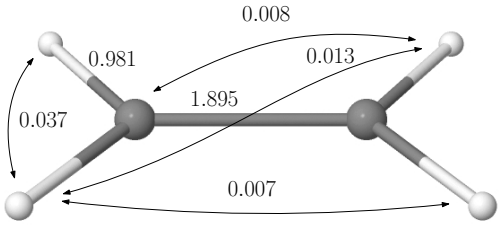
The mutual information can only be zero if the joint probabilities are reproduced by the product of the marginals, i.e. if the random variables are independent. Computing all the possible I_2 provides information about the degree of dependency among the set of variables. For instance, in our water case, $I(a,b) = I(b,c) = -0.3414$, and $I(a,c) = -0.00031$. This shows that, to a good approximation, the H atoms are not to be considered bonded. Dash chemical graphs can easily be drawn from I_2 tables.

6.2 Grasping the graph structure

It is our opinion that the non-additivity of the I_2 descriptors severely precludes their general use. Cumulants provide similar information and give direct chemical insight. A simple algorithmic recipe to identify the chemical graph has already been explored in terms of interatomic exchange-correlation energies [46], the energetic correlates of δ_2. Its equivalence in the present context is simple: *(i) obtain δ_2 values between all pairs of centers; (ii) cluster their values in groups; (iii) draw dashes for all pairs belonging to the largest δ_2 cluster.*

Scheme 15 shows how this is applied to the C_2H_4 case, with a QTAIM EDF computed from a GAMESS [41] HF optimization with the TZV(d,p) basis set. The DIs for the C-C, and C-H bonded interactions are 1.895 and 0.981, respectively. The C-H non-bonded DI is 0.008, and the non-bonded H-H DIs are 0.037 (geminal), 0.007 (cis), and 0.013 (trans). No statistical technique is needed to find manually two clusters of values. Actually, the values of the secondary cluster can be interpreted in terms of stereoelectronic effects, this being not important here.

The DI(C,C) = 1.895 value indicates that there are several electrons delocalizing between the two C atoms. A naïve identification of the bond order with DI would tell us that we have two 2c,2e bonds. However, we should be careful with this type of assignment, particularly in polar interactions. Simple EDF manipulations shed light on this. In the ethylene case, for instance, we can perform an EDF calculation if we divide the system into two CH_2 groups.



Scheme 15: Grasping the C_2H_4 graph structure from δ_2. See the text for details.

Table 4 shows the results.

Table 4: QTAIM EDF for a HF//TZV(d,p) calculation in C_2H_4. The system has been partitioned into either two equivalent CH_2 groups, A, B, or into a H and C_2H_3 components, C, D, respectively.

n_A	n_B	p	n_C	n_D	p
8	8	0.3650	1	15	0.4882
9	7	0.2476	0	16	0.2610
7	9	0.2476	2	14	0.2412
6	10	0.0687	3	13	0.0122
10	6	0.0687			
5	11	0.0031			
11	5	0.0031			

It is readily seen that to a good precision, the methylene groups are able to exchange two pairs of electrons, so their populations vary from 6 to 10. Similarly, the H atoms exchange basically a pair of electrons, their populations staying between 0 and 2. Moreover, in the symmetric CH_2-CH_2 partition, the probability distribution is almost binomial. The distribution corresponding to two pure 2c,2e bonds would provide, by direct product, probabilities equal to $6/16 = 0.375, 4/16 = 0.25$, and $1/16 = 0.0625$ for the $(8,8), (7,8)$, and $(6,10)$ structures, respectively. Similarly, the C-H link is seen to correspond rather well to a non-polar 2c,2e distribution. A simple strategy thus leads to the conclusion that the EDF can be very well represented by the graph shown in Scheme 15, with an almost non-polar double C-C bond and slightly polar C-H links. Both

the graph structure and the set of nc,2e bonds have been uncovered without any other help but the exam of the EDF statistics.

The present approach can be generalized, if necessary, to include nc,2e links. The simplest option is a top-down approach, identifying first the maximum number of centers implied in multicenter bonding and then descending the ladder down to 2c,2e links.

Let us put this to work in a clear three-center bonding case, the cyclopropenyl cation $C_3H_3{}^+$, calculated at the DFT/cc-pVTZ level. To simplify as much as possible, we can distinguish three CH interacting groups, which we label as a, b, and c. At the *pseudo*-SD level, we distinguish two independent 10 electron alpha or beta distributions. In a first approximation we can take for granted that two of them are localized in each CH group, corresponding to the carbon core and the C-H bond. This leaves 4 electrons to delocalize through the C_3 skeleton. If we suppose each σ C-C link to be pure, non-polar, and independent of the two others, it is easy to check by direct product that p^α for these three ideal non-polar bonds give rise to $p^\alpha(2,1,0) = 1/8$ (six equivalent of these exist), and $p^\alpha(1,1,1) = 2/8$. If we now add a pure 3c,1e alpha contribution, we arrive by direct product at (only inequivalent terms shown) $p^\alpha(3,1,0) = 1/24$, $p^\alpha(2,2,0) = 1/12$, $p^\alpha(2,1,1) = 1/6$.

Table 5: QTAIM alpha EDF for a DFT//cc-pVTZ calculation in $C_3H_3{}^+$. The system has been partitioned into three equivalent CH groups a, b, c.

n_a	n_b	n_c	p(exact)	p(approx)
4	3	3	0.1639	0.1667
4	4	2	0.0810	0.0833
5	3	2	0.0427	0.0416
6	2	2	0.0022	0.0000

It is rather clear that our approximation reproduces rather well the distribution, as shown in Table 5. From our present point of view, this structure would be recovered from computing first the cumulants. Separating again the system into the three a, b, c coupled CH groups, $\kappa_3(a,b,c) = 0.220$, $\delta(a,b) = 1.513$. This tells immediately about a large three-center component.

An unbiased analysis would start with the computation of the full spinless EDF of the system. Assuming that no four-center bonding exists, we may just obtain the several multicenter DIs among pairs and trios of atoms. For the sake of simplicity, let us denote the carbon atoms with capital letters A, B, C and their linked H counterparts with small letters, a, b, c, respectively. It turns out that the net charges of the C and H atoms are, respectively, 0.063 and 0.270 au. This implies a rather polar C-H link. The several two-center DIs are: $\delta(A,B) = 1.432$, $\delta(A,a) = 0.873$, $\delta(A,b) = 0.040$, $\delta(a,b) = 0.002$. Finally, the three-center DIs are $\delta(A,B,C) = 0.600$, $\delta(A,B,a) = 0.019$, $\delta(A,b,c) = 0.001$, $\delta(A,a,b) = 0.001$, $\delta(A,B,c) = 0.018$, and $\delta(a,b,c) = 0.000$.

From these data, we would reconstruct easily the standard chemical graph

with C-H and C-C bonds plus a very large three-center contribution. This last contribution can not only be due to a standard 3c,2e π bond, since the third order cumulant among the C atoms is larger than that of a pure 3c,2e link ($4/27 \approx 0.148$). We have to conclude that a small σ three-center term is also present in the system.

6.3 Partial EDFs

The complete (spinfull or spinless) EDF of a system contains a large amount of information about the statistical dependence or independence of atomic populations. However, as we have noticed, obtaining appropriate chemical models from them is not easy. Another simple approach to getting model parameters from the computed distributions is through the analysis of partial EDFs.

Let us illustrate this procedure with the H_2O example. Given the EDF of Table 3 we have shown that an immediate analysis of the two- and three-center DIs (δ(O,Ha) = 0.648, δ(Ha,Hc) = 0.008, δ(Ha,O,Hc) = 0.017) leads to consider this system in terms of two equivalent 2c,2e OH polar bonds. How we do obtain their q, f parameters (see Subsection 3.2) or, equivalently, what their 2c,2e local (partial) EDF is has not a unique answer. A general optimization solution will be offered in Section 7.

The simplest solution, that would be to consider the EDF provided by a Ha-OHc partition, leads to a biased vision, since it includes not only the Ha-O delocalization but also the Ha-Hb one, which in this case is very small. Moreover, this recipe cannot be applied to cycles. Anyway, doing so in H_2O gives rise to a δ(Ha,OHc) = 0.656 (the sum of the previous DIs) and to a two-electron EDF: $p(0, 10) = 0.645$, $p(1, 9) = 0.314$, and $p(2, 8) = 0.042$, which can of course be read in q, f coordinates easily.

As Scheme 5 shows, the small value of δ(Ha,Hc) implies that, statistically speaking, only a subset of the oxygen's electrons delocalize over each of the Ha, Hc atoms. Let us call them n^a_O and n^c_O. Now imagine that we consider only those structures in the total three-center EDF in which the population of the Hc atom is fixed. This *blocks* the O-Hc delocalization channel, in Bayesian networks parlance. All fluctuation of the O atom population is then due to the Ha-O delocalization, and its *partial EDF* can be easily obtained by renormalizing all the fixed n^c_O probabilities so that they add to one. In a sense, we take snapshots in which, given that the Hc population is fixed, so is n^c_O. This *blocked-channel* technique is very useful, but not unique. As δ_3 is small but not zero, the partial EDF depends slightly on the chosen n^c_O value. This can be used for chemical purposes to understand how a chemical bond depends on the electron distribution of its neighbors. A weighted average of the different distribution according to the probability of finding a given n^c_O may also be defined.

Using Table 3, we can select $n(\text{Hc}) = 0, 1, 2$. It is clear that for the sake of precision we should not trust much the results obtained from $n(\text{Hc}) = 2$, but that will not be a concern here. The renormalized EDF for the Ha-O interaction is obtained for each of these blocked structures easily. For instance, taking $n(\text{Hc}) = 0$, we have, as expected, three non-negligible structures, $(0, 10, 0), (1, 9, 0)$, and

$(2, 8, 0)$, which correspond to a pair of electrons delocalizing over the Ha-O pair, i.e. to the Ha-O 2c,2e link. Adding their probabilities, we recover $p(n(\mathrm{Hc}) = 0) = 0.644$, and renormalizing each of them by this value we get for the $n(\mathrm{Hc}) = 0$-blocked Ha-O EDF: $p(0, 10) = 0.6418$, $p(1, 9) = 0.3157$, and $p(2, 8) = 0.0425$. These have to be interpreted as the $(0, 2), (1, 1)$, and $(2, 0)$ components of a simple 2c,2e EDF.

Table 6: Partial QTAIM EDFs for a HF//6-311G(d,p) calculation in $H_2O \equiv$ Ha-O-Hc for the Ha-O pair. See the text for details.

n(Hc)	n(Ha)	n(O)	p
0	0	2	0.6418
	1	1	0.3157
	2	0	0.0425
1	0	2	0.6496
	1	1	0.3103
	2	0	0.0401
2	0	2	0.6601
	1	1	0.3028
	2	0	0.0371
avg.	0	2	0.6450
	1	1	0.3135
	2	0	0.0415

Table 6 contains the three different partial Ha-O distributions that we can get from Table 3. Notice that the Ha-O bond depends slightly on the number of electrons that have been blocked in the Hc atom. As expected from chemical grounds, when the Hc atom is negatively/positively charged and thus the oxygen bears a cancelling positive/negative net charge, the latter atom becomes more/less electronegative, thus the Ha-O bond becomes more/less polarized toward O. Since the probability of the blocked channel is known (in this case, the marginals for $n_c = 2, 1, 0$,respectively) we can also average each partial EDF with its marginal probability to get an average 2c,2e distribution.

The procedure sketched here is easy to generalize. Let us consider a $\mathrm{R}_1{-}\mathrm{A}{-}\mathrm{B}{-}\mathrm{R}_2$ fragment for which we want to isolate the A$-$B partial EDF. We can block both the R_1 and R_2 groups by fixing their populations, $n(\mathrm{R}_1), n(\mathrm{R}_2)$. If the $\mathrm{R}_1{-}$A, A$-$B and B$-\mathrm{R}_2$ links are assumed statistically independent, then by blocking $n(\mathrm{R}_1), n(\mathrm{R}_2)$ we effectively avoid communication of A with R_1 and of B with R_2. Denoting the populations of R_1, A, B, and R_2 as n_1, n_a, n_b, and n_2, respectively, we define the n_1, n_2-dependent A$-$B partial EDF probabilities as

$$p(\tilde{n}_a, \tilde{n}_b)^{n_1,n_2} = p(n_1, n_a, n_b, n_2)/p(n_1, n_2) \; \forall \; n_a, n_b \text{ at fixed } n_1, n_2.$$

In this expression, $\tilde{n}_a = n_a + n_1$, $\tilde{n}_b = n_b + n_2$ are total electron counts that may be used as proper references for the A-B fragments. For instance, $\tilde{n}(\mathrm{O}) =$

$n(\mathrm{O}) + 8$ in Table 6. The averaged partial EDF thus becomes

$$\tilde{p}(\tilde{n}_a, \tilde{n}_b) = \sum_{n_1,n_2} p(n_1, n_2) p(\tilde{n}_a, \tilde{n}_b)^{n_1,n_2} = \sum_{n_1+n_a=\tilde{n}_a, n_2+n_b=\tilde{n}_b} p(n_1, n_a, n_b, n_2),$$

which is seen to be nothing but the EDF that we would obtain by grouping together $\mathrm{R}_1-\mathrm{A}$ and $\mathrm{B}-\mathrm{R}_2$ as two fragments and examining their distribution function. In the water molecule case, the average distribution in Table 6 is nothing but the Ha-(OHc) EDF, that as previously noticed includes Ha-Hc delocalization. The reason why this is so is easy to unmask. For instance, when $n(\mathrm{Hc})= 2$ in Table 6, the $(0, 8, 2)$ structure can be interpreted both as coming from either a Ha atom that has delocalized its electron to the oxygen and an oxygen that delocalizes its OHc pair into the Hc atom, or as an spectator oxygen and a Ha atom that delocalizes its electron into Hc, i.e, as a direct Ha-Hc delocalization.

There is no general solution to this problem. If we would like to avoid these or similar $\mathrm{R}_1-\mathrm{R}_2$ direct transfers we must exclude cyclic delocalizations, like Ha donating to Hc that donates to O which in turn donates to Ha. If this is a good approximation, then taking the partial EDF for neutral (or reference) R_1, R_2 fragments minimizes the impact of direct $\mathrm{R}_1-\mathrm{R}_2$ transfers, for the only way for these two fragments to maintain their initial populations is a cyclic electron movement. In the $\mathrm{H}_2\mathrm{O}$ case, we would take the $n(\mathrm{Hc}) = 1$ partial EDF as defining the $\mathrm{Ha}-\mathrm{O}$ bond.

7 Fitting EDFs

At the beginning of the previous Section we identified two problems that needed to be solved before considering seriously how to extract chemical information from a given EDF. We have now shown how to discern the centers that are involved in strong links and where and among which centers to look for multicenter interactions. Fortunately, the quality (negligible or not) and quantity (single, multiple bonding) of any of these nc,2e links can be extracted from the values of the several cumulants κ_n.

Now we examine a simple recipe to shed light on what we can expect on the general problem of modelling an EDF. We will not recourse to the advanced statistical techniques or machine learning methodologies that are needed to appropriately tackle this problem, but we will use instead our previous insights coupled to standard minimization procedures. In line with the manuscript, we restrict to a general presentation, leaving any implementation details for future works.

Let us assume that an EDF and a molecular structure are provided. The fitting algorithm contains the following steps: (i) *Compute all relevant cumulants κ_n's starting with the largest multicenter interaction expected.* If three-center bonds are expected, for instance, compute κ_3 among all trios. Using the magnitude of κ_3, identify the centers involved in three-center bonds. (ii) *Descend the n ladder* until the κ_2 values between any pair of centers have been exhausted.

Identify all two-center bonds, including their multiplicity, and draw the chemical graph. (iii) *Build the model EDF as a direct product of the nc,2e interactions that have been found*. Each of the factors of the product depends on model variables. For instance, in a 2c,2e link we need the q, f parameters. Feed the EDF builder with appropriate starting values for all of them. (iv) *Use a robust minimizer* to achieve, in a least squares sense, an optimum set of parameters that best describes the full exact EDF.

Although the above algorithm is still fed manually, it is sufficient for our purpose here: showing how to attack, albeit in a simple manner, this difficult problem. We have implemented a module within the EDF code [22, 47] that minimizes the least squares discrepancy of the computed EDF from that of a model direct product using the above ideas.

Using our water molecule example, and noticing that both κ_3 and δ(Ha,Hc) are very small, we may decide to construct a model with two equivalent 2c,2e links, each characterized by a q, f pair. In this way, the oxygen atom bears 6 localized electrons and two two-center bonds so that $\boldsymbol{p}^{10} = \boldsymbol{p}^2(\text{Ha,O}) \otimes \boldsymbol{p}^2(\text{Hb,O}) \otimes i(\boldsymbol{p}^1)^6$, and we optimize the q, f parameters against the full EDF of Table 3. The optimum parameters are $q = 0.601$ and $f = -0.015$. The product EDF is found in Table 7

Table 7: Comparison of the exact and fitted QTAIM EDF for a HF//6-311G(d,p) calculation in H_2O. The optimum parameters are $q_{\text{opt}} = 0.601$ and $f_{\text{opt}} = -0.015$. Only inequivalent probabilities are shown.

n(Ha)	n(O)	n(Hc)	p(exact)	p(fitted)
0	10	0	0.4133	0.4139
1	9	0	0.2032	0.2024
1	8	1	0.0971	0.0989
2	8	0	0.0274	0.0271
1	7	2	0.0126	0.0133
2	6	2	0.0002	0.0002

It is easily checked that the reconstructed EDF improves slightly the one modeled in Table 3. This is done at the expense of introducing a small negative correlation factor that is used to absorb the neglected Ha,Hc delocalization. As demonstrated above, this product model provides a vanishing δ(Ha,Hc). The final quadratic deviation is rather small, 5×10^{-6}

Another simple example is provided by the CH_4 molecule. We use this time a valence CAS//TZV calculation that includes most of its static correlation. We again use three- and two-center cumulants to demonstrate that only the C−H links are relevant in a first approximation and construct a $(\boldsymbol{p}^2(\text{C-H}))^4 \otimes (\boldsymbol{p}^1)^2$ model with only one irreducible bond. Table 8 compares the exact and fitted spinless EDF.

Despite the crudeness of the model, and as we saw, the EDF is well reproduced. Interestingly, the average populations show opposing polarities. The

Table 8: Comparison of the exact and fitted QTAIM EDF for a CAS[4,8]//TZV calculation in CH_4. The optimum parameters for the C−H 2c,2e link are $q_{\rm opt} = 0.0062$ and $f_{\rm opt} = 0.0805$. Only inequivalent probabilities are shown.

n(C)	n(Ha)	n(Hb)	n(Hc)	n(Hd)	p(exact)	p(fitted)
6	1	1	1	1	0.08592	0.08519
7	1	1	1	0	0.03533	0.03673
5	2	1	1	1	0.03499	0.03575
6	2	1	1	0	0.01704	0.01542
8	1	1	0	0	0.01285	0.01584
4	2	2	1	1	0.01167	0.01501
7	2	1	0	0	0.00724	0.00664
5	2	2	1	0	0.00681	0.00647
6	2	2	0	0	0.00343	0.00279
3	2	2	2	1	0.00316	0.00630
8	2	0	0	0	0.00272	0.00286
4	2	2	2	0	0.00223	0.00271
10	0	0	0	0	0.00122	0.00294
2	2	2	2	2	0.00069	0.00264

exact and fitted $\bar{n}$(C) are 6.025 and 5.986 respectively. The positive net charge of the QTAIM carbon atom is a well known feature that has been criticized, since the standard Pauling electronegativity scale suggests opposite polarity. Here we show that if we neglect the H−H delocalization we recover the traditional polarity of the C−H bond, pointing to this non-standard exchange channel as the culprit of the *"wrong"* QTAIM polarity.

The fitting procedure is able to unmask that the C-H bonds are positively correlated, as expected. The f value in this case (0.081) has to be gauged, for instance, with the full-CI result in the H_2 molecule, where $f = 0.151$.

8 Conclusions

We have devoted this manuscript to present, in as close to a layman's language as possible, the statistical point of view in the theory of the chemical bond. This is born from the very probabilitic rules of quantum mechanics as soon as the electrons are associated to atoms. Since the electron number operator does not commute with a partial (atomic) Hamiltonian, the conclusion that the atomic electron population fluctuates follows immediately. If this fluctuations are connected with exchange, which has been many times pointed out as the root (or glue) of bonding [9], then an easy to grasp link among the fluctuation of the atomic electron populations, electron delocalization and chemical bonding appears. When two atoms are bonded, their populations fluctuate as the electrons delocalize between them.

Since the probabilities of finding an integer, given number of electrons in the

set of atoms comprising a molecule can be computed both for model as well as for production-ready wavefunctions, the analysis of the probabilistic distribution function, the EDF, provides a new, enlightening perspective of what a chemical bond means.

The EDF approach to chemical bonding does work without explicit mention to orbitals, although it can benefit from them. Actually, as we have shown, the recourse to orbital images provides a privileged window into how some rooted concepts in chemistry are related to the physics of bonding. Interestingly (this has been known for years within a specialized community working in chemical bonding in real space), the statistical moments of the electron distribution function measure nothing but the traditional bond orders of computational chemistry. Multicenter bonding from the statistical perspective is nothing but the mutual interdependence of the electron populations of several (more than two) atoms.

A simple mnemotecnic rule appears: if the populations of two atoms are statistically independent, they are not (covalently) bonded. Similarly, only if the populations of three centers cannot be disentangled, does a three-center bond exist. As we have shown, the cumulants moments of the EDF easily encode this information.

Much can be learnt from building EDFs from model wavefunctions. To that end, the zero differential overlap approximation with a Mulliken projection has proven very useful. This is simple enough to be taught in fresh inorganic or organic chemistry courses, needing only from the standard model orbitals that are found in every modern textbook. Using it we have shown the consequencies of the Pauli principle in real space, forbidding explicitly that two same spin electrons occupy the same atomic state thus shaping the behavior of the EDF. Using basic algebra the Aufbau principle in diatomic molecules is also found without effort. This leads to construct the approximate EDF of a molecule as a lego-like product of bonds.

EDFs allow a new, insightful classification of n-center chemical bonds in terms of a small set of parameters with clear physical meaning. We have only examined in detail the two-center and three-center one- and two-electron bonds, but it is not difficult to generalize to other possibilities. EDF classifications uncover many new exotic bonding regimes that have not been explored so far: bosonized bonds, in which electrons prefer to pair instead of separate themselves as expected; negative bond order links, where electrons appear or disappear simultaneously from two centers due to the presence of a third, etc. These are not only of academic interest. As we have shown, they do appear in real systems.

The use of EDFs also leads to uncovering the physical difference between the traditional 3c,2e and the Pimentel's 3c,4e bond. It is the different type of fluctuation, not the presence of a second orbital, that distinguishes them. To our knowledge, all these insights are new, and may open new avenues in the theory of the chemical bond.

Once the basic relation between the EDF point of view and the standard position has been clarified, we turned to show how to obtain chemical informa-

tion from a real EDF, computed by whatever methodology. We have illustrated that the problem requires advanced statistical techniques, out of the scope of this presentation, but that much can be learnt from simpler methods. The easiest way to uncover chemistry is to compute the several cumulant moments to construct the chemical graph and to decide which atoms are bonded, and what type of links exist among them. When this is done, two simple methods, the partial EDF and the fitting techniques allow to obtain the parameters of the different links into which we partition the distribution function.

Although the ideas here contained may in part be found interspersed in the recent literature, no self-contained non-mathematical account of them was available. The statistical point of view provides a different perspective of chemical phenomena, one in which many rooted chemical concepts like charge transfer, electron localization and delocalization, resonance structures, polarity, etc, can be directly imaged in terms of *"snapshots"* of the electron distribution. In these snapshots, electrons lie in particular atoms, with different probabilities that change in the course of a chemical process. We think that this perspective may be very fruitful, and we encourage the community to take it into account. The statistical approach is easy to teach and understand, admits from very simple models all up to the most advanced electronic structure methods without changing its narrative, and provides a bridge between the molecular orbital paradigm and the real space theories of chemical bonding. Both have been living relatively isolated for a long time, and we hope that this work may contribute to communicate both worlds better.

Acknowledgment

We thank the spanish MINECO, grant PGC2018-095953-B-I00, the FICYT, grant IDI-2018-000177, and the European Union FEDER funds for financial support.

Keywords

Chemical bonding, Atoms in Molecules, Bonding classification, Electron delocalization

TOC

Chemical Bonding (CB) can be understood from the point of view of the statistics of the electron distribution without recourse to the orbital paradigm and can be envisaged as a game in which atoms juggle electrons. Here we show how the complete electron distribution function (EDF) allows us to reconstruct/interpret basic concepts of CB theory, to classify all possible bonds of a given kind, and provides an extensive set of chemical bonding indicators.

Bibliography

[1] W. Heitler, F. London, *Z. Physik* **1927**, *44*, 455–472.

[2] F. L. Hirshfeld, *Theor. Chim. Acta* **1977**, *44*, 129–138.

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

[4] V. Vitek, *MRS Bulletin* **1996**, *21*, 20–23.

[5] F. Fantuzzi, D. W. O. de Sousa, M. A. C. Nascimento, *The Nature of the Chemical Bond from a Quantum Mechanical Interference Perspective*, **2017**.

[6] T. Berlin, *J. Chem. Phys.* **1950**, *19*, 208–213.

[7] K. Ruedenberg, *Rev. Mod. Phys.* **1962**, *34*, 326–376.

[8] M. Born, *Zeitschrift für Physik* **1926**, *38*, 803–827.

[9] S. Kurth, J. P. Perdew, *Int. J. Quantum Chem.* **2000**, *77*, 814–818.

[10] R. F. W. Bader, M. E. Stephens, *Chem. Phys. Lett.* **1974**, *26*, 445–449.

[11] M. Giambiagi, M. S. de Giambiagi, K. C. Mundim, *Struct. Chem.* **1990**, *1*, 423–427.

[12] K. C. Mundim, M. Giambiagi, M. S. de Giambiagi, *J. Phys. Chem.* **1994**, *98*, 6118–6119.

[13] X. Fradera, M. A. Austen, R. F. W. Bader, *J. Phys. Chem. A* **1999**, *103*, 304–314.

[14] X. Fradera, J. Poater, S. Simon, M. Duran, M. Solà, *Theor. Chem. Acc.* **2002**, *108*, 214–224.

[15] E. Francisco, A. Martín Pendás, M. García-Revilla, R. Álvarez Boto, *Comput. Theor. Chem.* **2013**, *1003*, 71–78.

[16] M. Rafat, P. L. A. Popelier in *The Quantum Theory of Atoms in Molecules. From Solid State to DNA and Drug Design*, C. F. Matta, R. J. Boyd (Eds.), WIley-VCH, **2007**, p. 121.

[17] E. Francisco, D. Menéndez-Crespo, A. Costales, A. Martín Pendás, *J. Comput. Chem.* **2017**, *38*, 816–829.

[18] E. Chamorro, P. Fuentealba, A. Savin, *J. Comp. Chem.* **2003**, *24*, 496–504.

[19] E. Francisco, A. Martín Pendás, M. A. Blanco, *J. Chem. Phys.* **2007**, *126*, 094102–1–094102–13.

[20] A. Martín Pendás, E. Francisco, M. A. Blanco, *J. Chem. Phys.* **2007**, *127*, 144103.

[21] A. Martín Pendás, E. Francisco, M. A. Blanco, *J. Phys. Chem. A* **2007**, *111*, 1084–1090.

[22] E. Francisco, A. Martín Pendás, M. A. Blanco, *Comp. Phys. Commun.* **2008**, *178*, 621–634.

[23] A. Martín Pendás, E. Francisco, M. A. Blanco, *Phys. Chem. Chem. Phys.* **2007**, *9*, 1087–1092.

[24] E. Francisco, A. Martín Pendás, M. A. Blanco, *Theor. Chem. Acc* **2011**, *128*, 433–444.

[25] E. Cancès, R. Keriven, F. Lodier, A. Savin, *Theor. Chem. Acc.* **2004**, *111*, 373–380.

[26] M. A. Blanco, A. Martín Pendás, E. Francisco, *J. Chem. Theory Comput.* **2005**, *1*, 1096–1109.

[27] A. Martín Pendás, E. Francisco, M. A. Blanco, *Chem. Phys. Lett.* **2007**, *437*, 287–292.

[28] M. García-Revilla, E. Francisco, P. L. A. Popelier, A. Martín Pendás, *ChemPhysChem* **2013**, *14*, 1211–1218.

[29] J. Jara-Cortés, J. M. Guevara-Vela, Á. Martín Pendás, J. Hernández-Trujillo, *J. Comput. Chem.* **2017**, *38*, 957–970.

[30] I. Mayer, *Bond Orders and Energy Components: Extracting Chemical Information from Molecular Wave Functions*, CRC Press, Boca Raton, FL, **2016**.

[31] W. N. Lipscomb, *J. Inorg. Nucl. Chem.* **1959**, *11*, 1–8.

[32] W. H. Eberhardt, B. Crawford, W. N. Lipscomb, *J. Chem. Phys.* **1954**, *22*, 989–1001.

[33] S. H. Bauer, *Chem. Rev.* **1942**, *31*, 43–75.

[34] E. Lukacs, *Adv. Appl. Probab.* **1972**, *38*, 1–38.

[35] J. Molina, J. A. Dobado, *Theor. Chem. Acc.* **2001**, *105*, 328–337.

[36] G. C. Pimentel, *J. Chem. Phys.* **1951**, *19*, 446–448.

[37] R. E. Rundle, *J. Phys. Chem.* **1957**, *61*, 45–50.

[38] A. Gallo-Bueno, E. Francisco, A. Martín Pendás, *Phys. Chem. Chem. Phys.* **2016**, *18*, 11772–11780.

[39] A. Gallo-Bueno, M. Kohout, A. Martín Pendás, *J. Chem. Theory Comput.* **2016**, *12*, 3053–3062.

[40] A. Martín Pendás, J. M. Guevara-Vela, D. Menéndez-Crespo, A. Costales, E. Francisco, *Phys. Chem. Chem. Phys.* **2017**, *19*, 1790–1797.

[41] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, *J. Comput. Chem.* **1993**, *14*, 1347–1363.

[42] C. K. Chow, C. N. Liu, *IEEE Transactions on Information Theory* **1968**, *14*, 462–467.

[43] J. Suzuki in *6th European Workshop on Probabilistic Graphical Models*, pp. 315–322.

[44] J. Pearl, *Causality: Models, Reasoning, and Inference, Second Edition*, Cambridge University Press, New York, **2011**.

[45] G. F. Cooper, *Artificial Intelligence* **1990**, *42*, 393–405.

[46] M. García-Revilla, P. L. A. Popelier, E. Francisco, A. Martín Pendás, *J. Chem. Theory Comput.* **2011**, *7*, 1704–1711.

[47] E. Francisco, A. Martín Pendás, *Comput. Phys. Commun.* **2014**, *185*, 2663 – 2682.