PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Information and complexity measures in the molecular reactivity studies.

Meressa A. Welearegay, Robert Balawender* and Andrzej Holas

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
DOI: 10.1039/b000000x

The analysis of the information and complexity measures as tools for the investigation of the chemical reactivity has been done in the spin-position and the position spaces, for the density and shape representations. The concept of the transferability and additivity of atoms or functional groups were used as "checkpoints" in analysis of obtained results. The shape function as an argument of various measures reveals less information than the spinor density. Use of the shape function can yield wrong conclusions when the information measures such as the Shannon entropy (SE, S), the Fisher information (FI, I), the Onicescu information (OI, D) and complexities based on them are used for the systems with different electron numbers. Results obtained in the spinor-density representation show the transferability and additivity (while lacking in the case of the shape representation). The group transferability is well illustrated on the example of the X–Y molecules and their benzene derivatives. Another example is the methyl group transferability presented on the alkane-alkene-alkyne set. Analysis of the results displayed on planes between the three information-theoretical (IT) based measures has shown that the $S-I$ plane provides "richer" information about the pattern, organization, similarity of used molecules than the $I-D$ and $D-S$ planes. The linear relation of high accuracy is noted between the kinetic energy and the FI and the OI measures. Another interesting regression was found between the atomization total energy and the atomization entropy. Unfortunately, the lack of the group electronic energy transferability results that no general relations between the IT measures and the chemical reactivity indices are observed. The molecular set chosen for the study includes different types of molecules with various functional groups (19 groups). The used set is enough large (more than 700 molecules) and diverse to improve the previous understating of molecular complexities and to generalize obtained conclusions.

I. Introduction

The development of methods for systematic and rational classification of chemical compounds and the search for compounds with desired properties is a fundamental task in chemical, material and pharmaceutical research. According to the density functional theory (DFT), a description of the ground state properties of a molecule/atom, including the electronic structure and chemical reactivity is possible using density functional reactivity theory, often called the conceptual DFT.[1-6] It is based on the study of the $E=E[N,v]$ (the energy as the function of the number of electrons N and a functional of the external potential $v(\mathbf{r})$), more precisely, on the change it undergoes when the number of electrons and /or the external potential is varied, concretized by the corresponding response functions. Traditionally conceptual DFT has been concerned primarily with the prediction of phenomena associated with electron transfer, i.e. responses to changes in the number of electrons. Recently, numerous workers have begun to focus on the response of a system to a change in the external potential,[7-9] which for example results from shifts of the nuclear positions in a molecule, from the approach of reagents or from changing the nuclear charges. [10-16]

The use of the information theory (IT) to describe the chemical bonding and the chemical reaction was stimulated by the fact that density function is ultimately a probability distribution. The derivation of the Schrödinger equation by minimizing the Fisher information was the first application of IT concept in quantum chemistry.[17, 18] Other significant result is recovering the periodicity of atomic properties within the information theoretical framework (see Ref.[19] for review). Since then, there has been a tremendous interest to apply IT to the electronic structure theory. The concepts of uncertainty, randomness, disorder or delocalization, are the basic quantities appearing in various chemical applications. In quantum chemistry, the IT objects were used to optimize and improve basis set, to measure the amount of correlation included in a wavefunction, [20-24] to measure the similarity,[25-28] and to perform very promising IT investigation of the molecular bond[29-34] and reaction path.[35-38] One of the challenges is the recognition of the chemical reactivity by employing IT based measures and statistical complexity.[39, 40]

As it follows from the texts concerning the complexity, there is

no unique and universal definition of the complexity for arbitrary distribution.[39] Complexity is used in very different fields, although there is no general agreement about its definition. Successes in its applications suggest that the characterization of complexity cannot be univocal and must be adequate for the type of structure or process we study. In general, the initial form of complexity is designed in such a way that it vanishes for the two extreme probability distributions: corresponding to the perfect order (represented by a Dirac-delta in the shape representation) and to the maximum disorder (associated with a highly flat distribution). Recent proposals have formulated this quantity as a product of two factors, taking into account *order/disequilibrium* and *disorder/uncertainty*, respectively. One simple measure of the complexity can be defined as the product of the exponential Shannon entropy and the Fisher information (FI), the Fisher-Shannon complexity (FS), Eq.(11). Another complexity by replacing the FI by the Onicescu information[41] (OI) (the disequilibrium measure) the López-Ruiz–Mancini–Cablet (LMC) shape complexity, Eq.(9).

The goal of the present study is two-fold: (i) answer the question, if it is preferable to use the electron density rather than the shape function as the functional argument and (ii) the recognition of the relations between the IT complexity measures, their components and the molecular reactivity. The first subject is related to the fact that all complexity measures along with their IT components are originally defined in terms of the statistical distribution functions which are normalized to unity and defined in the position space (or momentum space). This means that all properties and uncertainty relationships which are known from the statistic or the IT are valid for the $\sigma_{\mathrm{N}}(\mathbf{r})$ functional (the spinless shape), but not always for other distributions used in quantum chemistry, e.g., the spinless density. The second subject will be realized by analyzing relations between the IT measures and the DFT based chemical reactivity indices (the chemical potential and the hardness), energy components, the Pauli energy and the atomization energy.

The organization of the paper is as follows: in section 2, the theoretical and methodological framework, the complexity measures along with their IT components and the chemical reactivity descriptors used in this work are defined and shortly discussed. In the section 3, the subjects proposed above are calculated for a large set of molecules and thoroughly discussed. In section 4, conclusions are given. In Appendix, the relations among the IT measures in different representation (the functional dependence) are derived.

II. Theoretical and methodological framework

An *N*-particle system is described in quantum mechanics by means of its wavefunction $\Psi(\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_N)$, depending on the spin-position coordinates $\mathbf{x}_i=(\mathbf{r}_i;\kappa_i)$, where $\mathbf{r}=(x,y,z)$ is a position vector and $\kappa\in\{\alpha,\beta\}$ is a spin variable. The physical and chemical properties of the systems are described by spinor electron density defined as

$$\rho(\mathbf{x}_1)=N\int\left|\Psi(\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_N)\right|^2 d\mathbf{x}_2...d\mathbf{x}_N\,. \qquad (1)$$

This density can be reduced to the spinless one

$$\rho_{\mathrm{N}}(\mathbf{r})=\sum_{\kappa}\rho(\mathbf{r},\kappa)=\rho_{\alpha}(\mathbf{r})+\rho_{\beta}(\mathbf{r})\,, \qquad (2)$$

where $\rho_{\kappa}(\mathbf{r})\equiv\rho(\mathbf{r},\kappa)$ are the spin components of the spinor density. All these densities will be collectively denoted by $\zeta(\mathbf{q})$ with

$$\zeta(\mathbf{q})\in\{\rho(\mathbf{x}),\rho_{\mathrm{N}}(\mathbf{r}),\rho_{\alpha}(\mathbf{r}),\rho_{\beta}(\mathbf{r})\}\,. \qquad (3)$$

Their integral over the whole space (the spin-position space or the position space) is equal to the related electron number

$$\int\zeta(\mathbf{q})d\mathbf{q}=N_{\zeta},\qquad N_{\zeta}\in\{N,N,N_{\alpha},N_{\beta}\},\ N_{\alpha}+N_{\beta}=N\,. \qquad (4)$$

Their normalized to 1 forms, the so-called the shapes, are defined as follows

$$\sigma_{\zeta}(\mathbf{q})\equiv\frac{\zeta(\mathbf{q})}{N_{\zeta}},\quad \text{satisfying } \int\sigma_{\zeta}(\mathbf{q})d\mathbf{q}=1\,. \qquad (5)$$

The density $\zeta(\mathbf{q})$ can be characterized in different ways: as power moments $\langle r^k\rangle_{\zeta}$, the logarithmic moments $\langle r^k\ln r\rangle_{\zeta}$ or entropic moments $\langle\zeta^k\rangle_1\equiv\langle\zeta^{k-1}\rangle_{\zeta}$. The symbol $\langle f(\mathbf{q})\rangle_{\zeta}$ denotes the expectation value $\langle f(\mathbf{q})\rangle_{\zeta}\equiv\int f(\mathbf{q})\zeta(\mathbf{q})d\mathbf{q}$. For specific values of k, these moments are physically meaningful and, at times, experimentally accessible, e.g. the electron-nucleus attraction energy which is related to the nuclear magnetic screening constant or diamagnetic screening factor, is given by $-Z_A\langle r_A^{-1}\rangle_{\zeta}=-Z_A\int\zeta(\mathbf{q})/|\mathbf{r}-\mathbf{R}_A|d\mathbf{q}$, $\mathbf{R}_A-$ the nucleus position.[42] Another example is the limit of the entropic moments, the Shannon entropy (SE)

$$S[\zeta] \equiv -\int \zeta(\mathbf{q}) \ln \zeta(\mathbf{q}) d\mathbf{q} = -\langle \ln \zeta \rangle_\zeta = -\lim_{k\to 1} \frac{\partial \langle \zeta^{k-1} \rangle_\zeta}{\partial k}. \quad (6)$$

In quantum physics, $S[\rho_N]$ was introduced by von Neumann as an adaptation of the thermodynamic Gibbs entropy. For normalized argument case, it is the differential entropy $S[\sigma_N]$. It is worth noting that, unlike the discrete entropy, $-\sum_i p_i \ln p_i \geq 0$, the continuous one $S[\zeta]$ can have any value in $[-\infty, +\infty]$. Any sharp peak in $\zeta(\mathbf{q})$ will tend to make negative contribution to $S[\zeta]$, whereas positive contribution are provoked by a slowly decaying tail; hence $S[\zeta]$ is a localization measure of the density. However, the differential entropy is called very often the Shannon continuous entropy (or simply SE), but it is clear that the differential entropy does not share all properties of the discrete entropy. The SE was first introduced as a way to measure the information content (or the uncertainty) in a distribution. It is a measure of localization (delocalization) or structure that exists in its functional argument, $\zeta(\mathbf{q})$.[43] As a measure of the delocalization, the SE is relevant for the bonding theory.[44]

The Fisher information (FI) of the density $\zeta(\mathbf{q})$ is defined

$$I[\zeta] \equiv \left\langle \left| \nabla \ln \zeta \right|^2 \right\rangle_\zeta = \int \frac{\nabla \zeta(\mathbf{q}) \cdot \nabla \zeta(\mathbf{q})}{\zeta(\mathbf{q})} d\mathbf{q}, \quad (7)$$

where ∇ denotes the 3-dimensional gradient. The FI, contrary to the SE, is a measure sensitive to a local spreading of the density $\zeta(\mathbf{q})$, because it is a gradient functional of $\zeta(\mathbf{q})$. The higher is this gradient, the more localized is the density (the smaller is its uncertainty of localization) and the higher is the sharpness in estimating the localization of the electrons. The sharpness, concentration or delocalization of the electronic cloud is measured by the SE as well by the FI. They give complementary descriptions of the electron localization: SE is a measure sensitive to global aspects, while the FI – to local ones. In the atomic case, the FI is a measure of density compactness, in the case of molecule, it is a measure of "peakiness". This is related to the fact that the von Weizsäcker kinetic energy (equals to $I[\rho_N]/8$) is dominated by contributions of *K* electrons in atoms and molecules.[45] The FI was proposed also as the measure of the steric contribution to the total energy.[46]

The last measure, which is used here, is the Onicescu information (OI)[41] (the disequilibrium), the second-order entropic moment

$$D[\zeta] \equiv \left\langle \zeta^2 \right\rangle_1 = \int \zeta^2(\mathbf{q}) d\mathbf{q} = \langle \zeta \rangle_\zeta. \quad (8)$$

It quantifies the departure of $\zeta(\mathbf{q})$ from equiprobability. The concept of this measure is a finer measure of dispersion distribution than that of the Shannon entropy. So far, only the mathematical aspects of this concept have been developed,[47, 48] while the physical or chemical aspects were not investigated as yet.[19, 49, 50]

These three IT based measures are used to define two complexity measures, the LMC and FS complexities. In both, the Shannon entropy (more precisely, its exponent) is used as a measure of the total spreading of $\zeta(\mathbf{q})$. The other factor, measuring order, is the disequilibrium for the LMC complexity or the Fisher information for the FS complexities, respectively. The LMC complexity is defined[51] as

$$C_{\mathrm{LMC}}[\zeta] \equiv D[\zeta] H[\zeta] \quad (9)$$

where

$$H[\zeta] = \exp(S[\zeta]) \quad (10)$$

is the exponential Shannon entropy. The FS complexity is defined[52, 53] as

$$C_{\mathrm{FS}}[\zeta] \equiv I[\zeta] J[\zeta] \quad (11)$$

where

$$J[\zeta] = (2\pi e)^{-1} \exp(2S[\zeta]/3) = (2\pi e)^{-1} (H[\zeta])^{2/3} \quad (12)$$

is the modified exponential Shannon entropy (in three dimensional space).[54] All these definitions are valid also for the $\sigma_\zeta(\mathbf{q})$ functional argument.

The two complexity measures along with their IT components employed throughout this work were originally defined in terms of the statistical distribution functions which are normalized to unity and defined in the position space i.e. $\sigma_N(\mathbf{r})$ (or in the momentum space). However, these measures are calculated in our paper in the spin-position space ($\mathbf{q} = \mathbf{x}$) and in the position space ($\mathbf{q} = \mathbf{r}$) in two representations: the density one and the shape one (i.e. as functionals of the density or the shape). So, the considered arguments are $\rho(\mathbf{x})$ and $\sigma(\mathbf{x})$ in the spin-position space, while $\rho_N(\mathbf{r})$ and $\sigma_N(\mathbf{r})$ in the position space. The relations between the IT measures and the statistical complexities for the general case are presented in Appendix. When sets of the molecules chosen for study contain only singlet molecular systems i.e.,

Table I. A compilation of the relations between $F[\zeta]$ and $F[\zeta']$ for $F \in \{S, I, D, H, J, C_{\mathrm{LMC}}, C_{\mathrm{FS}}\}$ and $\zeta, \zeta' \in \{\sigma_{\mathrm{N}}, \sigma, \rho, \rho_{\mathrm{N}}\}$, limited to spin compensated systems (where $\rho_\alpha(\mathbf{r}) = \rho_\beta(\mathbf{r}) = \rho_{\mathrm{N}}(\mathbf{r})/2$, $\sigma_\alpha(\mathbf{r}) = \sigma_\beta(\mathbf{r}) = \sigma_{\mathrm{N}}(\mathbf{r})$).

$\sigma_{\mathrm{N}} \leftrightarrow \sigma$	$\sigma \leftrightarrow \rho$	$\rho \leftrightarrow \rho_{\mathrm{N}}$	$\rho_{\mathrm{N}} \leftrightarrow \sigma_{\mathrm{N}}$
$S[\sigma] = S[\sigma_{\mathrm{N}}] + \ln 2$	$S[\rho] = N\left(S[\sigma] - \ln N\right)$	$S[\rho_{\mathrm{N}}] = S[\rho] - N\ln 2,$	$S[\sigma_{\mathrm{N}}] = S[\rho_{\mathrm{N}}]/N + \ln N$
$I[\sigma] = I[\sigma_{\mathrm{N}}]$	$I[\rho] = N\, I[\sigma]$	$I[\rho_{\mathrm{N}}] = I[\rho]$	$I[\sigma_{\mathrm{N}}] = I[\rho_{\mathrm{N}}]/N$
$D[\sigma] = D[\sigma_{\mathrm{N}}]/2$	$D[\rho] = N^2 D[\sigma]$	$D[\rho_{\mathrm{N}}] = 2D[\rho]$	$D[\sigma_{\mathrm{N}}] = D[\rho_{\mathrm{N}}]/N^2$
$H[\sigma] = 2H[\sigma_{\mathrm{N}}]$	$H[\rho] = \left(H[\sigma]/N\right)^N$	$H[\rho_{\mathrm{N}}] = 2^{-N} H[\rho]$	$H[\sigma_{\mathrm{N}}] = N\left(H[\rho_{\mathrm{N}}]\right)^{-N}$
$J[\sigma] = 2J[\sigma_{\mathrm{N}}]$	$J[\rho] = (2\pi e)^{1-N} N^{2N/3} \left(J[\sigma]\right)^N$	$J[\rho_{\mathrm{N}}] = 2^{-2N/3} J[\rho]$	$J[\sigma_{\mathrm{N}}] = N^{2/3} (2\pi e)^{(N-1)/N} \left(J[\rho_{\mathrm{N}}]\right)^{1/N}$
$C_{\mathrm{LMC}}[\sigma] = C_{\mathrm{LMC}}[\sigma_{\mathrm{N}}]$	$C_{\mathrm{LMC}}[\rho] = N^{2-N} \left(D[\sigma]\right)^{1-N} \left(C_{\mathrm{LMC}}[\sigma]\right)^N$	$C_{\mathrm{LMC}}[\rho_{\mathrm{N}}] = 2^{1-N} C_{\mathrm{LMC}}[\rho]$	$C_{\mathrm{LMC}}[\sigma_{\mathrm{N}}] = N^{-1} \left(D[\rho_{\mathrm{N}}]\right)^{N+1} \left(C_{\mathrm{LMC}}[\rho_{\mathrm{N}}]\right)^{-N}$
$C_{\mathrm{FS}}[\sigma] = 2^{2/3} C_{\mathrm{FS}}[\sigma_{\mathrm{N}}]$	$C_{\mathrm{FS}}[\rho] = \frac{N^{(2N+3)/3}}{(2\pi e)^{N-1}} \left(I[\sigma]\right)^{-N} \left(C_{\mathrm{FS}}[\sigma]\right)^N$	$C_{\mathrm{FS}}[\rho_{\mathrm{N}}] = 2^{-2N/3} C_{\mathrm{FS}}[\rho]$	$C_{\mathrm{FS}}[\sigma_{\mathrm{N}}] = \frac{(2\pi e)^{(N-1)/N}}{N^{1/3}} \left(I[\rho_{\mathrm{N}}]\right)^{-1/N} \left(C_{\mathrm{FS}}[\rho_{\mathrm{N}}]\right)^{1/N}$

where $\rho_\alpha(\mathbf{r}) = \rho_\beta(\mathbf{r}) = \rho_N(\mathbf{r})/2$, these relations become simple. A compilation of the definitions, relationships among the various spaces and representations considered in this work is given in Table I. It is important to mention that the LMC and FS complexities are known to comply with the following lower bounds: $C_{LMC}[\sigma_N] \geq 1$, $C_{FS}[\sigma_N] \geq 3$ (in the spinless-shape representation) but in other representations these inequalities have to be reformulated using relations from Table I.

To interpret and understand the chemical nature of the IT measures, several reactivity indices are computed. The reactivity properties set used by Esquivel, and co-workers[39, 40] is extended by the atomization energy, energy components and the Pauli energy. In DFT, the total electronic energy is a functional of the electron density $\rho(\mathbf{x})$ and the external potential $v(\mathbf{r})$,

$$E[\rho, v] = T_s[\rho] + E_{es}[\rho] + E_{xc}[\rho] + \int v(\mathbf{r})\rho(\mathbf{x})d\mathbf{x}, \quad (12)$$

the sum of the kinetic energy, $T_s[\rho]$, of the non-interacting reference system, the electrostatic energy, $E_{es}[\rho]$, and the exchange-correlation energy, $E_{xc}[\rho]$. The ground-state energy, $E[N, v]$, is the minimum in the space of N-electron densities, $E[N, v] = \underset{\rho \to N}{\text{Min}} E[\rho, v] = E[\rho^{gs}, v]$. The minimizer, ρ^{gs}, represents the ground-state density of the system. The sum of the electronic energy E for $v = v[\{Z\}, \{\mathbf{R}\}]$ and the nuclear-nuclear repulsion energy V_{nn} is the total energy W of a system,

$$W[N, \{Z\}, \{\mathbf{R}\}] \equiv E[N, v[\{Z\}, \{\mathbf{R}\}]] + V_{nn}[\{Z\}, \{\mathbf{R}\}], \quad (13)$$

where $\{\mathbf{R}\}$ and $\{Z\}$ denote the nuclear locations and the corresponding nuclear charges. The total energy of a molecular system is often interpreted as stemming from three independent effects: steric, electrostatic, and quantum,[46] $W = W_s + W_{es} + W_q$. The electrostatic contribution is $W_{es} = E_{es}[\rho] + \int v(\mathbf{r})\rho(\mathbf{x})d\mathbf{x}$. The quantum contribution is due to the exchange-correlation and the Pauli energy, $W_q[N, \{Z\}, \{\mathbf{R}\}] = E_{xc}[\rho] + E_P[\rho]$. The Pauli energy represents the portion of the kinetic energy that embodies all the effects from the antisymmetric requirement of the total wavefunction by the Pauli principle[55-57] and it is defined as[58-60]

$$E_P[\rho] = T_s[\rho] - T_W[\rho], \quad (14)$$

where $T_W[\rho]$ is the von Weizsäcker kinetic energy functional

$$T_W[\rho] \equiv \frac{1}{8}\int \frac{\nabla\rho(\mathbf{x})\cdot\nabla\rho(\mathbf{x})}{\rho(\mathbf{x})}d\mathbf{x}. \quad (15)$$

After some algebraic rearrangements, it is found that the contribution from the steric effect to the total energy equals to the Weizsäcker kinetic energy,[46] $W_s = T_W[\rho]$. It easy to notice that the Weizsäcker kinetic energy is the rescaled Fisher information, Eq.(7), $T_W[\rho] = I[\rho]/8$.

The DFT-based chemical reactivity indices, the chemical potential and the chemical hardness are defined[3, 58] as

$$\mu = -(IE + EA)/2, \ \eta = (IE - EA)/2, \quad (16)$$

where IE and EA are the vertical ionization energy and the electron affinity, respectively. In general, the chemical potential and the chemical hardness are good global descriptors of chemical reactivity. The former measures the escaping tendency of the electrons from the system and is constant through a system in equilibrium, while the latter measures the global stability of the molecule (large values of η characterize less reactive molecules–more resistive to change of electron number). The electrophilicity index is given in terms of μ and η as $\omega = \mu^2/2\eta$. It is a descriptor of reactivity that allows for a quantitative classification of the global electrophilic nature of a molecule within a relative scale, it is the power of a system to 'soak up' electrons.[61]

In addition to these global reactivity indices, the atomization energy (activation energy) will be used as the molecule stability descriptor. It is simply defined as the energy change that accompanies the total separation of all atoms in a molecule

$$\Delta W_M = \sum_{A \in M} W_A - W_M \quad (17)$$

where W_M – the total energy of molecule, W_A – the total energy of an atom A.

The molecular set chosen for the study includes different types of molecules with various functional groups (see Scheme 1). This set can be divided into three subsets namely the alkane-alkene-alkyne set (A3 set, 39 molecules,), X–Y molecule set (XY set, 190 molecules) and the *ortho*-, *meta*-, *para*- substituted benzene derivatives (*d*-XY set, 170 molecules at each position). These sets represent a variety of chemical organic systems, including alcohol (-OH), carboxylic acid (-COOH), aldehyde (-CHO), ketone (-$COCH_3$), ether (-CH_2OCH_3), ester (-$COOCH_3$) and amide (-$CONH_2$), for both non-aromatic and aromatic molecules. These sets can be also split into a subset of molecules with non-hydrogen atom from the II row only, from the III row only and

$$C_nH_{2n+2}\,(\blacktriangle),\ C_nH_{2n}\,(\blacktriangledown),\ C_nH_{2n-2}\,(\blacktriangleright),\qquad 3\le n\le 10 \qquad \text{A3}$$

$$X-Y,\quad \{X,Y\}\in\left\{\begin{array}{l}-H,-CH_3,-NH_2,-OH,-F,-SiH_3,-PH_2,-SH,\\ -Cl,-CONH_2,-OCH_3,-CHO,-COCH_3,-C_6H_5,\\ -COOH,-COOCH_3,-CH_2OCH_3,-CN,-NO_2\end{array}\right\}(\bullet) \qquad \text{X–Y}$$

$$d-X-C_6H_4-Y,\quad d\in\{ortho\,(\square),meta\,(\blacksquare),para\,(\blacksquare)\} \qquad \text{d-XY}$$

Scheme 1 Molecular sets. The symbols used in figures are here in parenthesis. The set names are on the right. The color coding used in figures is as follows: red – the molecules with non-hydrogen atoms from II row only and the hydrogen molecule; green – the molecules with non-hydrogen atoms from III row only; blue – the remaining molecules.

the remaining molecules (with atoms from the II and III rows simultaneously).These subsets will be denoted by II, III, and m. The used total set is large enough and diverse to improve the understanding of molecular complexity and to generalize obtained conclusions. All molecules are spin singlet systems. All calculations were done using GAMESS program package.[62] As the exchange-correlation functional, the B3LYP functional was used, known as the most widely used functional in trouble-free situations (covalent interaction, excluding π-π stacking and no transition metal bonds). [63, 64] The cc-pVTZ basis set[65-67] was used as the optimal balance choice between accuracy and computation time. In addition, the XY set was tested with different basis sets (aug-cc-pVTZ) and different approximate exchange-correlation functionals, to see the impact on atomic and molecular values of Shannon entropy and Fisher information. No significant dependence on the choice of basis sets and functional forms has been observed.

For each molecule different information and complexity transferable from one molecule to another (different) molecule.[68, 69] These concepts of transferability and additivity will be tested on the example of separate atoms, the CH_2 group and C_6H_4 group contributions.

The atomic additivity and transferability is investigated using the linear fitting without a constant

$$\tilde{F}\left[\omega_M\right]=a\sum_{A\in M}F\left[\omega_A^0\right],\qquad F\in\{S,I,D\},$$

$$\omega(\mathbf{q})\in\left\{\sigma_N(\mathbf{r}),\sigma(\mathbf{x}),\rho_N(\mathbf{r}),\rho(\mathbf{x})\right\} \qquad (18)$$

to the molecular value $F\left[\omega_M\right]$, for a set of molecules. All three measures used here in the spinor-density representation, $\rho(\mathbf{x})$, show a very high-accuracy linear relation between the molecular measures, i.e. S, I, D, C_{LMC} and C_{FS} were calculated as the spinor-density functional as well as the spinless-shape functionals all in the configuration space. The ionic forms of molecules and atomic data were calculated at the same level. The atomic information measures were calculated for the equi-ensemble density (the average over the spin and spatial degeneracy). All quantities are given in atomic units throughout this work.

III. Results and discussion

III.A. Additivity and transferability of the IT measures

As the argument in discussion about whether more information is revealed using the shape function or using the electron density, the transferability and additivity concepts will be used. They are central to chemistry: the idea that every property of a molecule is approximated by the sum of the contributions from each of its constituent atoms or groups. The observation of "experimental group additivity" requires that in addition to the properties of the groups being additive, the group and its properties have to be measure value and the sum of the atomic measure values, the linear fitting data are presented in Table II. Combination of this fit with the atomization measure, which is defined as

$$\Delta F\left[\omega\right]=\sum_{A\in M}F\left[\omega_A^0\right]-F\left[\omega_M\right], \qquad (18)$$

yields

$$\Delta F\left[\omega\right]=(1-a)\sum_{A\in M}F\left[\omega_A^0\right]+\left(\tilde{F}\left[\omega_M\right]-F\left[\omega_M\right]\right), \qquad (19)$$

where ω_A^0 is the ω-type argument for a free atom.

This means that if the slope value is close to one, the atomization value is close to a fitting error, and observed transferability is the result of compensatory transferability wherein the changes in the properties of one atom/group are compensated for by equal but opposite change in the properties of the adjoining atom/group.[69]

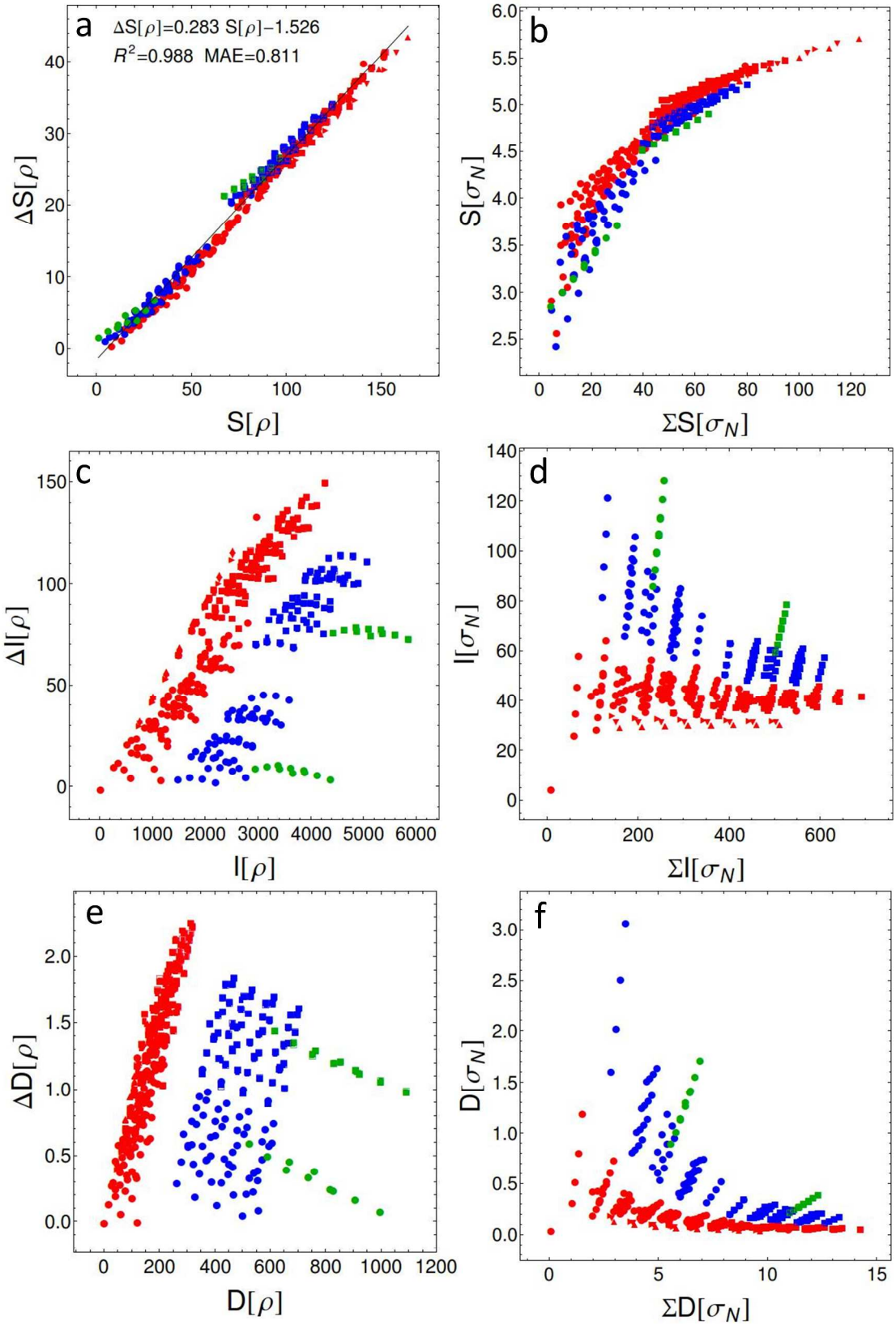

Fig.1 The IT measures of the atomization in the spinor density representation (left panels) and the atomic additivity of the IT measures in the shape representation (right panel). For axis labels see the text, for symbols used in figures see Scheme 1.

In Fig.1, the relation between the molecular value and the atomization value are presented in left panels. The atomization values for all measures are positive. In the case of the FI and the OI, the slope is close to one (see Table II). Small atomization values in comparison with the molecular values for the FI and the OI confirm that the compensating changes are small. Using the fact that the major contribution to the Weizsäcker kinetic energy comes from the *K*-shell electrons for neutral atoms,[45] the *K*-shells model to approximate the FI is tested (see Table II)

$$I_K = 4\sum_A n_A^K Z_A^2, \qquad n_A^K = \{1 \text{ for } A = H;\ 2 \text{ for } A \neq H\}, \tag{20}$$

where n_A^K is the *K*-shell occupation number (one for hydrogen atom, two for the rest) and Z_A is the nuclear charge of *A* atom. The high-accuracy of this model confirms the major contribution to the FI coming from core electron density. Similar model for the OI, $D_K = \sum_A n_A^K Z_A^3 / 8\pi$, based on the hydrogenic atoms[70] produces large errors.

In the case of the SE, the slope significantly differs from one and, what is more, the $\Delta S[\rho]$ can be fitted as the linear function of $S[\rho]$ (see the black line in ΔS vs S). The slope value ($a = 0.788$) should be interpreted as the ratio of the average entropy of the atoms in molecule to the average entropy of free atoms which are constituents of molecules used in this work. It depends on the used set, e.g. for the X–Y-III set and d-XY-III, it is equal 0.806 and 0.777, respectively, but still with high correlation. This behavior is due to the fact that during bond formation, the major changes in the density are observed for the valence part.

As can be noted from the right panels of Fig.1, compared with the relations given in Table II, the transferability and addititivity of atoms shown by the IT measures in the spinor-density representation disappear in the spinless-shape representation. The highly accurate linear relation between the SE of the molecule and the sum of the atomic SEs in the ρ representation (see Table II), is lost in the ρ_N representation as can be easily shown using transformation from Table I,

$$S[\rho_M] = a\left(\sum_{A\in M} S[\rho_A]\right) \Rightarrow \quad S[\rho_N^M] = a\left(\sum_{A\in M} S[\rho_N^A]\right) + (a-1)N_M \ln 2 . \tag{21}$$

The transferability of the methylene group, CH_2 has been extensively studied.[69, 71-73] In Fig.2, the data for the A3 set is presented in both representations. The most remarkable feature that may be observed from these figures is that the density representation (right panels) preserves the group transferability and the size extensiveness. For all three measures, we observed an increasing value with increasing molecular size. Using the electron number as a molecule size index, the excellent fittings to linear relationships are obtained (see Table III). Almost the same slopes of subsets imply for the given measure the conservation of the methylene properties – the same $8a$ contribution to this measure. Zooming in the molecule values, we find for the hydrocarbons with the same carbon number the following trend: from alkane to alkene to alkyne, their value of three measures used here decrease.

Using transformation from the ρ representation to the σ_N representation (Table I, the $\rho \to \rho_N \to \sigma_N$ mapping), we obtain for the σ_N

$$S[\sigma_N] = (a - \ln 2) + \frac{b}{N} + \ln N, \quad I[\sigma_N] = a + \frac{b}{N}, \quad D[\sigma_N] = \frac{2a}{N} + \frac{2b}{N^2}, \tag{22}$$

where a and b are the coefficients of the linear fitting $f[N] = aN + b$ for ρ representation (see Table III). The result for the SE is consistent with the approximate linearity of the sum of Shannon entropies in the position and the momentum space with $\ln N$ observed empirically for atoms.[19] The large N limits are associated with the asymptotic behavior of the OI at zero value and of the FI at the a values, e.g. 31.391 a.u in the alkenes case. As can be noticed form Fig.2 and Eq.(22), with increasing size N of hydrocarbons the $I[\sigma_N]$ values for the alkanes and alkynes become finally the same as the alkenes values. Very often the SE and the FI are considered to be the different sides of the same coins. But after examination of the trends between alkanes, alkenes and alkynes in the shape representation, we conclude that the FI value for alkynes is always higher than for alkenes and for alkenes is higher than for alkanes irrespective of the number of carbon atoms they contain. For a given number of carbon atoms, the entropy trend is opposite to the disequilibrium trend. This observation is in contradiction with the meaning of both measures, because the OI is considered to be finer measure of dispersion distribution than the SE. The FI is interpreted as the steric effect measure. The FI results for the A3 set are inconsistent with the expectation that the steric effect needs to be extensive in size (the FI for alkenes are almost constant with

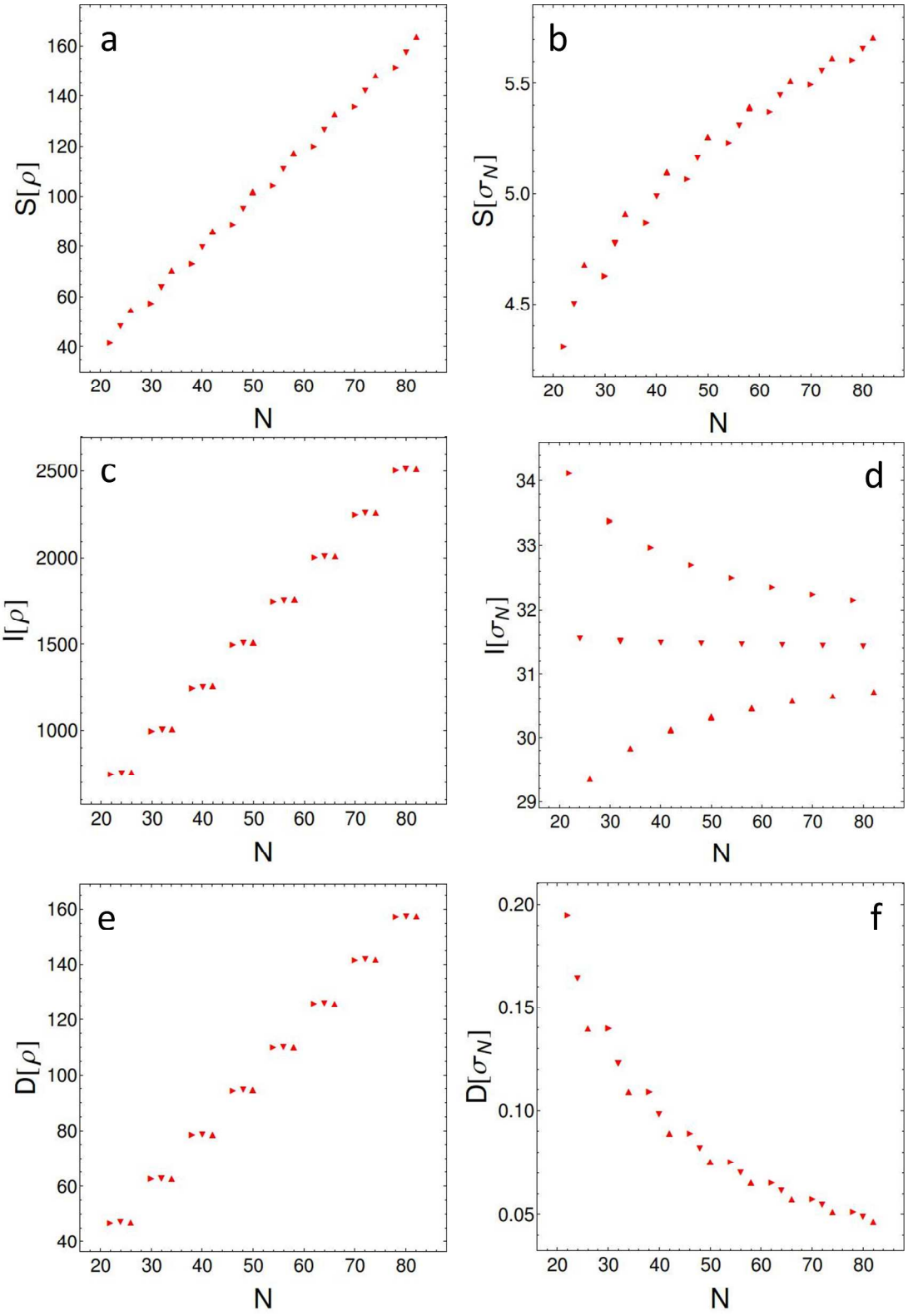

Fig.2 The IT measures vs. electron number for the hydrocarbons (the A3 set) in the spinor density representation (left panels) and the shape representation (right panel). See Fig.1 for details

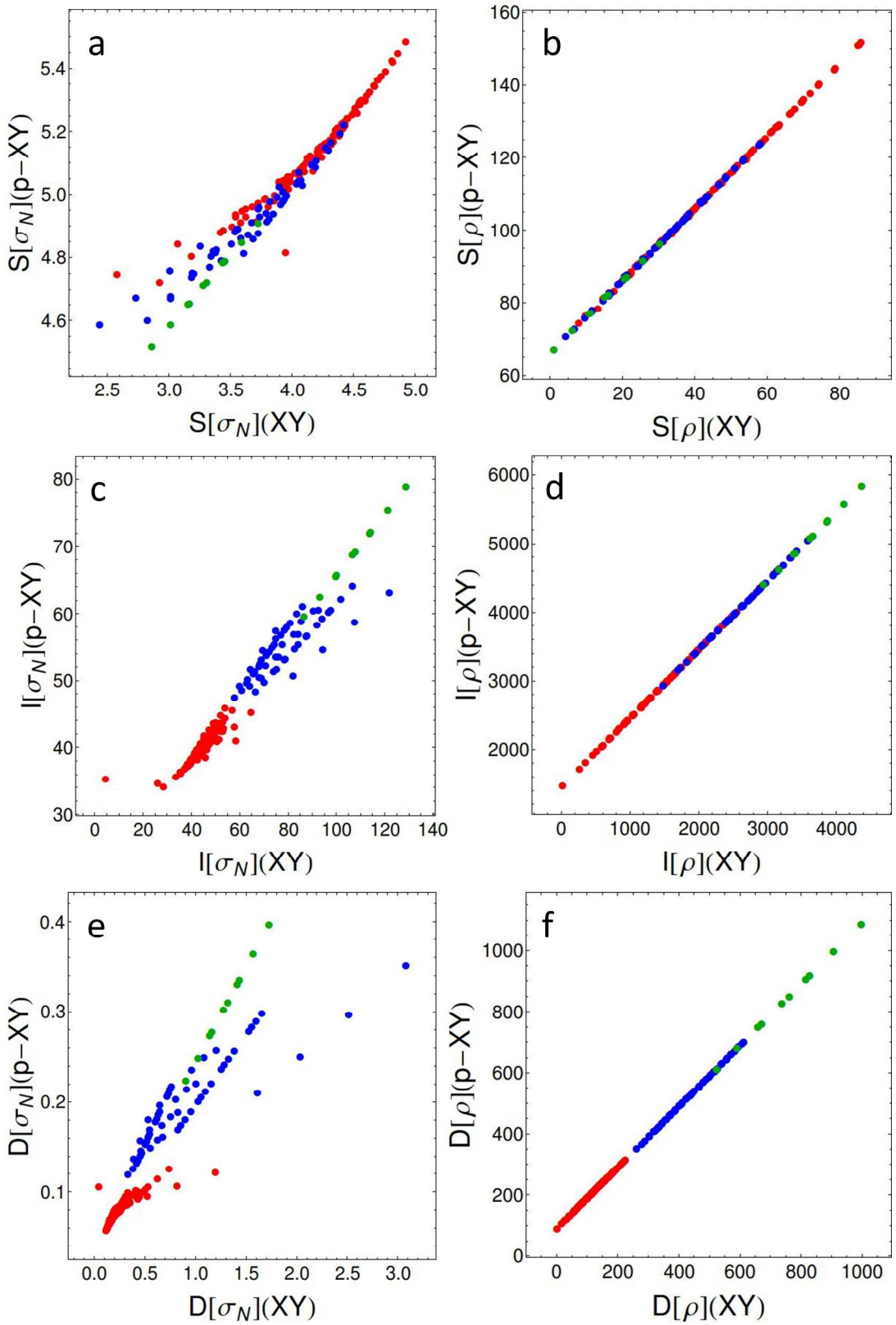

Fig.3 The $p-C_6H_6$ group transferability of the IT measures between the X–Y set and *p*-XY set in the spinor density representation (right panels) and the shape representation (left panel). See Fig.1 for details.

Table II Fitting of the molecular value vs. the sum of atomic values for the Shannon entropy, the Fisher information, the Onicescu information, all in the spinor-density representation, as a linear function, $F[\rho_M] = a\left(\sum_{A\in M} F[\rho_A]\right)$, and the Fisher information approximated by *K*-shells model (see Eq.(20)). The mean absolute error (MAE) and the mean absolute percentage error (MAPE). $R^2 = 1.000$ for each fitting.

$F[\rho]$	a	MAE	MAPE
$S[\rho]$	0.788	0.722	1.441
$I[\rho]$	0.974	25.638	0.926
$D[\rho]$	0.997	0.720	0.339
I_K	0.880	67.464	2.345

Table III The Shannon entropy, the Fisher information, the Onicescu information, all in the spinor-density representation, fitted as a linear function of the electron number, $f[N] = aN + b$, and the mean absolute error, for the subsets of the A3 set. All in a.u.

	$S[\rho]$		$I[\rho]$		$D[\rho]$	
alkanes	$1.946N + 4.738$	0.146	$31.370N - 51.710$	0.257	$1.973N - 3.909$	0.018
alkenes	$1.955N + 1.578$	0.058	$31.390N + 4.091$	0.170	$1.972N + 0.045$	0.008
alkynes	$1.961N - 1.099$	0.020	$31.400N + 59.990$	0.096	$1.972N + 3.947$	0.007

increasing size in the shape representation). All discussed contradictions are observed for the shape representations only.

Another example, which illustrate that the conclusions derived from the shape representation are inconsistent with the group transferability, is presented in Fig. 3. The group additivity and transferability is tested on the examples of the C_6H_4 group present in the d-XY set as compared with the X–Y set. This group transferability was tested using the linear fitting

$$\tilde{F}[\rho_{d-XY}] = a_F F[\rho_{XY}] + b_F \qquad (23)$$

where ρ_{XY} and ρ_{d-XY} are the densities of the X–Y molecule and its benzene derivative. These relations in the case of substituents at the *para* position, in the spinor-density representation, are perfectly linear with $R^2 = 1.000$ (Fig.3, right panels)

$$\begin{aligned} S[\rho_{p-XY}] &= 0.997\, S[\rho_{XY}] + 66.409, && \text{MEA=0.160,} \\ I[\rho_{p-XY}] &= 1.000\, I[\rho_{XY}] + 1477.000, && \text{MEA=0.937,} \\ D[\rho_{p-XY}] &= 1.000\, D[\rho_{XY}] + 94.430, && \text{MEA=0.034.} \end{aligned} \qquad (24)$$

This demonstrates additivity of X and Y groups, in the X–Y molecules and the *p*-XY molecules, and transferability of the $p-C_6H_4$ group. The intercepts in Eq.(24) are the IT measure values for the free p-C_6H_4 group. Similar to Eq.(24), perfectly linear relations ($R^2 = 1.000$) are obtained in the case of substituents at the *meta* and *ortho* position. The difference between the free m-C_6H_4 and the p-C_6H_4 are -0.022, 0.291, 0.001 for SE, FI and OI, respectively. For the o-C_6H_4, these values are -0.072, 0.139 and -0.006. Such small differences indicate the transferability of the C_6H_4 group with respect to various positions for binding substituent.

Recalling the relation between the IT measures in the shape representation and the spin-density representation, $S[\rho] = N\left(S[\sigma_N] + \ln 2 - \ln N\right)$, $I[\rho] = N\, I[\sigma_N]$ and $D[\rho] = N^2\, D[\sigma_N]/2$, it is obvious why each linear relation, Eq.(24), is not observed in the shape representation (see Fig.3, left panels).

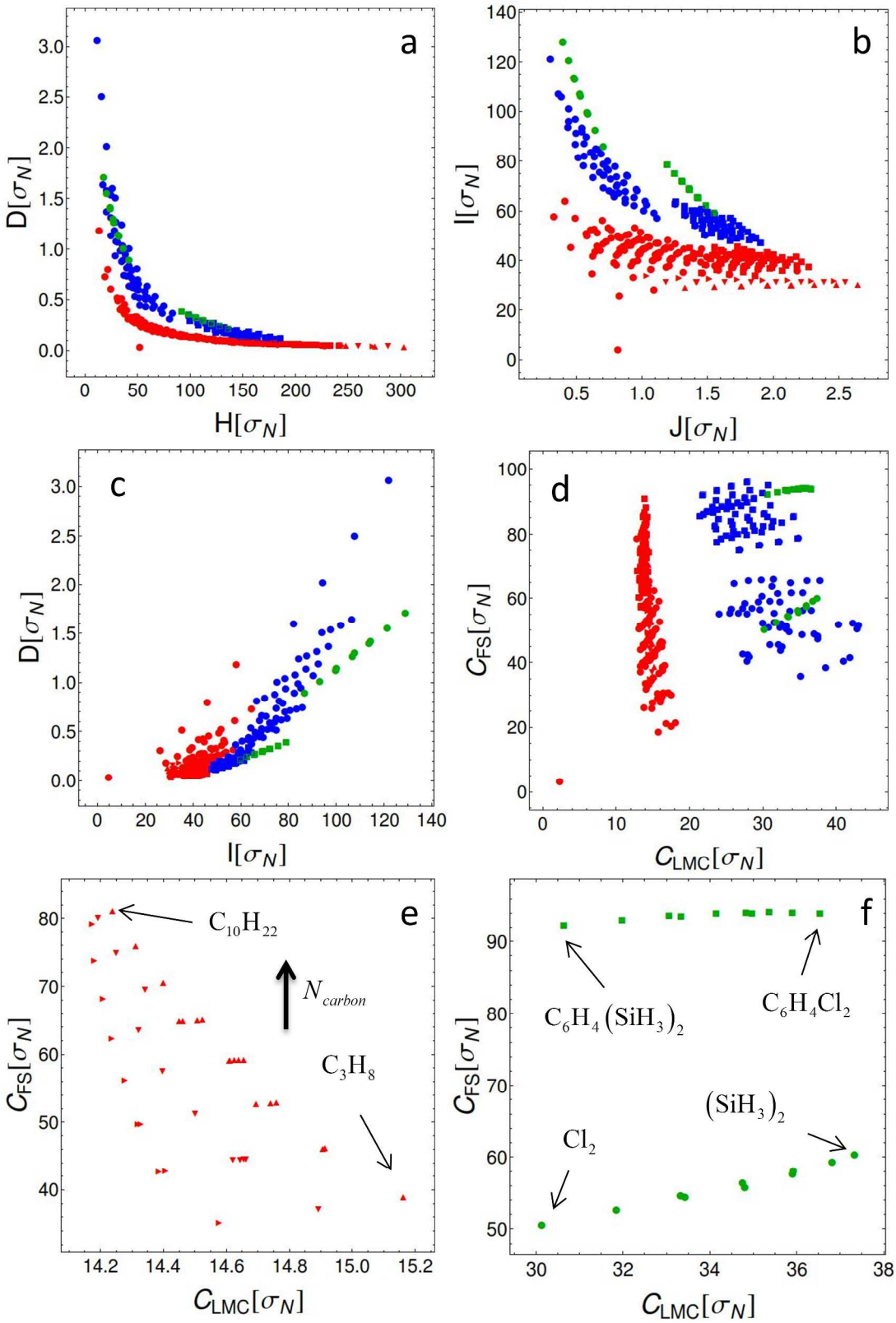

Fig.4 The IT measures and complexities planes in the shape representation (panels a-d). The inconsistence between $C_{\mathrm{LMC}}[\sigma_{\mathrm{N}}]$ and $C_{\mathrm{FS}}[\sigma_{\mathrm{N}}]$ for the A3 set (panel e) and for the X–Y-III and d-XY-III sets (panel f). See Fig.1 for details.

Table IV The Shannon entropy, the Fisher information, the Onicescu information, the electron-electron repulsion energy, the electron-nucleus attraction energy, for X–Y molecules with $X,Y \in \{-SiH_3,-PH_2,-SH,-Cl\}$ and for hexane isomers. All data in a.u.

	$S[\rho]$	$I[\rho]$	$D[\rho]$	$E_{ee}=E_{es}+E_{xc}$	E_{en}
X–Y molecule					
SiH3–SiH3	30.18	2939.61	522.46	312.99	-1566.62
SiH3–PH2	25.57	3161.13	589.89	331.54	-1687.13
PH2–PH2	21.05	3382.89	657.33	349.09	-1805.64
SiH3–SH	20.47	3399.77	668.43	351.50	-1818.38
PH2–SH	16.09	3621.68	735.88	368.04	-1934.82
SiH3–Cl	15.08	3656.17	759.16	372.28	-1959.33
SH–SH	11.23	3860.45	814.47	386.22	-2062.51
PH2–Cl	10.82	3878.24	826.64	388.13	-2074.39
SH–Cl	6.08	4117.09	905.20	405.63	-2200.76
Cl–Cl	0.95	4373.62	995.95	425.21	-2339.49
hexane isomers (C_6H_{14})					
3-methylpentane	102.30	1517.54	94.71	318.53	-1034.03
2,3-dimethylbutane	102.24	1517.21	94.72	327.93	-1052.83
isohexane	102.19	1517.07	94.71	330.62	-1058.19
2,2-dimethylbutane	102.15	1516.78	94.72	336.82	-1070.60
hexane	102.14	1516.64	94.72	340.17	-1077.32

III.B. The information theory planes and complexities.

The concept of the complexity seems to be very attractive in the chemical reactivity investigations. To the best of our knowledge, there is only one example of the complexity analysis from the molecular reactivity point of view.[39, 40] To begin investighation, it is interesting to analyze the relation between the complexities components. This is done in Fig. 4 by means of the information planes. On the $D-H$ plane, Fig.4a, we notice an approximate $D \sim 1/H$ relation between the entropy power, $H[\sigma_N]$, and the disequilibrium, $D[\sigma_N]$ for a subset of molecules without the III row atoms (the red points, the unfit point is the hydrogen molecule), yielding values close to 16 for the $C_{LMC}=DH$ complexity (see the $C_{FS}-C_{LMC}$ plane, Fig.4d). A larger dispersion around $1/H$ behavior for the molecules containing the III-row atoms (blue and green points) results in a wide range of the $C_{LMC}[\sigma_N]$ values. Although the values for the benzene derivatives, the *d*-XY set (blue and green squares) are located at $H>85$, while their X–Y set counterparts (blue and green circles) are at $H<85$, nevertheless this results in the same wide range of C_{LMC} values (Fig.4d). In the case of the C_{FS} complexity, $C_{FS}=IJ$, the range of the values for the molecules

indicated by the red points in Fig.4d is significantly wider than in the C_{LMC} case, but for the molecules marked by the green squares (the benzene derivative with heavy atom in substituent group only from III rows) is quite narrow.

The inconsistency between two complexities can be easily shown on the example of the A3 set. Insertion of the relations from Eq.(22) into the complexities definitions yields

$$C_{\mathrm{LMC}}\left[\sigma_{\mathrm{N}}\right]=\left(\frac{a_{\mathrm{D}}N+b_{\mathrm{D}}}{N}\right)\exp\left(a_{\mathrm{S}}+\frac{b_{\mathrm{S}}}{N}\right),$$
$$C_{\mathrm{FS}}\left[\sigma_{\mathrm{N}}\right]=\frac{\left(a_1N+b_1\right)N^{2/3}}{2^{1/3}\pi eN}\exp\left(\frac{2}{3}\left(a_{\mathrm{S}}+\frac{b_{\mathrm{S}}}{N}\right)\right), \quad (25)$$

where a and b are the fitting coefficients for a given measure, $F\in\{S,I,D\}$, Table III. As the hydrocarbon size N increases, the C_{LMC} complexity decreases while the C_{FS} complexity increases, two opposite effects (Fig.4e).

Another example is the complexity for the X–Y-III set and the d-X–Y-III set, in Fig.4f. The molecules from the X–Y-III set are isoelectronic molecules with $N=34$ electrons. Assuming that the A molecule has higher complexity than the B molecule, the following relation occurs for the C_{LMC} complexity

$$C_{\mathrm{LMC}}\left[\sigma_{\mathrm{N}}^{A,\mathrm{XY}}\right]>C_{\mathrm{LMC}}\left[\sigma_{\mathrm{N}}^{B,\mathrm{XY}}\right]$$
$$\Leftrightarrow\left(S\left[\sigma_{\mathrm{N}}^{A,\mathrm{XY}}\right]-S\left[\sigma_{\mathrm{N}}^{B,\mathrm{XY}}\right]\right)>\ln\frac{D\left[\sigma_{\mathrm{N}}^{B,\mathrm{XY}}\right]}{D\left[\sigma_{\mathrm{N}}^{A,\mathrm{XY}}\right]} \quad (26)$$
$$\Leftrightarrow\left(S\left[\rho_{\mathrm{XY}}^{A}\right]-S\left[\rho_{\mathrm{XY}}^{B}\right]\right)>N_{\mathrm{XY}}\ln\frac{D\left[\rho_{\mathrm{XY}}^{B}\right]}{D\left[\rho_{\mathrm{XY}}^{A}\right]},$$

and for the C_{FS} complexity

$$C_{\mathrm{FS}}\left[\sigma_{\mathrm{N}}^{A,\mathrm{XY}}\right]>C_{\mathrm{FS}}\left[\sigma_{\mathrm{N}}^{B,\mathrm{XY}}\right]$$
$$\Leftrightarrow\frac{2}{3}\left(S\left[\sigma_{\mathrm{N}}^{A,\mathrm{XY}}\right]-S\left[\sigma_{\mathrm{N}}^{B,\mathrm{XY}}\right]\right)>\ln\frac{I\left[\sigma_{\mathrm{N}}^{B,\mathrm{XY}}\right]}{I\left[\sigma_{\mathrm{N}}^{A,\mathrm{XY}}\right]} \quad (27)$$
$$\Leftrightarrow\left(S\left[\rho_{\mathrm{XY}}^{A}\right]-S\left[\rho_{\mathrm{XY}}^{B}\right]\right)>\frac{3}{2}N_{\mathrm{XY}}\ln\frac{I\left[\rho_{\mathrm{XY}}^{B}\right]}{I\left[\rho_{\mathrm{XY}}^{A}\right]}.$$

Using the data from Table IV, the lowest complexity in the X–Y-III set is for chlorine molecule and the highest for disilane (the complexity for the X–Y-III molecules decreases). The relations from Eqs.(26) and (27) for the benzene derivative at position *para* can be rewritten using Eq.(24)

$$C_{\mathrm{LMC}}\left[\sigma_{\mathrm{N}}^{A,p-\mathrm{XY}}\right]>C_{\mathrm{LMC}}\left[\sigma_{\mathrm{N}}^{B,p-\mathrm{XY}}\right]\Leftrightarrow$$
$$0.997\left(S\left[\rho_{\mathrm{XY}}^{A}\right]-S\left[\rho_{\mathrm{XY}}^{B}\right]\right)>N_{p-\mathrm{XY}}\ln\frac{D\left[\rho_{\mathrm{XY}}^{B}\right]+94.430}{D\left[\rho_{\mathrm{XY}}^{A}\right]+94.430}, \quad (28)$$

$$C_{\mathrm{FS}}\left[\sigma_{\mathrm{N}}^{A,p-\mathrm{XY}}\right]>C_{\mathrm{FS}}\left[\sigma_{\mathrm{N}}^{B,p-\mathrm{XY}}\right]\Leftrightarrow$$
$$0.997\left(S\left[\rho_{\mathrm{XY}}^{A}\right]-S\left[\rho_{\mathrm{XY}}^{B}\right]\right)>\frac{3}{2}N_{p-\mathrm{XY}}\ln\frac{I\left[\rho_{\mathrm{XY}}^{B}\right]+1477}{I\left[\rho_{\mathrm{XY}}^{A}\right]+1477}, \quad (29)$$

and with data from Table IV, the lowest complexity is for dichlorobenzene and the highest for disilylbenzene (the complexity for the p-X–Y-III molecules has the opposite ordering as compared to the X–Y-III molecules). With almost the same range of C_{LMC} for both sets, the C_{FS} values are well separated, Fig.4f.

In the spinor density representation the IT planes possess more pattern and organization information than those presented in Fig. 4. In Fig.5a-c, the information planes between the three IT based measures are plotted. We can observe similar patterns between the X–Y set and the d-XY set (circles and squares, respectively), which can be interpreted as the C_6H_4 group transferability effect. A linear behavior can be observed for molecules with the same X and different Y group and are isoelectronic molecules, e.g. green points which are related to the X–Y-III and d-XY-III set molecules. For such molecules, ordered as is written in Table IV, we observe the decrease in uncertainty, $S[\rho]$, accompanied by increase of order and organization, $D[\rho]$ and $I[\rho]$, respectively. We can see that for the molecular set used in this work, the hydrocarbons from the A3 set and hydrogen molecule are the lower bounds in all IT planes. Comparing the planes presented in Fig.5, it can be observed that the $S-I$ plane provides "richer" information about the pattern, organization, similarity of molecules than the other planes. Note that the strong correlation between SE and the FI information observed for the atoms[74] and the hydrocarbons,[75] does not reflect a general relation (see Fig.5a).

Let us consider separately the isolelectronic systems, taking as an example the X–Y-III set and the hexane isomers, Table IV. The FI value increases with increasing sum of Z_A^2 (see the *K*-shell model from Eq.(20)), in particular, hexane isomers have a lower FI value than the X–Y-III set molecules. The sharpness, concentration or delocalization of the electronic cloud is measured by the SE as well by the Fisher information (FI). In the case of the X–Y-III set, entropy decreases significantly with decreasing number of hydrogens (decreasing number of bonds). In the case of the isosters, the entropy is greater for the system with the higher symmetry, eg. diphospane and silanethiol, 3-

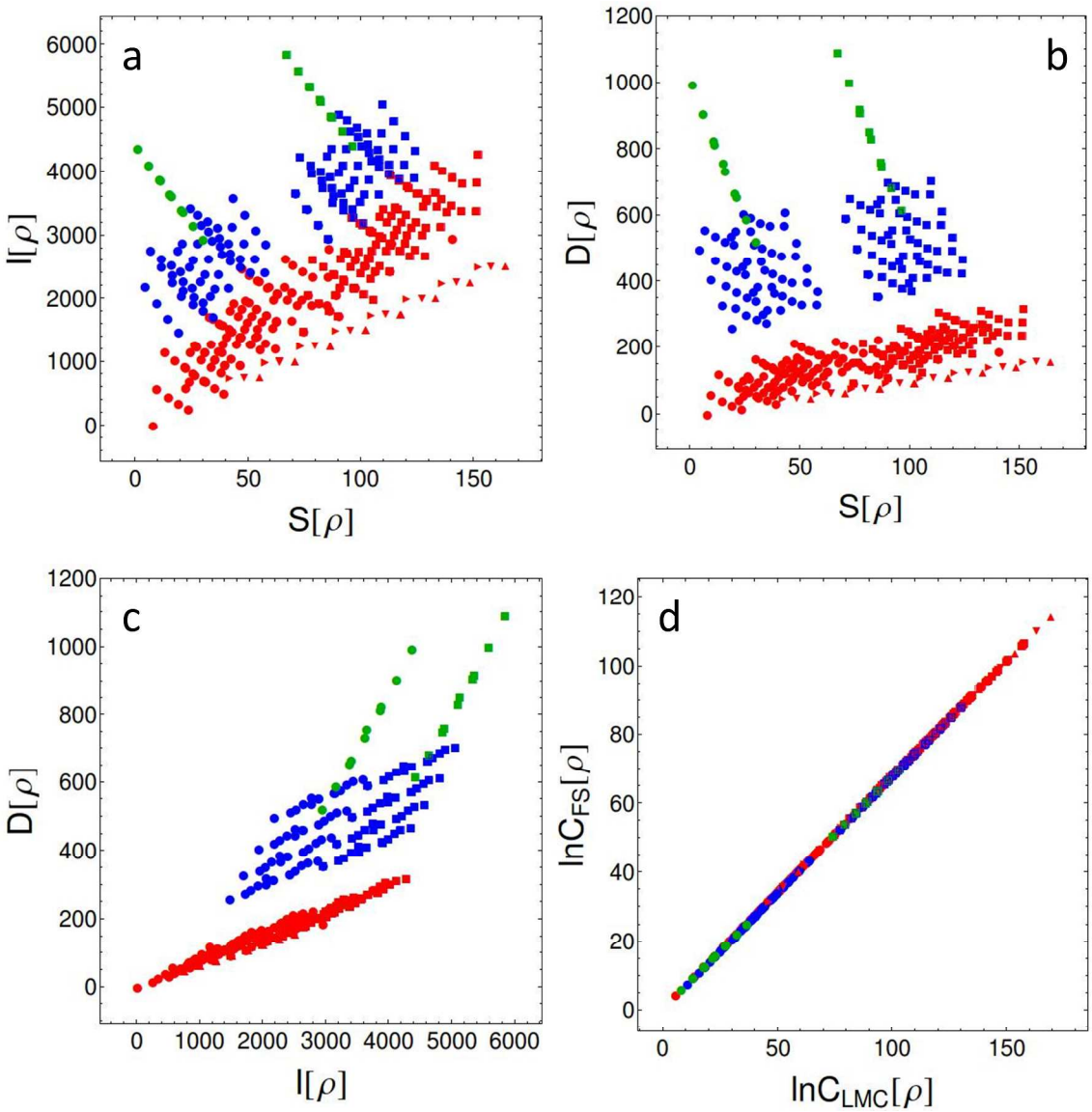

Fig.5 The IT measure planes (panels a-c) and the plane between logarithms of $C_{\mathrm{LMC}}[\rho]$ and $C_{\mathrm{FS}}[\rho]$ (panel d). All panels are in the spinor density representation. See Fig.1 for details.

methylpentane and hexane. The results for the entropy in the density representation are consistent with the interpretation that increasing of the electron-nucleus attraction energy $|E_{\mathrm{en}}|$ (reflecting the "structure" addition to the distribution) is thereby lowering the entropy, whereas decreasing of the electron-electron repulsion energy would raise the entropy[25] (in Table IV, the energy components for the X–Y-III set and hexane isomers are shown). Comparing the hexane data (Table IV), the OI is the less effective tool to conformers distinction. It is worth mentioning that the discussed regularities are for the systems with the same electron number so these conclusions are also valid in the shape representation. More discussion about the group effect on the IT can be find elsewhere.[76]

Unfortunately, the extension of the definition of complexities used here to the spinor-density dependent functionals is inconvenient because it yields complexity values as very large numbers.Therefore logarithms of them are conveniet for analysis. In Fig.5d, an excellent linear relation between the logarithms of both complexities is observed. However, this linearity is due to the dominant role of the SE dependent factors of both complexities. When rewritten as $S[\rho] = a\ln(I[\rho]) + b\ln(D[\rho]) + c$, it is satisfied with very low accuracy.

We can conclude these two parts that the shape function

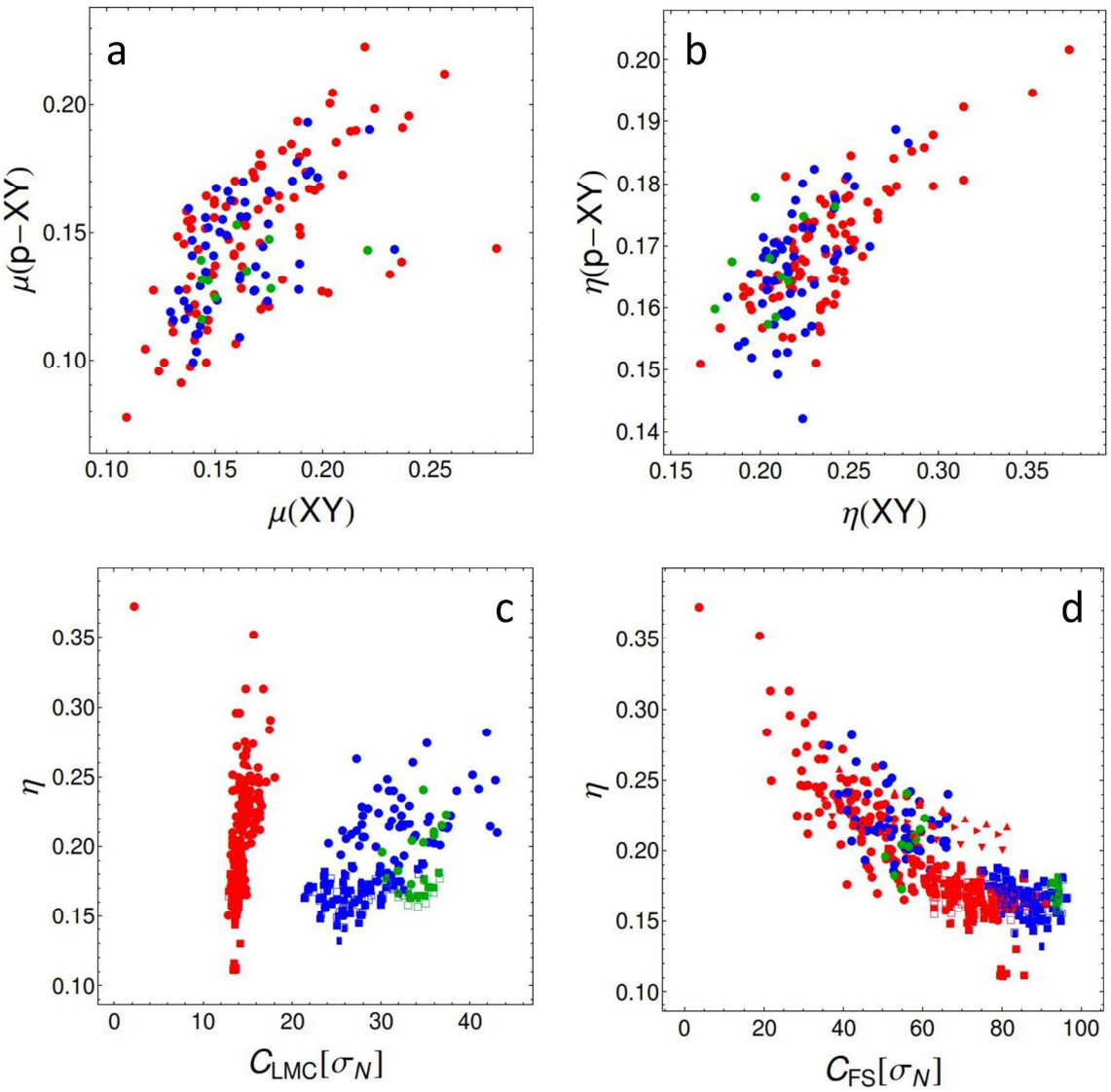

Fig.6 The lack of the group transferability for the chemical potential and hardness on the $p-C_6H_6$ example in the spinor density representation (panels a-b). The chemical hardness vs. complexities in the shape representation (panels c-d). See Fig.1 for details.

contains less information than the spinor density and this can yield wrong conclusion when the information measure is used for the system with different electron numbers. Secondly, both complexities used in this work do not give chemically reasonable descriptions.

III.C. The IT measures and reactivity indices.

In this part, we would like to find some correlation between the IT measures and the chemical reactivity indices. When the transferability concept is valid for the IT measures, the correlated properties should also possess such properties as a desired condition for obtaining a good correlation, e.g.

$$\begin{aligned} R[\rho_{d-XY}] &= aF[\rho_{d-XY}]+b \\ &= aF[\rho_{XY}]+b+a\left(F[\rho_{d-XY}]-F[\rho_{XY}]\right) \\ &= R[\rho_{XY}]+aF\left[\rho_{d-C_6H_4}\right]=R[\rho_{XY}]+R\left[\rho_{d-C_6H_4}\right] \end{aligned} \quad (30)$$

where $R[\rho]$ is a correlated property, $d-C_6H_4$ is the C_6H_4 group lacking hydrogen atoms at *d* position.

hardness is illustrated in Fig.6a-b. Only a general decreasing trend can be observed between the hardness and the SE or the FI. In the shape representation these relations are less visible. The observed trends between hardness and complexities are rather inconsistent. In the C_{LMC} complexity case, there are no direct relations, e.g. the molecules with non-hydrogen atoms from II row have very close complexity but wide hardness range (Fig.6c).

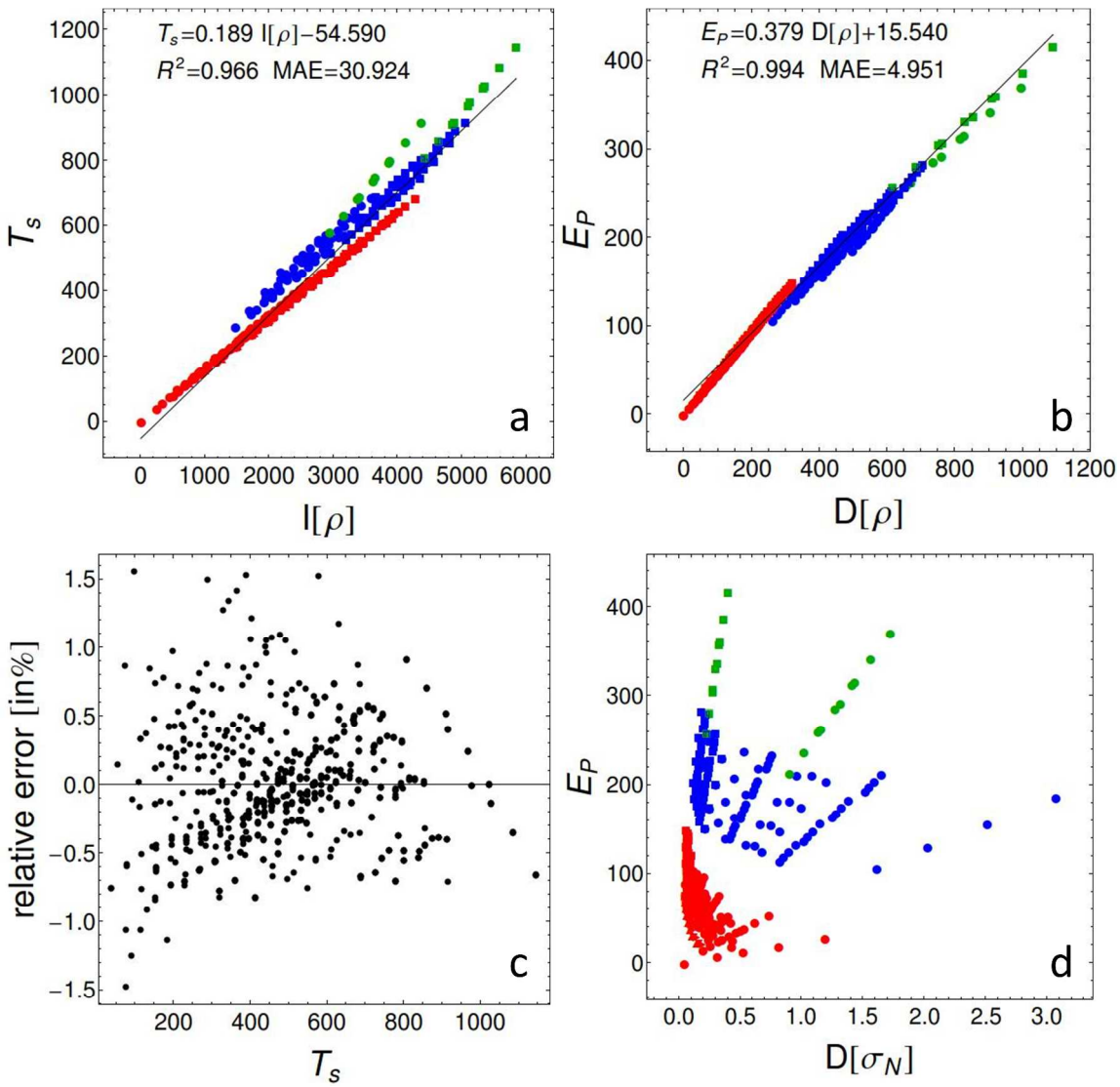

Fig.7 The correlation between the kinetic energy of non-interacting electrons and the Fisher information (T_s vs. $I[\rho]$, panel a); between the Pauli energy and the Onicescu information (E_P vs. $D[\rho]$, panel b); the relative error $(\tilde{T}_s - T_s)/T_s$ vs. T_s, of $\tilde{T}_s$ fitting, Eq.(31) (panel c); between the Pauli energy and the Onicescu information in the shape representation (E_P vs. $D[\sigma_N]$, panel d). See Fig.1 for details.

The hardness vs. C_{FS} complexity shows a general decreasing trend (Fig.6d).

In contrary to the electronic energy, the total energy shows the group transferability, thus the analysis of IT measures versus the total energy and its components was performed. Only the FI shows a linear behavior with the total energy, the kinetic energy and the nucleus-electron attraction energy with correlation coefficients 0. 967, 0.967 and 0.947, respectively. This behavior can be explained by recalling that the expectation value of the kinetic energy of non-interacting electrons of the Kohn-Sham problem is proportional to the Fisher information.[77, 78] When the virial theorem is satisfied with reasonable accuracy, the linear relation between the FI and the total energy is no surprise. The main contribution to the FI comes from each region close to nucleus, the same is true in the case of the nucleus-electron attraction energy. Relations between the two other measures and the energy and its components have the structural behavior related to the chemical group transferability.[76]

In addition to the linear behavior of T_s vs. $I[\rho]$ characterized by $R^2 = 0.966$ (Fig.7a), surprisingly, we observe a similar linear behavior in the case of the Pauli energy vs. the Onicescu information (disequilibrium) with $R^2 = 0.994$ (see Fig.7.b). These two linearities combined with the Pauli energy definition, Eq.(14), suggest T_s to be linear in $I[\rho]$ and $D[\rho]$. A multi-linear regression (through the origin) between the kinetic energy

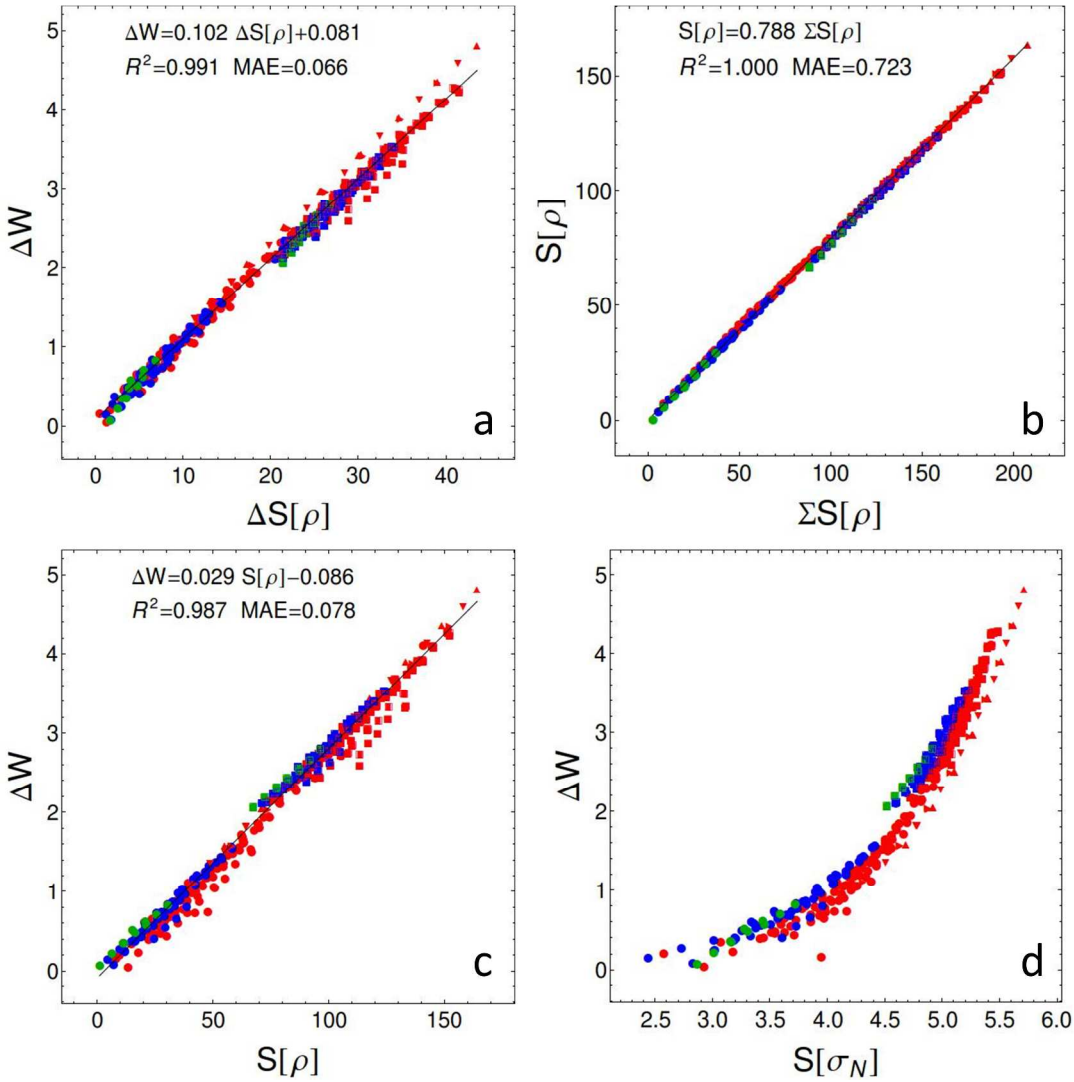

Fig.8 The correlation between: the atomization energy and the atomization entropy (ΔW vs. $\Delta S[\rho]$, panel a); the molecular entropy and the sum of atomic entropies ($S[\rho]$ vs. $\sum_i S[\rho]$, panel b); the atomization energy and the molecular entropy (ΔW vs. $S[\rho]$, panel c); the atomization energy and the molecular entropy in the shape representation (ΔW vs. $S[\sigma_N]$, panel d). See Fig.1 for details.

and the FI and the OI was calculated

$$\tilde{T}_s = 0.1347\, I[\rho] + 0.3367\, D[\rho], \quad R^2 = 1.0000,\ \text{MAE}=1.5745 \qquad (31)$$

The plot of the relative error, $\left(\tilde{T}_s - T_s\right)/T_s$ vs. T_s of this fitting is presented in Fig.7c. It is very small (except the hydrogen molecule error, equal -10%, not displayed). This fitting, Eq.(31), due to the relation $T_W = I/8$ can be rewritten for the Pauli energy as $\tilde{E}_P = 0.0097\, I[\rho] + 0.3367\, D[\rho]$. Since the virial theorem is approximately satisfied in DFT, Eq.(31) provides a model for the total energy $W \approx -T_s$ with quite good accuracy. It is noteworthy that these linear statistical models disappear in the shape representation, e.g. compare E_P vs. $D[\sigma_N]$, Fig.7d, with E_P vs. $D[\rho]$, Fig.7b.

Another interesting regression found, is the atomization energy ΔW vs. the atomization entropy $\Delta S[\rho]$, Fig.8a. The atomization energy is the energy required to split a molecule into its constituent atoms, in other words to broke all the bonds, Eq.(17). The atomization entropy is defined as

$$\Delta S[\rho] = \sum_{A \in M} S\left[\rho_A^0\right] - S\left[\rho_M\right] \qquad (32)$$

where ρ_M is the molecule density and ρ_A^0 is the electron density computed for the isolated atom A. This ΔS is a modification of the SE due to the atomization (or the negative value of the SE change due to the formation of chemical bonds). The trend observed in Fig.8a is in line with the interpretation of the entropy

change during the bond formation from the promolecule.[79] As a consequence of the observed linear relation between the sum of atomic entropies and the molecular entropy, Fig.8b, a linear relation between the atomization energy and the molecular entropy is observed, Fig.8c. A positive slope indicates that the increase of molecular SE is accompanied by the increase of the energetical stability ΔW. However, in the case of the constitutional isomers, e.g. the benzene derivatives for a given X and Y substituents, these observations are deemed to be inconsistent. When the sum of atomic quantities is the same for constitutional isomers, from Fig.8a and Eq.(32) follows $S_A > S_B \Rightarrow \Delta S_A < \Delta S_B \Rightarrow \Delta W_A < \Delta W_B$. In other words, if the entropy of molecule A is higher than the entropy of molecule B, then molecule B is more stable than molecule A, in contradiction with the observed relation in Fig.8c. This indicates that the linear relation ΔW vs. $\Delta \mathrm{S}[\rho]$ is only a theoretical idealization of a general trend and this relation should not be used to compare molecules with small entropy difference.

Conclusions

The analysis of the information and complexity measures as the tool in the investigation of the chemical reactivity and similarity has been done for the density representation in the spin-position space and the shape representation in the position space. The used set of molecules was enough large and diverse to improve the previous understanding of IT measures in chemical applications and to generalize the obtained conclusions. The concepts of the transferability and addititivity of atoms or functional groups were used as "checkpoints" in analysis of the obtained results. It was shown that in the shape representation, observed trends are in contradiction with chemical intuition, e.g. the FI value for alkynes/alkenes is always higher than for alkenes/ alkenes irrespectively of the number of carbon atoms they contain. The perfect linear relation between the XY molecules and benzene with the X and Y substituents observed in the spinor-density representation is devastated in the shape representation. In the case of the hydrocarbons investigated in this work, the entropy trend is opposite to the disequilibrium trend for a given carbon number. When the OI is considered as a finer measure of dispersion distribution than the SE, the previous observation is in contradiction with the meaning of both measures. The FI results are inconsistent with the requirements that the steric effect needs to be extensive in size (the FI for alkenes are almost constant with increasing size in the shape representation). The obtained values for the C_{LMC} and the C_{FS} complexities were mutually inconsistent, e.g. (i) in the case of A3 set, with increasing hydrocarbon size C_{LMC} is decreasing and while C_{FS} is increasing; (ii) the complexity trend observed for the changing substituents of the X–Y-III set is opposite to trend observed for the benzene derivatives, the d-X–Y-III set (the lowest complexity in the X–Y-III set is for chlorine molecule and the highest for disilane, while in the case of the d-X–Y-III set, the lowest complexity is for dichlorobenzene and the highest for disilylbenzene). In general, we can conclude that the shape function contains less information than the spinor density. Use of the shape function can lead to wrong conclusions when the information measure is analizeed for the systems with different electron numbers.

In contrary to the results from the shape representation, the measures S, I, D in the spinor density representation show the transferability and additivity. The group transferability is well illustrated on the example of the X–Y molecules and their benzene derivatives. As it was shown, a perfect linear relation exists between the SE for the X–Y molecules and the SE for their derivatives. Another example is the methylene group transferability presented on the alkane-alkene-alkyne set. Analysis of the information planes between the three IT based measures has shown that the $S-I$ plane provides "richer" information about the pattern, organization, similarity of used molecules than $S-D$ and $I-D$ planes. In the spinor-density representation, the C_{LMC} complexity and the C_{FS} complexity are inconvenient quantities due to their large values. They do not provide a chemically reasonable description of the molecular complexity. Unfortunately, no general relations between the IT measures and the chemical reactivity indices (like the chemical potential and the chemical hardness) are observed.

Surprisingly, in addition to the linear behavior of T_{s} vs. $I[\rho]$, a similar linear behavior in the case of the Pauli energy vs. the Onicescu information (disequilibrium) is observed. Based on these relations, the highly accurate linear relation is noted between the kinetic energy and the FI and the OI measures. Another interesting regression was found between the atomization energy and the atomization entropy. To the best of our knowledge, these linear behaviors have not been observed before. Finally, the atomic additivity in the spinor density

representation proves the advantages of this representation over the shape representations. Of course the shape representation is still useful in work with systems which have the same electron number, e.g. for the reaction path analysis, quantifying **c**hirality.[80]

The results obtained in this work reveal that the IT measures can be used in the chemical reactivity investigation, but only as the source of the information about the pattern, organization, similarity of molecules, while not as direct indicators of their reactivity. In efforts to improve our understanding of the IT measures in the chemical applications, it seems necessary to take into account the substituent group influence on the IT measures and their relations with the reactivity change of substituted molecule.

Acknowledgment

This work was supported by the International PhD Projects Programme of the Foundation for Polish Science, co-financed from European Regional Development Fund within Innovative Economy Operational Programme "Grants for innovation" and the Interdisciplinary Center for Mathematical and Computational Modeling computational grant

Appendix 1

Relations between the IT measures calculated in the spin-position and the position spaces for the density and shape representations are shortly presented. The relation between these functionals can be obtained with the help of the related functionals of the spin density components, $\rho_\alpha(\mathbf{r})$ and $\rho_\beta(\mathbf{r})$.

Each considered spinor density functional ($S[\rho]$, $I[\rho]$, $D[\rho]$) happens to be a local functional of $\rho(\mathbf{x})$ and $\nabla\rho(\mathbf{x})$, $F[\rho]=\int f\left(\rho(\mathbf{x}),\nabla\rho(\mathbf{x})\right)d\mathbf{x}$, therefore it is a sum of spin-component functionals

$$F[\rho]=\sum_\kappa\int f\left(\rho(\mathbf{r},\kappa),\nabla\rho(\mathbf{r},\kappa)\right)d\mathbf{r}=F[\rho_\alpha]+F[\rho_\beta]. \quad \text{(A.33)}$$

Analogous relation for the spinless density, Eq.(2), can be written formally as

$$F[\rho_{\mathrm{N}}]=F[\rho_\alpha]+F[\rho_\beta]+\tilde{F}\left[\rho_{\mathrm{N}}(\mathbf{r}),\rho_\alpha(\mathbf{r}),\rho_\beta(\mathbf{r})\right], \quad \text{(A.34)}$$

where $\tilde{F}\left[\rho_{\mathrm{N}}(\mathbf{r}),\rho_\alpha(\mathbf{r}),\rho_\beta(\mathbf{r})\right]\equiv F[\rho_{\mathrm{N}}]-\left(F[\rho_\alpha]+F[\rho_\beta]\right)$.

The transformation to the shape functionals, Eq.(5), goes via the relation

$$F[\zeta]=F\left[N_\zeta\sigma_\zeta\right]. \quad \text{(A.35)}$$

In terms of $\sigma_\zeta(\mathbf{q})=\zeta(\mathbf{q})/N_\zeta$, the following relations hold

$$\sigma(\mathbf{r},\kappa)=\frac{1}{N}\rho_\kappa(\mathbf{r})=\frac{N_\kappa}{N}\sigma_\kappa(\mathbf{r}) \quad , \quad \text{i.e.} \quad \sigma(\mathbf{r},\kappa)\neq\sigma_\kappa(\mathbf{r}) \quad \text{(A.36)}$$

$$\sigma_{\mathrm{N}}(\mathbf{r})=\frac{1}{N}\left(\rho_\alpha(\mathbf{r})+\rho_\beta(\mathbf{r})\right)=\sigma_\alpha(\mathbf{r})\frac{N_\alpha}{N}+\sigma_\beta(\mathbf{r})\frac{N_\beta}{N}=\sum_\kappa\sigma(\mathbf{r},\kappa) \quad \text{(A.37)}$$

Basing on these relations, we can pass between different representations and spaces (here N and N_κ are also involved)

$$\begin{array}{ccccc}\rho & \leftrightarrow & \{\rho_\alpha,\rho_\beta\} & \leftrightarrow & \rho_{\mathrm{N}}\\ \updownarrow & & \updownarrow & & \updownarrow \\ \sigma & \leftrightarrow & \{\sigma_\alpha,\sigma_\beta\} & \leftrightarrow & \sigma_{\mathrm{N}}\end{array} \quad \text{(A.38)}$$

The relations between the Shannon entropies (SEs) in the density representations are

$$\begin{aligned}S[\rho]&=S[\rho_\alpha]+S[\rho_\beta],\\ S[\rho_{\mathrm{N}}]&=S[\rho]+\tilde{S}_{\mathrm{KL}}[\rho_\alpha,\rho_{\mathrm{N}}]+\tilde{S}_{\mathrm{KL}}[\rho_\beta,\rho_{\mathrm{N}}],\end{aligned} \quad \text{(A.39)}$$

where

$$\tilde{S}_{\mathrm{KL}}[\rho_\kappa,\rho_{\mathrm{N}}]\equiv\int\rho_\kappa(\mathbf{r})\ln\frac{\rho_\kappa(\mathbf{r})}{\rho_{\mathrm{N}}(\mathbf{r})}d\mathbf{r}, \quad \text{(A.40)}$$

known as the Kullback-Leibler entropy.[75]

Due to the properties that $\rho_\kappa(\mathbf{r})/\left(\rho_\alpha(\mathbf{r})+\rho_\beta(\mathbf{r})\right)\le 1$ and $\rho_\kappa(\mathbf{r})\ge 0$, $\tilde{S}_{\mathrm{KL}}[\rho_\kappa,\rho_{\mathrm{N}}]$ is non-positive quantity, so $S[\rho]\ge S[\rho_{\mathrm{N}}]$. Only in the fully spin-polarized case, $\rho_\alpha(\mathbf{r})=\rho_{\mathrm{N}}(\mathbf{r})$, $\rho_\beta(\mathbf{r})=0$, these entropies are equal).

The Fisher informations (FIs) are related as follows

$$\begin{aligned}I[\rho]&=I[\rho_\alpha]+I[\rho_\beta],\\ I[\rho_{\mathrm{N}}]&=I[\rho]-\int\frac{\rho_\alpha(\mathbf{r})\rho_\beta(\mathbf{r})}{\rho_{\mathrm{N}}(\mathbf{r})}\left|\nabla\ln\left(\frac{\rho_\beta(\mathbf{r})}{\rho_\alpha(\mathbf{r})}\right)\right|^2 d\mathbf{r}.\end{aligned} \quad \text{(A.41)}$$

so $I[\rho]\ge I[\rho_{\mathrm{N}}]$

The Onicescu informations (OIs) (the disequilibriums) are related as

$$\begin{aligned}D[\rho]&=D[\rho_\alpha]+D[\rho_\beta],\\ D[\rho_{\mathrm{N}}]&=D[\rho]+2\int\rho_\alpha(\mathbf{r})\rho_\beta(\mathbf{r})d\mathbf{r}\end{aligned}, \quad \text{(A.42)}$$

so $D[\rho_{\mathrm{N}}]\ge D[\rho]$.

To pass from the density representation to the shape

representation, Eq.(A.35) is used,

$$S[\sigma_\zeta] = S[\zeta]/N_\zeta + \ln N_\zeta , \tag{A.43}$$

$$I[\sigma_\zeta] = I[\zeta]/N_\zeta , \tag{A.44}$$

$$D[\sigma_\zeta] = D[\zeta]/N_\zeta^2 . \tag{A.45}$$

The relations between the spinor-shape functionals and spinless-shape functionals are more complicated than their counterparts in the density representation. For the SE relations, we have

$$\begin{aligned} S[\sigma] &= \sum_\kappa \frac{N_\kappa}{N}\left(S[\sigma_\kappa] - \ln N_\kappa\right) + \ln N, \\ S[\sigma_{\mathrm{N}}] &= S[\sigma] + \sum_\kappa \frac{N_\kappa}{N}\left(\tilde{S}_{\mathrm{KL}}[\sigma_\kappa,\sigma_{\mathrm{N}}] + \ln\frac{N_\kappa}{N}\right) \\ &= \sum_\kappa \frac{N_\kappa}{N}\left(S[\sigma_\kappa] + \tilde{S}_{\mathrm{KL}}[\sigma_\kappa,\sigma_{\mathrm{N}}]\right), \end{aligned} \tag{A.46}$$

see Eq.(A.40) for $\tilde{S}_{\mathrm{KL}}$. Since $\sigma_\alpha(\mathbf{r})$ is less (greater) than $\sigma_{\mathrm{N}}(\mathbf{r})$ when $\sigma_\alpha(\mathbf{r})$ is less (greater) than $\sigma_\beta(\mathbf{r})$, the sign of $\tilde{S}_{\mathrm{KL}}[\sigma_\kappa,\sigma_{\mathrm{N}}]$ is unpredictable.

The relations for FI are as follows

$$\begin{aligned} I[\sigma] &= \frac{N_\alpha}{N} I[\sigma_\alpha] + \frac{N_\beta}{N} I[\sigma_\beta], \\ I[\sigma_{\mathrm{N}}] &= I[\sigma] - \frac{N_\alpha N_\beta}{N^2}\int \frac{\sigma_\alpha(\mathbf{r})\sigma_\beta(\mathbf{r})}{\sigma_{\mathrm{N}}(\mathbf{r})}\left|\nabla \ln\left(\frac{\sigma_\beta(\mathbf{r})}{\sigma_\alpha(\mathbf{r})}\right)\right|^2 d\mathbf{r}, \end{aligned} \tag{A.47}$$

and the relations for OI are

$$\begin{aligned} D[\sigma] &= \frac{1}{N^2}\left(N_\alpha^2 D[\sigma_\alpha] + N_\beta^2 D[\sigma_\beta]\right), \\ D[\sigma_{\mathrm{N}}] &= D[\sigma] + 2\frac{N_\alpha N_\beta}{N^2}\int \sigma_\alpha(\mathbf{r})\sigma_\beta(\mathbf{r})\, d\mathbf{r}. \end{aligned} \tag{A.48}$$

Using Eq.(A.39), the relations for the exponential SE in the density representation are

$$\begin{aligned} H[\rho] &= H[\rho_\alpha]H[\rho_\beta], \\ H[\rho_{\mathrm{N}}] &= H[\rho]\exp\left(\sum_\kappa \tilde{S}_{\mathrm{KL}}[\rho_\kappa,\rho_{\mathrm{N}}]\right), \end{aligned} \tag{A.49}$$

and

$$\begin{aligned} J[\rho] &= (2\pi e) J[\rho_\alpha]J[\rho_\beta], \\ J[\rho_{\mathrm{N}}] &= J[\rho]\exp\left(\frac{2}{3}\sum_\kappa \tilde{S}_{\mathrm{KL}}[\rho_\kappa,\rho_{\mathrm{N}}]\right). \end{aligned} \tag{A.50}$$

In the shape representation, using Eq.(A.46)

$$\begin{aligned} H[\sigma] &= N\prod_\kappa \left(H[\sigma_\kappa]/N_\kappa\right)^{\frac{N_\kappa}{N}}, \\ H[\sigma_{\mathrm{N}}] &= H[\sigma]\prod_\kappa \left(\frac{N_\kappa}{N}\exp\left(\tilde{S}_{\mathrm{KL}}[\sigma_\kappa,\sigma_{\mathrm{N}}]\right)\right)^{\frac{N_\kappa}{N}} \end{aligned} \tag{A.51}$$

and

$$\begin{aligned} J[\sigma] &= N\left(\frac{1}{2\pi e}\right)^{1-\frac{N_\alpha N_\alpha}{N^2}} \prod_\kappa \left(N_\kappa^{-2/3} J[\sigma_\alpha]\right)^{\frac{N_\alpha}{N}}, \\ J[\sigma_{\mathrm{N}}] &= J[\sigma]\exp\left(\sum_\kappa \frac{N_\kappa}{N}\left(\tilde{S}_{\mathrm{KL}}[\sigma_\kappa,\sigma_{\mathrm{N}}] + \ln\frac{N_\kappa}{N}\right)\right). \end{aligned} \tag{A.52}$$

For the LMC complexity, we have the following relationships between $C_{\mathrm{LMC}}[\sigma_{\mathrm{N}}]$ and $C_{\mathrm{LMC}}[\sigma]$, using Eq.(A.48) and Eq.(A.51)

$$\begin{aligned} C_{\mathrm{LMC}}[\sigma_{\mathrm{N}}] &= C_{\mathrm{LMC}}[\sigma]\left(\frac{N_\alpha N_\alpha}{N^2}\right)\exp\left(\sum_\kappa \tilde{S}_{\mathrm{KL}}[\sigma_\kappa,\sigma_{\mathrm{N}}]\right) \\ &\quad + 2H[\sigma]\left(\frac{N_\alpha N_\alpha}{N^2}\right)\exp\left(\sum_\kappa \tilde{S}_{\mathrm{KL}}[\sigma_\kappa,\sigma_{\mathrm{N}}]\right) \\ &\quad \times\left(\frac{N_\alpha N_\beta}{N^2}\int \sigma_\alpha(\mathbf{r})\sigma_\beta(\mathbf{r})\, d\mathbf{r}\right), \end{aligned} \tag{A.53}$$

between $C_{\mathrm{LMC}}[\sigma_{\mathrm{N}}]$ and $C_{\mathrm{LMC}}[\rho_{\mathrm{N}}]$, using Eqs. (A.43)

$$\begin{aligned} C_{\mathrm{LMC}}[\sigma_{\mathrm{N}}] &= D[\sigma_{\mathrm{N}}]H[\sigma_{\mathrm{N}}] = \frac{1}{N}D[\rho_{\mathrm{N}}]\exp\left(S[\rho_{\mathrm{N}}]\right)^{\frac{1}{N}} \\ &= \frac{H[\rho_N]^{\frac{1-N}{N}}}{N} C_{\mathrm{LMC}}[\rho_{\mathrm{N}}], \end{aligned} \tag{A.54}$$

between $C_{\mathrm{LMC}}[\rho_{\mathrm{N}}]$ and $C_{\mathrm{LMC}}[\rho]$, using Eq. (A.42) and Eq.(A.49)

$$\begin{aligned} C_{\mathrm{LMC}}[\rho_{\mathrm{N}}] &= \left(C_{\mathrm{LMC}}[\rho] + 2\int \rho_\alpha(\mathbf{r})\rho_\beta(\mathbf{r})\, d\mathbf{r}\right) \\ &\quad \times\exp\left(\sum_\kappa \tilde{S}_{\mathrm{KL}}[\rho_\alpha,\rho_{\mathrm{N}}]\right). \end{aligned} \tag{A.55}$$

The realtions for the FS complexity can be obtained in a similar way.

It should be noted that for the spin-compensated system, where $\sigma_\alpha(\mathbf{r}) = \sigma_\beta(\mathbf{r}) = \sigma_{\mathrm{N}}(\mathbf{r})$, and $\rho_\alpha(\mathbf{r}) = \rho_\beta(\mathbf{r}) = \rho_{\mathrm{N}}(\mathbf{r})/2$ all these relations are simpler due to the fact that

$$\tilde{S}_{\mathrm{KL}}[\sigma_N,\sigma_{\mathrm{N}}] = 0 \tag{A.56}$$

and

$$\begin{aligned} \tilde{S}_{\mathrm{KL}}[\rho_\kappa,\rho_{\mathrm{N}}] &= \tilde{S}_{\mathrm{KL}}[\rho_{\mathrm{N}}/2,\rho_{\mathrm{N}}] \\ &= \frac{1}{2}\int \rho_{\mathrm{N}}(\mathbf{r})\ln\frac{\rho_{\mathrm{N}}}{2\rho_{\mathrm{N}}} d\mathbf{r} = -\frac{N\ln 2}{2} . \end{aligned} \tag{A.57}$$

A compilation of all relations presented in this Appendix for spin-compensated systems is displayed in Table I.

Notes and references

Institute of Physical Chemistry, Polish Academy of Sciences, Kasprzaka 44/52, PL-01-224 Warsaw, Poland, E-mail: rbalawender@ichf.edu.pl.

† Electronic Supplementary Information (ESI) available: [All data used in this work and the interactive file with plots are accessible from ESI]. See

DOI: 10.1039/b000000x/

References

1. P. W. Ayers, J. S. M. Anderson and L. J. Bartolotti, *Int. J. Quantum Chem.*, 2005, **101**, 520-534.
2. R. G. Parr, *Annu. Rev. Phys. Chem.*, 1983, **34**, 631-656.
3. P. Geerlings, F. De Proft and W. Langenaeker, *Chem. Rev.*, 2003, **103**, 1793-1873.
4. H. Chermette, *J. Comput. Chem.*, 1999, **20**, 129-154.
5. J. L. Gazquez, *J. Mex. Chem. Soc.*, 2008, **52**, 3-10.
6. P. Geerlings and F. De Proft, *Phys. Chem. Chem. Phys.*, 2008, **10**, 3028-3042.
7. A. Cedillo, R. Contreras, M. Galvan, A. Aizman, J. Andres and V. S. Safont, *J. Phys. Chem. A*, 2007, **111**, 2442-2447.
8. S. Liu, T. Li and P. W. Ayers, *J. Chem. Phys.*, 2009, **131**, 114106.
9. P. W. Ayers, S. B. Liu and T. L. Li, *Chem. Phys. Lett.*, 2009, **480**, 318-321.
10. O. A. von Lilienfeld, R. D. Lins and U. Rothlisberger, *Phys. Rev. Lett.*, 2005, **95**, 153002.
11. O. A. von Lilienfeld and M. E. Tuckerman, *J. Chem. Phys.*, 2006, **125**.
12. O. A. von Lilienfeld and M. E. Tuckerman, *J. Chem. Theory Comput.*, 2007, **3**, 1083-1090.
13. D. Sheppard, G. Henkelman and O. A. von Lilienfeld, *J. Chem. Phys.*, 2010, **133**, 084104.
14. O. A. von Lilienfeld, *Int. J. Quantum Chem.*, 2013, **113**, 1676-1689.
15. M. Lesiuk, R. Balawender and J. Zachara, *J. Chem. Phys.*, 2012, **136**, 034104.
16. R. Balawender, M. A. Welearegay, M. Lesiuk, F. De Proft and P. Geerlings, *J. Chem. Theory Comput.*, 2013.
17. S. B. Sears, R. G. Parr and U. Dinur, *Isr. J. Chem.*, 1980, **19**, 165-173.
18. R. F. Nalewajski, *Int. J. Quantum Chem.*, 2008, **108**, 2230-2252.
19. C. P. Panos, K. C. Chatzisavvas, C. C. Moustakidis, N. Nikolaidis, S. E. Massen and K. D. Sen, in *Statistical Complexity*, ed. K. D. Sen, Springer Netherlands, 2011, pp. 49-64.
20. G. Maroulis, M. Sana and G. Leroy, *Int. J. Quantum Chem.*, 1981, **19**, 43-60.
21. A. M. Simas, A. J. Thakkar and V. H. Smith, *Int. J. Quantum Chem.*, 1983, **24**, 527-550.
22. S. R. Gadre, S. B. Sears, S. J. Chakravorty and R. D. Bendale, *Phys. Rev. A*, 1985, **32**, 2602-2606.
23. M. H. Ho, R. P. Sagar, V. H. Smith and R. O. Esquivel, *J. Phys. B-At. Mol. Opt. Phys.*, 1994, **27**, 5149-5157.
24. M. H. Ho, R. P. Sagar, J. M. Perezjorda, V. H. Smith and R. O. Esquivel, *Chem. Phys. Lett.*, 1994, **219**, 15-20.
25. M. Ho, V. H. Smith, D. F. Weaver, C. Gatti, R. P. Sagar and R. O. Esquivel, *J. Chem. Phys.*, 1998, **108**, 5469-5475.
26. P. Geerlings and A. Borgoo, *Phys. Chem. Chem. Phys.*, 2011, **13**, 911-922.
27. R. F. Nalewajski, *J. Math. Chem.*, 2009, **45**, 607-626.
28. L. Tarko, *J. Math. Chem.*, 2011, **49**, 2330-2344.
29. R. F. Nalewajski, *Mol. Phys.*, 2006, **104**, 365 - 375.
30. R. F. Nalewajski, *Mol. Phys.*, 2006, **104**, 3339 - 3370.
31. R. F. Nalewajski, *J. Math. Chem.*, 2008, **43**, 780-830.
32. R. F. Nalewajski, *Int. J. Quantum Chem.*, 2009, **109**, 2495-2506.
33. R. F. Nalewajski, *J. Math. Chem.*, 2009, **45**, 709-724.
34. R. F. Nalewajski, *J. Math. Chem.*, 2011, **49**, 546-561.
35. M. Molina-Espiritu, R. O. Esquivel, J. C. Angulo and J. S. Dehesa, *Entropy*, 2013, **15**, 4084-4104.
36. M. Molina-Espiritu, R. O. Esquivel, J. C. Angulo, J. Antolin and J. S. Dehesa, *J. Math. Chem.*, 2012, **50**, 1882-1900.
37. M. Molina-Espiritu, R. O. Esquivel, J. C. Angulo, J. Antolin, C. Iuga and J. S. Dehesa, *Int. J. Quantum Chem.*, 2013, **113**, 2589-2599.
38. A. Borgoo, P. Jaque, A. Toro-Labbe, C. Van Alsenoy and P. Geerlings, *Phys. Chem. Chem. Phys.*, 2009, **11**, 476-482.
39. J. C. Angulo, J. Antolín and R. O. Esquivel, in *Statistical Complexity*, ed. K. D. Sen, Springer Netherlands, 2011, pp. 167-213.
40. R. O. Esquivel, J. C. Angulo, J. Antolin, J. S. Dehesa, S. Lopez-Rosa and N. Flores-Gallegos, *Phys. Chem. Chem. Phys.*, 2010, **12**, 7108-7116.
41. O. Onicescu, *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences Serie A*, 1966, **263**, 841.
42. J. S. Dehesa, S. López-Rosa and D. Manzano, in *Statistical Complexity*, ed. K. D. Sen, Springer Netherlands, 2011, pp. 129-166.
43. I. Bialynickibirula and J. Mycielski, *Communications in Mathematical Physics*, 1975, **44**, 129-132.
44. W. Kutzelnigg, *Angewandte Chemie International Edition in English*, 1973, **12**, 546-562.
45. P. K. Acharya, L. J. Bartolotti, S. B. Sears and R. G. Parr, *Proceedings of the National Academy of Sciences of the United States of America-Physical Sciences*, 1980, **77**, 6978-6982.
46. S. B. Liu, *J. Chem. Phys.*, 2007, **126**, 244103.
47. P. K. Bhatia, *Information Sciences*, 1997, **97**, 233-240.
48. S. M. Taheri and R. Azizi, *Information Sciences*, 2007, **177**, 3871-3881.
49. M. Alipour and A. Mohajeri, *Mol. Phys.*, 2012, **110**, 403-405.
50. K. C. Chatzisavvas and C. P. Panos, *International Journal of Modern Physics E-Nuclear Physics*, 2005, **14**, 653-661.
51. R. LopezRuiz, H. L. Mancini and X. Calbet, *Phys. Lett. A*, 1995, **209**, 321-326.
52. J. C. Angulo, J. Antolin and K. D. Sen, *Phys. Lett. A*, 2008, **372**, 670-674.
53. K. D. Sen, J. Antolin and J. C. Angulo, *Phys. Rev. A*, 2007, **76**.
54. A. Dembo, T. M. Cover and J. A. Thomas, *Ieee Transactions on Information Theory*, 1991, **37**, 1501-1518.
55. N. H. March, *Phys. Lett. A*, 1986, **113**, 476-478.
56. A. Holas and N. H. March, *Phys. Rev. A*, 1991, **44**, 5521-5536.
57. A. Savin, A. D. Becke, J. Flad, R. Nesper, H. Preuss and H. G. Vonschnering, *Angew. Chem.-Int. Edit.*, 1991, **30**, 409-412.
58. R. G. Parr and W. Yang, *Density Functional Theory of Atoms and Molecules*, Oxford University Press, New York, 1989.
59. E. Engel and R. M. Dreizler, *Density Functional Theory. An Advanced Course*, Springer-Verlag Berlin, 2011.
60. H. Eschrig, *The Fundamentals of Density Functional Theory*, EAG.LE., Leipzig, 2003.

61. R. G. Parr, L. v. Szentpaly and S. Liu, *J. Am. Chem. Soc.*, 1999, **121**, 1922-1924.
62. 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 and J. A. Montgomery, *J. Comput. Chem.*, 1993, **14**, 1347-1363.
63. A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098-3100.
64. C. T. Lee, W. T. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785-789.
65. T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007-1023.
66. D. E. Woon and T. H. Dunning, *J. Chem. Phys.*, 1995, **103**, 4572-4585.
67. R. A. Kendall, T. H. Dunning and R. J. Harrison, *Vib. Spectrosc.*, 1992, **96**, 6796-6806.
68. R. F. W. Bader, P. L. A. Popelier and T. A. Keith, *Angew. Chem.-Int. Edit.*, 1994, **33**, 620-631.
69. R. F. W. Bader and D. Bayles, *J. Phys. Chem. A*, 2000, **104**, 5579-5589.
70. J. S. Dehesa, S. Lopez-Rosa and D. Manzano, *European Physical Journal D*, 2009, **55**, 539-548.
71. L. Lorenzo and R. A. Mosquera, *Chem. Phys. Lett.*, 2002, **356**, 305-312.
72. F. Cortes-Guzman and R. F. W. Bader, *Chem. Phys. Lett.*, 2003, **379**, 183-192.
73. F. Cortes-Guzman and R. F. W. Bader, *J. Phys. Org. Chem.*, 2004, **17**, 95-99.
74. S. B. Liu, *J. Chem. Phys.*, 2007, **126**, 191107.
75. C. Y. Rong, T. Lu and S. B. Liu, *J. Chem. Phys.*, 2014, **140**.
76. in *preparation*.
77. A. Nagy, *Chem. Phys. Lett.*, 2007, **449**, 212-215.
78. R. F. Nalewajski, *Chem. Phys. Lett.*, 2003, **367**, 414-422.
79. R. F. Nalewajski and E. Broniatowska, *J. Phys. Chem. A*, 2003, **107**, 6270-6280.
80. S. Janssens, A. Borgoo, C. Van Alsenoy and P. Geerlings, *J. Phys. Chem. A*, 2008, **112**, 10560-10569.