

# Hirshfeld partitioning from non-extensive entropies

Farnaz Heidar-Zadeh<sup>1,2,3</sup>  · Ivan Vinogradov<sup>1</sup> · Paul W. Ayers<sup>1</sup>

Received: 23 January 2017 / Accepted: 28 February 2017 / Published online: 30 March 2017  
© Springer-Verlag Berlin Heidelberg 2017

**Abstract** We show that the statistical divergence measures associated with non-extensive thermodynamic entropy functions—specifically the Tsallis, Rényi, Sharma–Mittal, supraextensive, and  $H$ -divergences—are associated with the Hirshfeld atoms-in-molecules partitioning. This extends the treatment of Nalewajski and Parr (J Phys Chem A 109:3957–3959, 2005), (for the extensive Shannon entropy) to non-extensive entropy measures. It also extends the work of Heidar-Zadeh and Ayers (J Chem Phys 142(4):044107, 2015), (for divergence measures that are local density functionals) to non-local functionals. These results dramatically extend the mathematical framework that one can use for similarity-based atoms-in-molecules partitioning.

**Keywords** Atoms in molecules · Hirshfeld partitioning · Stockholder population analysis · Atomic charges · Information theory · Non-extensive entropy · Tsallis entropy · Rényi entropy · Sharma–Mittal entropy · Supraextensive entropy ·  $H$ -divergence

## 1 Background

In the language of chemistry, molecules are built from atoms and functional groups [1–5]. Although the atoms and functional groups are deformed (or “promoted”) when they are combined, they nonetheless maintain their quiddity. This is why, for example, the periodic table, along with tables of atomic properties (like the electronegativity, hardness, and polarizability) [6–8], is essential to practicing chemists. This motivated the strategy first proposed by Nalewajski, Parr, and Ayers: define the electron density of an atom in a molecule to maximize its resemblance to the electron densities of the isolated reference atoms and atomic ions enshrined in the periodic table [9–12]. The electron density is chosen as the fundamental descriptor of atoms because of the Hohenberg–Kohn theorem [13, 14] and inspired by the pioneering work of Richard Bader [3–5, 15, 16]. The measure of “resemblance” between the atom-in-molecule’s density,  $\rho_A(\mathbf{r})$ , and the isolated reference pro-atom’s density,  $\rho_A^0(\mathbf{r})$ , was originally taken to be the Kullback–Leibler directed divergence [11],

$$I[\rho_A | \rho_A^0] = \int \rho_A(\mathbf{r}) \ln \left( \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \right) d\mathbf{r}. \quad (1)$$

The Kullback–Leibler divergence is just the Shannon information/entropy gain relative to a reference distribution and has been used in many chemical and biochemical contexts beyond the atomic partitioning problem [17–73]. If one minimizes the total divergence of all atoms, subject to the constraint that the sum of all the atom-in-molecule densities is equal to the total molecular density,

✉ Farnaz Heidar-Zadeh  
farnazhz@gmail.com

✉ Paul W. Ayers  
ayers@mcmaster.ca

<sup>1</sup> Department of Chemistry and Chemical Biology, McMaster University, Hamilton, ON, Canada

<sup>2</sup> Department of Inorganic and Physical Chemistry, Ghent University, Krijgslaan 281 (S3), 9000 Ghent, Belgium

<sup>3</sup> Center for Molecular Modeling, Ghent University, Technologiepark 903, 9052 Zwijnaarde, Belgium

$$\min_{\{\rho_A(\mathbf{r})\}_{\sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) = \rho_{\text{mol}}(\mathbf{r})}} \sum_{A=1}^{N_{\text{atoms}}} I[\rho_A | \rho_A^0], \quad (2)$$

then one obtains the “stockholder” partitioning that Hirshfeld proposed on heuristic grounds in 1977 [74], and upon which many state-of-the-art atomic population methods are based [75–89].

It was rapidly recognized that the Hirshfeld partitioning can also be obtained from more general directed and undirected divergences [12] including the non-extensive Tsallis entropy [90], generalized Hellinger-Bhattacharyya distances [91], and general  $f$ -divergences [92]. Reference [92] proves a strong result: any divergence measure that can be written as a local density functional,

$$I[\rho_A | \rho_A^0] = \int h(\rho_A(\mathbf{r}), \rho_A^0(\mathbf{r})) \mathbf{d}\mathbf{r} \quad (3)$$

and which furthermore satisfies

$$I[\rho_A | \rho_A^0] > I[\rho_A^0 | \rho_A^0] = 0 \quad (4)$$

whenever  $\rho_A(\mathbf{r}) \neq \rho_A^0(\mathbf{r})$  for the electron densities with the same number of electrons,

$$\int \rho_A(\mathbf{r}) \mathbf{d}\mathbf{r} = N_A = N_A^0 = \int \rho_A^0(\mathbf{r}) \mathbf{d}\mathbf{r}, \quad (5)$$

and when minimized according to Eq. (2) gives the Hirshfeld partitioning is an  $f$ -divergence.

The goal of this paper is to consider divergence measures that cannot be written as local functionals of the electron density. Recall that all local functionals of the electron density can be written as

$$F_{\text{local}}[\rho] = \int f(\rho(\mathbf{r})) \mathbf{d}\mathbf{r} \quad (6)$$

where  $f(x) : \mathbb{R}^+ \rightarrow \mathbb{R}$  is an ordinary function. Equivalently, to evaluate the functional derivative of a local functional at a point, one needs to only know the electron density at that point,

$$\frac{\delta F_{\text{local}}[\rho]}{\delta \rho(\mathbf{r})} = \left. \frac{df(x)}{dx} \right|_{x=\rho(\mathbf{r})} \quad (7)$$

In this paper, we will consider non-local functionals which are functions of local functionals,

$$G_{\text{non-local}}[\rho] = g\left(F_{\text{local}}^{(1)}[\rho], F_{\text{local}}^{(2)}[\rho], \dots\right) \quad (8)$$

Of particular interest are divergence measures that are based on non-extensive functionals for the thermodynamic entropy. (Entropy functionals which are non-local are inherently non-extensive.) These are not *obviously*

$f$ -divergences, and for this reason, it is interesting to assess whether or not the Hirshfeld partitioning is recovered. Specifically, we consider the directed divergence measures associated with the Tsallis divergence [90, 93, 94],

$$I_T^\alpha[\{\rho_A\} | \{\rho_A^0\}] = \frac{1}{\alpha - 1} \left[ \int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) \left( \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \right)^{\alpha-1} \mathbf{d}\mathbf{r} - \int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) \mathbf{d}\mathbf{r} \right] \quad (9)$$

the Rényi divergence [94–96],

$$I_R^\alpha[\{\rho_A\} | \{\rho_A^0\}] = \frac{1}{\alpha - 1} \ln \left( \frac{\int \sum_{A=1}^{N_{\text{atoms}}} \left( \rho_A(\mathbf{r}) \left( \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \right)^{\alpha-1} \right) \mathbf{d}\mathbf{r}}{\int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) \mathbf{d}\mathbf{r}} \right) \quad (10)$$

the Sharma–Mittal divergence [97–101],

$$I_{\text{SM}}^{\alpha, \beta}[\{\rho_A\} | \{\rho_A^0\}] = \frac{1}{\beta - 1} \left[ \left( \int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) \left( \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \right)^{\alpha-1} \mathbf{d}\mathbf{r} \right)^{\frac{\beta-1}{\alpha-1}} - \left( \int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) \mathbf{d}\mathbf{r} \right)^{\frac{\beta-1}{\alpha-1}} \right] \quad (11)$$

a recently proposed supraextensive divergence [102],

$$I_{\text{SE}}^{\alpha, \beta}[\{\rho_A\} | \{\rho_A^0\}] = \frac{1}{\alpha - 1} \left[ \left( \int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) \mathbf{d}\mathbf{r} + \frac{\beta-1}{\alpha-1} \ln \left( \frac{\int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) \left( \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \right)^{\alpha-1} \mathbf{d}\mathbf{r}}{\int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) \mathbf{d}\mathbf{r}} \right) \right)^{\frac{\alpha-1}{\beta-1}} - \left( \int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) \mathbf{d}\mathbf{r} \right)^{\frac{\alpha-1}{\beta-1}} \right] \quad (12)$$

and the very general family of  $H$ -divergences [103],

$$H_{h, \varphi_1, \varphi_2}[\{\rho_A\} | \{\rho_A^0\}] = h \left( \frac{\int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) \varphi_1 \left( \frac{\rho_A^0(\mathbf{r})}{\rho_A(\mathbf{r})} \right) \mathbf{d}\mathbf{r}}{\int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) \varphi_2 \left( \frac{\rho_A^0(\mathbf{r})}{\rho_A(\mathbf{r})} \right) \mathbf{d}\mathbf{r}} \right) \quad (13)$$

These  $H$ -divergences are not a valid divergence measure for every choice for the functions  $\varphi_1(x)$ ,  $\varphi_2(x)$ , and  $h(x)$ . It suffices, however, for  $\varphi_1(x)$  to be convex with  $\varphi_1(1) = 0$  (as for an  $f$ -divergence),  $\varphi_2(x) > 0$ , and  $h(x)$  to be monotonic ( $h'(x) > 0$ ) with  $h(0) = 0$ .

There are further extensions (e.g., corresponding to position-dependent values,  $\alpha(\mathbf{r})$ , for the parameter in Tsallis divergence) [103, 104], but we will choose not to explore those generalizations here. We also omit consideration of divergence measures that are invariant to coordinate rotations (e.g., the total Bregman divergence) [105–107]. Finally, we note that divergence measures in Eqs. (9)–(13) are slightly different from the usual form of these divergence measures. This revision is needed because atomic electron densities are normalized to the number of electrons, while the traditional divergence measures only apply to probability distribution functions that are normalized to one.

The Tsallis divergence is known to be an  $f$ -divergence and, in particular, is closely related to the special type of  $f$ -divergences called the  $\alpha$ -divergences [92, 94, 108–111],

$$\begin{aligned} I_f^\alpha[\{\rho_A\}|\{\rho_A^0\}] &= \int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) \left( \left( \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \right)^{\alpha-1} - 1 \right) d\mathbf{r} \\ &= \int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) \left( \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \right)^{\alpha-1} d\mathbf{r} - N_{\text{mol}} \end{aligned} \quad (14)$$

where

$$N_{\text{mol}} = \int \rho_{\text{mol}}(\mathbf{r}) d\mathbf{r} = \int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) d\mathbf{r} \quad (15)$$

is the number of electrons in the molecule. For convenience, we have chosen a different normalization of the  $\alpha$ -divergence from the usual form. While we regard Eq. (14) as merely a notational convenience, we note that in the absence of prefactors,  $I_f^\alpha$  is not a valid divergence measure for  $0 \leq \alpha \leq 1$ , because it is not convex.

Specifically, the Tsallis divergence is proportional to the  $\alpha$ -divergence [94]

$$I_T^\alpha = \frac{I_f^\alpha}{\alpha - 1}, \quad (16)$$

Similarly, the Rényi divergence can be written as

$$I_R^\alpha = \frac{1}{\alpha - 1} \ln \left( 1 + \frac{I_f^\alpha}{N_{\text{mol}}} \right). \quad (17)$$

The  $\alpha$ -divergence is also closely related to the Sharma–Mittal divergence,

$$I_{\text{SM}}^{\alpha,\beta} = \frac{N_{\text{mol}}^{\frac{\beta-1}{\alpha-1}}}{\beta - 1} \left[ \left( 1 + \frac{I_f^\alpha}{N_{\text{mol}}} \right)^{\frac{\beta-1}{\alpha-1}} - 1 \right], \quad (18)$$

and the supraextensive divergence,

$$I_{\text{SE}}^{\alpha,\beta} = \frac{N_{\text{mol}}^{\frac{\alpha-1}{\beta-1}}}{\alpha - 1} \left[ \left( 1 + \frac{\beta - 1}{N_{\text{mol}}(\alpha - 1)} \ln \left( 1 + \frac{I_f^\alpha}{N_{\text{mol}}} \right) \right)^{\frac{\alpha-1}{\beta-1}} - 1 \right]. \quad (19)$$

Notice that the Rényi, Sharma–Mittal, and supraextensive divergences are functions of local functionals [cf. Eq. (8)]. They are therefore non-local density functionals, not  $f$ -divergences.

The goal of this paper is to show that these more general families of divergences, which are closely related to the  $\alpha$ -divergence, give rise to Hirshfeld atoms. This extends the results of Ref. [92] and is the first time that the Hirshfeld partitioning has been obtained from non-local divergence functionals. This gives a more general approach to Hirshfeld-inspired atomic partitioning. As Verstraelen et al. [75] have noted, this also has potential applications in computational algorithms for electronic structure theory, e.g., density fitting [112–116].

## 2 Non-extensive information models and Hirshfeld atoms

Suppose one is given an information loss function that has the general form,

$$I_{\text{gen}}^\alpha[\{\rho_A\}|\{\rho_A^0\}] = g(N_{\text{mol}}, I_f^\alpha) \quad (20)$$

This form clearly encompasses and generalizes the Tsallis, Rényi, Sharma–Mittal, and supraextensive divergence measures. We then determine the atoms in molecule by the usual procedure,

$$\min_{\{\rho_A(\mathbf{r}) | \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) = \rho_{\text{mol}}(\mathbf{r})\}} I_{\text{gen}}^\alpha[\{\rho_A\}|\{\rho_A^0\}] \quad (21)$$

Introducing the constraint with a Lagrange multiplier, the Lagrangian is:

$$\Lambda[\{\rho_A\}] = g(N_{\text{mol}}, I_f^\alpha) - \int \lambda(\mathbf{r}) \left( \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) - \rho_{\text{mol}}(\mathbf{r}) \right) d\mathbf{r} \quad (22)$$

and the stationary condition for the minimum is

$$0 = \frac{\delta \Lambda}{\delta \rho_B(\mathbf{r})} = \frac{\partial g}{\partial N_{\text{mol}}} \frac{\delta N_{\text{mol}}}{\delta \rho_B(\mathbf{r})} + \frac{\partial g}{\partial I_f^\alpha} \frac{\delta I_f^\alpha}{\delta \rho_B(\mathbf{r})} - \lambda(\mathbf{r}) \quad (23)$$

where

$$\frac{\delta N_{\text{mol}}}{\delta \rho_B(\mathbf{r})} = 1 \quad (24)$$

$$\frac{\delta I_f^\alpha}{\delta \rho_B(\mathbf{r})} = (\alpha - 1) \left( \frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})} \right)^{\alpha-1} \quad (25)$$

The equation can then be written

$$\lambda(\mathbf{r}) - \frac{\partial g}{\partial N_{\text{mol}}} = (\alpha - 1) \frac{\partial g}{\partial I_f^\alpha} \left( \frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})} \right)^{\alpha-1} \quad (26)$$

As long as  $\alpha \neq 1$  and  $\partial g / \partial I_f^\alpha \neq 0$ , this identity gives the key relation from which the Hirshfeld atom is derived, namely that  $\rho_B(\mathbf{r}) / \rho_B^0(\mathbf{r})$  is the same for all atoms. (For example, it is sufficient to have a strictly monotonic  $g(N_{\text{mol}}, I_f^\alpha)$  with respect to  $I_f^\alpha > 0$ .) For the Tsallis, Rényi, and Sharma–Mittal divergences,

$$\frac{\partial g_T}{\partial I_f^\alpha} = \frac{1}{\alpha - 1} \neq 0 \quad (27)$$

$$\frac{\partial g_R}{\partial I_f^\alpha} = \frac{1}{(\alpha - 1)(N_{\text{mol}} + I_f^\alpha)} \neq 0 \quad (28)$$

$$\frac{\partial g_{\text{SM}}}{\partial I_f^\alpha} = \frac{N_{\text{mol}}^{\frac{\beta-\alpha}{\alpha-1}}}{\alpha - 1} \left[ \left( 1 + \frac{I_f^\alpha}{N_{\text{mol}}} \right)^{\frac{\beta-\alpha}{\alpha-1}} \right] \neq 0 \quad (29)$$

For the supraextensive entropy,

$$\begin{aligned} \frac{\partial g_{\text{SE}}}{\partial I_f^\alpha} &= \frac{N_{\text{mol}}^{\frac{\alpha-1}{\beta-1}}}{\beta - 1} \left[ \left( N_{\text{mol}} + \frac{\beta - 1}{N_{\text{mol}}(\alpha - 1)} \ln \left( 1 + \frac{I_f^\alpha}{N_{\text{mol}}} \right) \right)^{\frac{\alpha-\beta}{\beta-1}} \right] \\ &\times \left( \frac{\beta - 1}{(\alpha - 1)(N_{\text{mol}} + I_f^\alpha)} \right) \end{aligned} \quad (30)$$

Since this expression cannot be equal to zero, one must have  $\beta \neq 1$ . In all these expressions, we have used the fact that  $I_f^\alpha \geq 0$ , which presumes that the sum of the atomic densities and the sum of the reference pro-atomic densities have the same normalization.

For local divergence functionals, one sometimes uses the fact that the densities of the reference pro-atoms,  $\{\rho_A^0(\mathbf{r})\}_{A=1}^{N_{\text{atoms}}}$ , can be optimized to make the density of the so-called pro-molecule,

$$\rho_{\text{mol}}^0(\mathbf{r}) = \sum_{A=1}^{N_{\text{atoms}}} \rho_A^0(\mathbf{r}) \quad (31)$$

as close as possible to the density of the molecule,  $\rho_{\text{mol}}(\mathbf{r})$  [75, 92]. (This can remove the ambiguity associated with picking the reference pro-atom(s).) This can also be done for these measures. To see this, notice that the key Hirshfeld criterion,

$$\frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})} = h(\mathbf{r}), \quad (32)$$

for some function  $h(\mathbf{r})$ , can be written as

$$\sum_{B=1}^{N_{\text{atoms}}} \rho_B(\mathbf{r}) = h(\mathbf{r}) \sum_{B=1}^{N_{\text{atoms}}} \rho_B^0(\mathbf{r}) \quad (33)$$

Therefore

$$h(\mathbf{r}) = \frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})} = \frac{\rho_{\text{mol}}(\mathbf{r})}{\rho_{\text{mol}}^0(\mathbf{r})} \quad (34)$$

and Eq. (14) can be rewritten as

$$\begin{aligned} I_f^\alpha [\rho_{\text{mol}} | \rho_{\text{mol}}^0] &= \int \rho_{\text{mol}}(\mathbf{r}) \left( \frac{\rho_{\text{mol}}(\mathbf{r})}{\rho_{\text{mol}}^0(\mathbf{r})} \right)^{\alpha-1} d\mathbf{r} \\ &= I_f^\alpha [\{\rho_A\} | \{\rho_A^0\}] \end{aligned} \quad (35)$$

where in Eqs. (33) and (35) we have used the constraint that the atom-in-molecule densities add up to the total molecular density. The pro-molecule density can therefore be optimized by the two-step procedure,

$$\begin{aligned} &\underbrace{\min}_{\{\rho_A^0(\mathbf{r})\}} \underbrace{\min}_{\{\rho_A(\mathbf{r}) | \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) = \rho_{\text{mol}}(\mathbf{r})\}} I_f^\alpha [\{\rho_A\} | \{\rho_A^0\}] \\ &= \underbrace{\min}_{\{\rho_A^0(\mathbf{r})\}} I_f^\alpha \left[ \rho_{\text{mol}} \left| \sum_{A=1}^{N_{\text{atoms}}} \rho_A^0 \right. \right] \end{aligned} \quad (36)$$

Identity (35) and the strategy in Eq. (36) clearly extend to any of the generalized  $\alpha$ -divergences in this paper. While these formulas generalize the  $f$ -divergences considered in Ref. [92] somewhat, they do not contradict the results in that paper because these divergences are not local functionals of the electron density. Their generalizations are also not very consequential, since one still obtains the Hirshfeld partitioning. However, while the Tsallis and Rényi divergences give the same pro-atoms (because both objective functions are minimized when  $I_f^\alpha [\rho_{\text{mol}} | \rho_{\text{mol}}^0]$  is made as small as possible), this is not necessarily true for the Sharma–Mittal and supraextensive divergences.

### 3 Atoms in molecules from $H$ -divergences

The divergence measures we considered in the previous section are all based on non-extensive entropy formulas.

The  $H$ -divergence formula in Eq. (13) generalizes these equations and also the  $f$ -divergence. For example, the  $H$ -divergence is an  $f$ -divergence (up to a choice of normalization) if  $\varphi_1(1) = 0$ ,  $\varphi_1(x)$  is convex,  $\varphi_2(x) = 1$ , and  $h(x) = x$ .

As mentioned before, not every choice of functions in Eq. (13) is allowed. In this paper, we consider only  $H$ -divergences which satisfy the requirements:

- $h(x)$  is monotonically increasing,  $h'(x) > 0$ , and  $h(0) = 0$ .
- $\varphi_1(x)$  is convex,  $\varphi_1''(x) > 0$ , and  $\varphi_1(1) = 0$ .
- $\varphi_2(x) > 0$ .

This gives  $H_{h,\varphi_1,\varphi_2}[\{\rho_A\}|\{\rho_A^0\}] \geq H_{h,\varphi_1,\varphi_2}[\{\rho_A^0\}|\{\rho_A^0\}] = 0$  for densities with the same normalization, which is one of the essential properties of a divergence measure. The analogous  $H$ -divergence derivation of the Hirshfeld atoms-in-molecules partitioning is found by minimizing

$$\min_{\{\rho_A(\mathbf{r})\}_{\sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) = \rho_{\text{mol}}(\mathbf{r})}} H_{h,\varphi_1,\varphi_2}[\{\rho_A\}|\{\rho_A^0\}] \quad (37)$$

with the solution

$$\lambda(\mathbf{r}) = \frac{1}{G_{\varphi_2}^2} \cdot h' \left( \frac{G_{\varphi_1}}{G_{\varphi_2}} \right) \cdot \left[ G_{\varphi_2} \frac{\delta G_{\varphi_1}}{\delta \rho_B(\mathbf{r})} - G_{\varphi_1} \frac{\delta G_{\varphi_2}}{\delta \rho_B(\mathbf{r})} \right] \quad (38)$$

where we have defined the convenient notation,

$$G_{\varphi_1} = \int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) \varphi_1 \left( \frac{\rho_A^0(\mathbf{r})}{\rho_A(\mathbf{r})} \right) d\mathbf{r} \quad (39)$$

$$G_{\varphi_2} = \int \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) \varphi_2 \left( \frac{\rho_A^0(\mathbf{r})}{\rho_A(\mathbf{r})} \right) d\mathbf{r},$$

Note that by requiring that  $\varphi_1(x)$  is a convex function with  $\varphi_1(1) = 0$ , we ensure that  $G_{\varphi_1}$  is an  $f$ -divergence.  $G_{\varphi_2}$  is not an  $f$ -divergence, but a type of normalization factor. Possible choices include  $\varphi_2(x) = x^\alpha$  ( $0 \leq \alpha \leq 1$ ),  $\varphi_2(x) = x/(x+1)$ ,  $\varphi_2(x) = \ln(x+1)$ ,  $\varphi_2(x) = \tanh(x)$ . All of these functions are concave for  $x \geq 0$ ,  $\varphi_2''(x) < 0$ . This is not required for  $H$ -divergence to be a valid divergence measure, but later it will turn out to be useful.

Inserting the functional derivatives,

$$\frac{\delta G_{\varphi_1}}{\delta \rho_B(\mathbf{r})} = \varphi_1 \left( \frac{\rho_B^0(\mathbf{r})}{\rho_B(\mathbf{r})} \right) - \frac{\rho_B^0(\mathbf{r})}{\rho_B(\mathbf{r})} \varphi_1' \left( \frac{\rho_B^0(\mathbf{r})}{\rho_B(\mathbf{r})} \right)$$

$$\frac{\delta G_{\varphi_2}}{\delta \rho_B(\mathbf{r})} = \varphi_2 \left( \frac{\rho_B^0(\mathbf{r})}{\rho_B(\mathbf{r})} \right) - \frac{\rho_B^0(\mathbf{r})}{\rho_B(\mathbf{r})} \varphi_2' \left( \frac{\rho_B^0(\mathbf{r})}{\rho_B(\mathbf{r})} \right) \quad (40)$$

into Eq. (38), we obtain the expression

$$\lambda(\mathbf{r}) = h' \left( \frac{G_{\varphi_1}}{G_{\varphi_2}} \right) \left( \frac{1}{G_{\varphi_2}} \cdot \left( \varphi_1 \left( \frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})} \right) - \frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})} \varphi_1' \left( \frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})} \right) \right) \right. \\ \left. - \frac{G_{\varphi_1}}{G_{\varphi_2}^2} \cdot \left( \varphi_2 \left( \frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})} \right) - \frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})} \varphi_2' \left( \frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})} \right) \right) \right) \quad (41)$$

Equation (32), which leads to the Hirshfeld partitioning, is a solution to this equation. However, it may not be the only solution. In general, Eq. (41) gives an equation relating the densities of every atom pair in the molecule,

$$g \left( \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \right) = g \left( \frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})} \right) \quad (42)$$

where

$$g(x) = \frac{1}{G_{\varphi_2}} (\varphi_1(x) - x\varphi_1'(x)) - \frac{G_{\varphi_1}}{G_{\varphi_2}^2} (\varphi_2(x) - x\varphi_2'(x)). \quad (43)$$

If  $g(x)$  is invertible for  $x \geq 0$ , then the unique solution to Eq. (42) is

$$\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} = \frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})}, \quad (44)$$

which leads to the Hirshfeld partitioning. If we assume that all the functions are at least twice differentiable, it is sufficient that  $g(x)$  be monotonic. Therefore, for  $x > 0$ ,

$$g'(x) = \frac{x}{G_{\varphi_2}^2} (G_{\varphi_2} \varphi_1''(x) - G_{\varphi_1} \varphi_2''(x)) > 0 \quad (45)$$

The conditions stipulated at the beginning of this section are almost sufficient to satisfy this equation because they ensure that  $G_{\varphi_2}$  is positive, that  $\varphi_1''(x)$  are positive, and that  $G_{\varphi_1}$  is nonnegative. If we further require  $\varphi_2''(x)$  to be non-positive, then the Hirshfeld partitioning is the unique solution to variational procedure (37). These conditions also suffice to derive the analogue of the identity in Eq. (35), namely that for the atom-in-molecule densities obtained from Eq. (37),

$$I_{h,\varphi_1,\varphi_2}[\{\rho_A\}|\{\rho_A^0\}] = I_{h,\varphi_1,\varphi_2}[\rho_{\text{mol}}|\rho_{\text{mol}}^0] \\ = h \left( \frac{\int \rho_{\text{mol}}(\mathbf{r}) \varphi_1 \left( \frac{\rho_{\text{mol}}^0(\mathbf{r})}{\rho_{\text{mol}}(\mathbf{r})} \right) d\mathbf{r}}{\int \rho_{\text{mol}}(\mathbf{r}) \varphi_2 \left( \frac{\rho_{\text{mol}}^0(\mathbf{r})}{\rho_{\text{mol}}(\mathbf{r})} \right) d\mathbf{r}} \right). \quad (46)$$

## 4 Summary

This work was initiated when our numerical investigations revealed that optimizing the pro-atoms,



$$\min_{\{\rho_A^0(\mathbf{r})\}} I \left[ \rho_{\text{mol}} \left| \sum_{A=1}^{N_{\text{atoms}}} \rho_A^0 \right. \right], \quad (47)$$

gave the same results for the Tsallis and Rényi divergences. We were also surprised that the Rényi divergence, even though it is not an  $f$ -divergence, gave back the Hirshfeld partitioning. This led us to explore what other sorts of non-local divergence measures would recover the Hirshfeld partitioning. This paper reports the results of that exploration.

Reference [92] shows that the only *local* density functionals that lead to the popular Hirshfeld partitioning are  $f$ -divergences. This paper explores divergences that are *non-local* density functionals but which also give the Hirshfeld partitioning. In particular, we observe that the Rényi, Sharma–Mittal, and supraextensive divergence measures all give the Hirshfeld partitioning. Moreover, all of these functionals are very closely linked to the  $\alpha$ -divergence. This is desirable insofar that it ensures that these measures are closely linked to a very popular and useful family of divergence measures, but it is undesirable insofar as it means that optimizing the pro-atom densities [using Eq. (47)] does not give significantly different results for these approaches.

The  $H$ -divergence in Eq. (13) is much more general. While it is difficult to find necessary conditions for the  $H$ -divergence that gives the Hirshfeld atom, it is sufficient to require the following properties for  $x > 0$ :

- $h(x)$  is monotonically increasing,  $h'(x) > 0$ . Also  $h(0) = 0$ .
- $\varphi_1(x)$  is convex,  $\varphi_1''(x) > 0$ . Also  $\varphi_1(1) = 0$ . This is the same as the requirements for an  $f$ -divergence.
- $\varphi_2(x) > 0$  and is non-convex,  $\varphi_2''(x) \leq 0$ .

Note that this family of  $H$ -divergences is closely related to the  $f$ -divergences, but extends that set in a non-trivial way. We anticipate that the family of  $H$ -divergences could be used to define new, and more effective, alternatives to variational Hirshfeld-based methods like the minimal basis iterated stockholder (MBIS) partitioning [75].

While our primary interest in divergence measures is motivated by the problem of atomic partitioning and, more generally, fitting molecular densities [cf. Eq. (35)], the mathematical tools presented here are suitable for measuring the divergence between other probability distribution functions that arise in quantum chemistry. For example, there has been a significant recent interest in approaches that use the shape function [90, 117–119], instead of the electron density, to describe chemical phenomena [70, 120–123].

**Acknowledgements** The authors thank NSERC and Compute Canada for funding. FHZ acknowledges support from Vanier-CGS fellowship and Ghent University Scholarship for a Joint Doctorate.

## References

1. Parr RG, Ayers PW, Nalewajski RF (2005) What is an atom in a molecule? *J Phys Chem A* 109:3957–3959
2. Matta CF, Bader RFW (2006) An experimentalist's reply to "What is an atom in a molecule?". *J Phys Chem A* 110:6365–6371
3. Bader RFW, Nguyendang TT (1981) Quantum-theory of atoms in molecules: Dalton revisited. *Adv Quantum Chem* 14:63–124
4. Bader RFW (1990) *Atoms in molecules: a quantum theory*. Clarendon, Oxford
5. Bader RFW (1985) *Atoms in molecules*. *Acc Chem Res* 18:9–15
6. Huheey JE, Keiter EA, Keiter RL (1993) *Inorganic chemistry: principles of structure and reactivity*. HarperCollins, New York
7. Cardenas C, Heidar-Zadeh F, Ayers PW (2016) Benchmark values of chemical potential and chemical hardness for atoms and atomic ions (including unstable anions) from the energies of isoelectronic series. *PCCP* 18(36):25721–25734. doi:10.1039/c6cp04533b
8. Parr RG, Yang W (1989) *Density-functional theory of atoms and molecules*. Oxford University Press, New York
9. Nalewajski RF, Switka E, Michalak A (2002) Information distance analysis of molecular electron densities. *Int J Quantum Chem* 87(4):198–213
10. Nalewajski RF, Parr RG (2001) Information theory thermodynamics of molecules and their Hirshfeld fragments. *J Phys Chem A* 105(31):7391–7400
11. Nalewajski RF, Parr RG (2000) Information theory, atoms in molecules, and molecular similarity. *Proc Natl Acad Sci* 97:8879–8882
12. Ayers PW (2000) Atoms in molecules, an axiomatic approach. I. Maximum transferability. *J Chem Phys* 113:10886–10898
13. Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. *Phys Rev* 136:B864–B871
14. Johnson PA, Bartolotti LJ, Ayers PW, Fievez T, Geerlings P (2012) Charge density and chemical reactivity: a unified view from conceptual DFT. In: Gatti C, Macchi P (eds) *Modern charge density analysis*. Springer, New York, pp 715–764
15. Heidarzadeh F, Shahbazian S (2011) The quantum divided basins: a new class of quantum subsystems. *Int J Quantum Chem* 111:2788–2801. doi:10.1002/qua.22629
16. Heidar Zadeh F, Shahbazian S (2011) Toward a fuzzy atom view within the context of the quantum theory of atoms in molecules: quasi-atoms. *Theor Chem Acc* 128(2):175–181. doi:10.1007/s00214-010-0811-x
17. Kullback S (1997) *Information theory and statistics*. Dover, Mineola
18. Kullback S, Leibler RA (1951) On information and sufficiency. *Ann Math Stat* 22:79–86
19. Gadre SR, Sears SB, Chakravorty SJ, Bendale RD (1985) Some novel characteristics of atomic information entropies. *Phys Rev A* 32:2602–2606
20. Sears SB, Gadre SR (1981) An information theoretic synthesis and analysis of Compton profiles. *J Chem Phys* 75(9):4626–4635
21. Acharya PK, Bartolotti LJ, Sears SB, Parr RG (1980) An atomic kinetic energy functional with full Weizsacker correction. *Proc Natl Acad Sci* 77:6978–6982
22. Sears SB, Parr RG, Dinur U (1980) On the quantum-mechanical kinetic energy as a measure of the information in a distribution. *Isr J Chem* 19(1–4):165–173
23. Sears SB (1980) University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

24. Heidar-Zadeh F, Fuentealba P, Cardenas C, Ayers PW (2014) An information-theoretic resolution of the ambiguity in the local hardness. *PCCP* 16(13):6019–6026. doi:[10.1039/c3cp52906a](https://doi.org/10.1039/c3cp52906a)
25. Kaplan H, Levine RD, Manz J (1976) Microscopic reversibility and probability matrices for molecular collisions—information theoretic synthesis. *Mol Phys* 31(6):1765–1782. doi:[10.1080/00268977600101401](https://doi.org/10.1080/00268977600101401)
26. Procaccia I, Levine RD (1976) Cross-sections for rotational energy transfer—an information-theoretic synthesis. *J Chem Phys* 64(2):808–817. doi:[10.1063/1.432227](https://doi.org/10.1063/1.432227)
27. Jensen CC, Steinfeld JJ, Levine RD (1978) Information theoretic analysis of multi-photon excitation and collisional deactivation in polyatomic molecules. *J Chem Phys* 69(4):1432–1439. doi:[10.1063/1.436757](https://doi.org/10.1063/1.436757)
28. Levine RD (1978) Information-theory approach to molecular reaction dynamics. *Annu Rev Phys Chem* 29:59–92. doi:[10.1146/annurev.pc.29.100178.000423](https://doi.org/10.1146/annurev.pc.29.100178.000423)
29. Alhassid Y, Levine RD, Karp JS, Steadman SG (1979) Information-theoretic analysis of energy disposal in heavy-ion transfer-reactions. *Phys Rev C* 20(5):1789–1813. doi:[10.1103/PhysRevC.20.1789](https://doi.org/10.1103/PhysRevC.20.1789)
30. Levine RD (1980) Information theoretical approach to inversion problems. *J Phys A Math Gen* 13(1):91–108. doi:[10.1088/0305-4470/13/1/011](https://doi.org/10.1088/0305-4470/13/1/011)
31. Zamir E, Haas Y, Levine RD (1980) Laser enhanced addition-reactions between hydrogen halides and unsaturated hydrocarbons—an information theoretic approach. *J Chem Phys* 73(6):2680–2687. doi:[10.1063/1.440481](https://doi.org/10.1063/1.440481)
32. Alhassid Y, Levine RD (1980) Experimental and inherent uncertainties in the information theoretic approach. *Chem Phys Lett* 73(1):16–20. doi:[10.1016/0009-2614\(80\)85192-x](https://doi.org/10.1016/0009-2614(80)85192-x)
33. Engel YM, Levine RD, Thoman JW, Steinfeld JJ, McKay R (1988) Information theoretic analysis of quantum fluctuations in fluorescence lifetimes. *J Phys Chem* 92(19):5497–5500. doi:[10.1021/j100330a032](https://doi.org/10.1021/j100330a032)
34. Levine RD (1990) The chemical shape of molecules—an introduction to dynamic stereochemistry. *J Phys Chem* 94(26):8872–8880. doi:[10.1021/j100389a006](https://doi.org/10.1021/j100389a006)
35. Levine RD (1998) Monte Carlo, maximum entropy and importance sampling. *Chem Phys* 228(1–3):255–264. doi:[10.1016/s0301-0104\(97\)00334-0](https://doi.org/10.1016/s0301-0104(97)00334-0)
36. Gross A, Levine RD (2006) A mechanical representation of entropy for a large finite system. *J Chem Phys.* doi:[10.1063/1.2357149](https://doi.org/10.1063/1.2357149)
37. Remacle F, Levine RD (2009) The elimination of redundant constraints in surprisal analysis of unimolecular dissociation and other endothermic processes. *J Phys Chem A* 113(16):4658–4664. doi:[10.1021/jp811463h](https://doi.org/10.1021/jp811463h)
38. Graeber TG, Heath JR, Skaggs BJ, Phelps ME, Remacle F, Levine RD (2010) Maximal entropy inference of oncogenicity from phosphorylation signaling. *Proc Natl Acad Sci* 107(13):6112–6117. doi:[10.1073/pnas.1001149107](https://doi.org/10.1073/pnas.1001149107)
39. Remacle F, Kravchenko-Balasha N, Levitzki A, Levine RD (2010) Information-theoretic analysis of phenotype changes in early stages of carcinogenesis. *Proc Natl Acad Sci* 107(22):10324–10329. doi:[10.1073/pnas.1005283107](https://doi.org/10.1073/pnas.1005283107)
40. Kravchenko-Balasha N, Remacle F, Gross A, Rotter V, Levitzki A, Levine RD (2011) Convergence of logic of cellular regulation in different premalignant cells by an information theoretic approach. *BMC Syst Biol.* doi:[10.1186/1752-0509-5-42](https://doi.org/10.1186/1752-0509-5-42)
41. Shin YS, Remacle F, Fan R, Hwang K, Wei W, Ahmad H, Levine RD, Heath JR (2011) Protein signaling networks from single cell fluctuations and information theory profiling. *Biophys J* 100(10):2378–2386. doi:[10.1016/j.bpj.2011.04.025](https://doi.org/10.1016/j.bpj.2011.04.025)
42. Remacle F, Graeber TG, Levine RD (2011) Whose entropy: a maximal entropy analysis of phosphorylation signaling. *J Stat Phys* 144(2):429–442. doi:[10.1007/s10955-011-0215-x](https://doi.org/10.1007/s10955-011-0215-x)
43. Gross A, Levine RD (2013) Surprisal analysis of transcripts expression levels in the presence of noise: a reliable determination of the onset of a tumor phenotype. *PLoS ONE.* doi:[10.1371/journal.pone.0061554](https://doi.org/10.1371/journal.pone.0061554)
44. Zadran S, Remacle F, Levine RD (2013) miRNA and mRNA cancer signatures determined by analysis of expression levels in large cohorts of patients. *Proc Natl Acad Sci* 110(47):19160–19165. doi:[10.1073/pnas.1316991110](https://doi.org/10.1073/pnas.1316991110)
45. Kravchenko-Balasha N, Wang J, Remacle F, Levine RD, Heath JR (2014) Glioblastoma cellular architectures are predicted through the characterization of two-cell interactions. *Proc Natl Acad Sci* 111(17):6521–6526. doi:[10.1073/pnas.1404462111](https://doi.org/10.1073/pnas.1404462111)
46. Zadran S, Arumugam R, Herschman H, Phelps ME, Levine RD (2014) Surprisal analysis characterizes the free energy time course of cancer cells undergoing epithelial-to-mesenchymal transition. *Proc Natl Acad Sci* 111(36):13235–13240. doi:[10.1073/pnas.1414714111](https://doi.org/10.1073/pnas.1414714111)
47. Kravchenko-Balasha N, Simon S, Levine RD, Remacle F, Exman I (2014) Computational surprisal analysis speeds-up genomic characterization of cancer processes. *PLoS ONE.* doi:[10.1371/journal.pone.0108549](https://doi.org/10.1371/journal.pone.0108549)
48. Kravchenko-Balasha N, Shin YS, Sutherland A, Levine RD, Heath JR (2016) Intercellular signaling through secreted proteins induces free-energy gradient-directed cell movement. *Proc Natl Acad Sci* 113(20):5520–5525. doi:[10.1073/pnas.1602171113](https://doi.org/10.1073/pnas.1602171113)
49. Levine RD, Bernstein RB (1974) Energy consumption and energy disposal in elementary chemical reactions: The information theoretic approach. *Acc Chem Res* 7:393
50. Levine RD (2005) *Molecular reaction dynamics*. Cambridge University Press, Cambridge
51. Liu SB (2016) Information-theoretic approach in density functional reactivity theory. *Acta Phys Chim Sin* 32(1):98–118. doi:[10.3866/pkumhxb201510302](https://doi.org/10.3866/pkumhxb201510302)
52. Liu SB, Rong CY, Lu T (2014) Information conservation principle determines electrophilicity, nucleophilicity, and regioselectivity. *J Phys Chem A* 118(20):3698–3704. doi:[10.1021/jp5032702](https://doi.org/10.1021/jp5032702)
53. Godo B, Nagy A (2016) Fisher information and Rényi dimensions: a thermodynamical formalism. *Chaos.* doi:[10.1063/1.4959908](https://doi.org/10.1063/1.4959908)
54. Nagy A (2013) Shannon entropy density as a descriptor of Coulomb systems. *Chem Phys Lett* 556:355–358. doi:[10.1016/j.cplett.2012.11.065](https://doi.org/10.1016/j.cplett.2012.11.065)
55. Nagy A (2015) Fisher and Shannon information in orbital-free density functional theory. *Int J Quantum Chem* 115(19):1392–1395. doi:[10.1002/qua.24812](https://doi.org/10.1002/qua.24812)
56. Nagy A, Romera E (2009) Relative Rényi entropy for atoms. *Int J Quantum Chem* 109(11):2490–2494. doi:[10.1002/qua.21962](https://doi.org/10.1002/qua.21962)
57. Nagy A, Romera E (2009) Maximum Rényi entropy principle and the generalized Thomas–Fermi model. *Phys Lett A* 373(8–9):844–846. doi:[10.1016/j.physleta.2009.01.004](https://doi.org/10.1016/j.physleta.2009.01.004)
58. Nagy A, Romera E (2014) Fisher and Shannon information from one-matrix. Link to the kinetic energy. *Chem Phys Lett* 597:139–142. doi:[10.1016/j.cplett.2014.02.032](https://doi.org/10.1016/j.cplett.2014.02.032)
59. Nagy A, Romera E, Liu SB (2013) Local coordinate, wave vector, Fisher and Shannon information in momentum representation. *Phys Lett A* 377(3–4):286–290. doi:[10.1016/j.physleta.2012.11.018](https://doi.org/10.1016/j.physleta.2012.11.018)
60. Ayers PW, Parr RG, Nagy A (2002) Local kinetic energy and local temperature in the density-functional theory of electronic structure. *Int J Quantum Chem* 90(1):309–326. doi:[10.1002/qua.989](https://doi.org/10.1002/qua.989)

61. Chattaraj PK, Chamorro E, Fuentealba P (1999) Chemical bonding and reactivity: a local thermodynamic viewpoint. *Chem Phys Lett* 314(1–2):114–121. doi:10.1016/S0009-2614(99)01114-8
62. Nagy A, Parr RG (1994) Density-functional theory as thermodynamics. *Proc Indian Acad Sci Chem Sci* 106(2):217–227
63. Nagy A, Parr RG (1996) Information entropy as a measure of the quality of an approximate electronic wave function. *Int J Quantum Chem* 58(4):323–327. doi:10.1002/(sici)1097-461x(1996)58:4<323:aid-qual>3.3.co;2-t
64. Nagy A, Parr RG (2000) Remarks on density functional theory as a thermodynamics. *J Mol Struct THEOCHEM* 501:101–106
65. Nagy A, Sen KD (2006) Atomic Fisher information versus atomic number. *Phys Lett A* 360(2):291–293. doi:10.1016/j.physleta.2006.08.025
66. Romera E, Nagy A (2008) Rényi information of atoms. *Phys Lett A* 372(29):4918–4922. doi:10.1016/j.physleta.2008.05.029
67. Liu SB (2015) Quantifying reactivity for electrophilic aromatic substitution reactions with Hirshfeld charge. *J Phys Chem A* 119(12):3107–3111. doi:10.1021/acs.jpca.5b00443
68. Wu WJ, Wu ZM, Rong CY, Lu T, Huang Y, Liu SB (2015) Computational study of chemical reactivity using information-theoretic quantities from density functional reactivity theory for electrophilic aromatic substitution reactions. *J Phys Chem A* 119(29):8216–8224. doi:10.1021/acs.jpca.5b04309
69. Wu ZM, Rong CY, Lu T, Ayers PW, Liu SB (2015) Density functional reactivity theory study of S(N)2 reactions from the information-theoretic perspective. *PCCP* 17(40):27052–27061. doi:10.1039/c5cp04442a
70. Rong CY, Lu T, Ayers PW, Chattaraj PK, Liu SB (2015) Scaling properties of information-theoretic quantities in density functional reactivity theory. *PCCP* 17(7):4977–4988. doi:10.1039/c4cp05609d
71. Liu SB (2014) Where does the electron go? The nature of ortho/para and meta group directing in electrophilic aromatic substitution. *J Chem Phys*. doi:10.1063/1.4901898
72. Zhou X-Y, Rong C-Y, Lu T, Liu S-B (2014) Hirshfeld charge as a quantitative measure of electrophilicity and nucleophilicity: nitrogen-containing systems. *Acta Phys Chim Sin* 30(11):2055–2062. doi:10.3866/pku.whxb201409193
73. Liu SB, Rong CY, Wu ZM, Lu T (2015) Rényi entropy, Tsallis entropy and Onicescu information energy in density functional reactivity theory. *Acta Phys Chim Sin* 31(11):2057–2063. doi:10.3866/pku.whxb201509183
74. Hirshfeld FL (1977) Bonded-atom fragments for describing molecular charge densities. *Theor Chim Acta* 44:129–138
75. Verstraelen T, Vandenbrande S, Heidar-Zadeh F, Vanduyfhuys L, Van Speybroeck V, Waroquier M, Ayers PW (2016) Minimal basis iterative stockholder: atoms in molecules for force-field development. *J Chem Theory Comp* 12(8):3894–3912. doi:10.1021/acs.jctc.6b00456
76. Bultinck P, Van Alsenoy C, Ayers PW, Carbó-Dorca R (2007) Critical analysis and extension of the Hirshfeld atoms in molecules. *J Chem Phys* 126:144111. doi:10.1063/1.2715563
77. Bultinck P, Ayers PW, Fias S, Tiels K, Van Alsenoy C (2007) Uniqueness and basis set dependence of iterative Hirshfeld charges. *Chem Phys Lett* 444:205–208
78. Verstraelen T, Ayers PW, Van Speybroeck V, Waroquier M (2013) Hirshfeld-E partitioning: AIM charges with an improved trade-off between robustness and accurate electrostatics. *J Chem Theory Comput* 9:2221–2225
79. Verstraelen T, Ayers PW, Van Speybroeck V, Waroquier M (2014) Extended Hirshfeld: atomic charges that combine accurate electrostatics with transferability. *Abstr Pap Am Chem Soc* 247:1
80. Ghillemijn D, Bultinck P, Van Neck D, Ayers PW (2011) A self-consistent Hirshfeld method for the atom in the molecule based on minimization of information loss. *J Comput Chem* 32:1561–1567. doi:10.1002/jcc.21734
81. Manz TA, Sholl DS (2010) Chemically meaningful atomic charges that reproduce the electrostatic potential in periodic and nonperiodic materials. *J Chem Theory Comput* 6(8):2455–2468. doi:10.1021/ct100125x
82. Manz TA, Sholl DS (2012) Improved atoms-in-molecule charge partitioning functional for simultaneously reproducing the electrostatic potential and chemical states in periodic and non-periodic materials. *J Chem Theory Comput* 8(8):2844–2867. doi:10.1021/ct3002199
83. Lillestolen TC, Wheatley RJ (2009) Atomic charge densities generated using an iterative stockholder procedure. *J Chem Phys* 131:144101. doi:10.1063/1.3243863
84. Lillestolen TC, Wheatley RJ (2008) Redefining the atom: atomic charge densities produced by an iterative stockholder approach. *Chem Commun* 45:5909–5911. doi:10.1039/b812691g
85. Verstraelen T, Ayers PW, Van Speybroeck V, Waroquier M (2012) The conformational sensitivity of iterative stockholder partitioning schemes. *Chem Phys Lett* 545:138–143. doi:10.1016/j.cplett.2012.07.028
86. Lee LP, Limas NG, Cole DJ, Payne MC, Skylaris CK, Manz TA (2014) Expanding the scope of density derived electrostatic and chemical charge partitioning to thousands of atoms. *J Chem Theory Comput* 10(12):5377–5390. doi:10.1021/ct500766v
87. Limas NG, Manz TA (2016) Introducing DDEC6 atomic population analysis: part 2. Computed results for a wide range of periodic and nonperiodic materials. *RSC Adv* 6(51):45727–45747. doi:10.1039/c6ra05507a
88. Manz TA, Limas NG (2016) Introducing DDEC6 atomic population analysis: part 1. Charge partitioning theory and methodology. *RSC Adv* 6(53):47771–47801. doi:10.1039/c6ra04656h
89. Marenich AV, Jerome SV, Cramer CJ, Truhlar DG (2012) Charge model 5: an extension of Hirshfeld population analysis for the accurate description of molecular interactions in gaseous and condensed phases. *J Chem Theory Comput* 8(2):527–541. doi:10.1021/ct200866d
90. Ayers PW (2006) Information theory, the shape function, and the Hirshfeld atom. *Theor Chem Acc* 115:370–378
91. Heidar-Zadeh F, Ayers PW, Bultinck P (2014) Deriving the Hirshfeld partitioning using distance metrics. *J Chem Phys* 141:094103
92. Heidar-Zadeh F, Ayers PW (2015) How pervasive is the Hirshfeld partitioning? *J Chem Phys* 142(4):044107. doi:10.1063/1.4905123
93. Tsallis C (1988) Possible generalization of Boltzmann–Gibbs statistics. *J Stat Phys* 52(1–2):479–487
94. Nielsen F, Nock R (2011) On the Rényi and Tsallis entropies and divergences for exponential families. [arXiv:1105.3259](https://arxiv.org/abs/1105.3259)
95. Lenzi EK, Mendes RS, da Silva LR (2000) Statistical mechanics based on Rényi entropy. *Phys A* 280(3–4):337–345
96. Rényi A (1960) On measures of information and entropy. In: *Proceedings of the fourth Berkeley symposium on mathematics, statistics, and probability*, pp 547–561
97. Nielsen F, Nock R (2012) A closed-form expression for the Sharma–Mittal entropy of exponential families. *J Phys A Math Theor*. doi:10.1088/1751-8113/45/3/032003
98. Gupta HC, Sharma BD (1976) On non-additive measures of inaccuracy. *Czechoslov Math J* 26:584–595
99. Sharma BD, Taneja IJ (1975) Entropy of type (alpha, beta) and other generalized measures in information theory. *Metrika* 22:205–215



100. Sharma BD, Mittal DP (1975) New nonadditive measures of entropy for discrete probability distributions. *J Math Sci* 10:28–40
101. Sharma BD, Mittal DP (1977) New nonadditive measures of relative information. *J Comb Inf Syst Sci* 2:122–132
102. Masi M (2005) A step beyond Tsallis and Rényi entropies. *Phys Lett A* 338(3–5):217–224. doi:[10.1016/j.physleta.2005.01.094](https://doi.org/10.1016/j.physleta.2005.01.094)
103. Esteban MD, Morales D (1995) A summary on entropy statistics. *Kybernetika* 31(4):337–346
104. Tsekouras GA, Tsallis C (2005) Generalized entropy arising from a distribution of  $q$  indices. *Phys Rev E*. doi:[10.1103/PhysRevE.71.046144](https://doi.org/10.1103/PhysRevE.71.046144)
105. Liu MZ, Vemuri BC, Amari SI, Nielsen F (2010) Total Bregman divergence and its applications to shape retrieval. In: 2010 IEEE conference on computer vision and pattern recognition. IEEE Computer Society, Los Alamitos, pp 3463–3468. doi:[10.1109/cvpr.2010.5539979](https://doi.org/10.1109/cvpr.2010.5539979)
106. Nielsen F, Nock R (2015) Total Jensen divergences: definition, properties, and clustering. In: 2015 IEEE international conference on acoustics, speech, and signal processing. International conference on acoustics speech and signal processing ICASSP, pp 2016–2020
107. Nock R, Nielsen F, Amari SI (2016) On conformal divergences and their population minimizers. *IEEE Trans Inf Theory* 62(1):527–538. doi:[10.1109/tit.2015.2448072](https://doi.org/10.1109/tit.2015.2448072)
108. Ali SM, Silvey SD (1966) A general class of coefficients of divergence of one distribution from another. *J R Stat Soc Ser B Methodol* 28:131–142
109. Morimoto T (1963) Markov processes and the H-theorem. *J Phys Soc Jpn* 18:328–331
110. Csiszár I (1963) Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Magyar Tud Akad Mat Kutato Int Kozl* 8:85–108
111. Chernoff H (1952) A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann Math Stat* 23:493–507
112. Dunlap BI, Rosch N, Trickey SB (2010) Variational fitting methods for electronic structure calculations. *Mol Phys* 108:3167–3180. doi:[10.1080/00268976.2010.518982](https://doi.org/10.1080/00268976.2010.518982)
113. Dunlap BI (2000) Robust and variational fitting: Removing the four-center integrals from center stage in quantum chemistry. *J Mol Struct THEOCHEM* 529:37–40
114. Dunlap BI (2000) Robust variational fitting: Gaspar’s variational exchange can accurately be treated analytically. *J Mol Struct THEOCHEM* 501:221–228
115. Dunlap BI (2000) Robust and variational fitting. *PCCP* 2(10):2113–2116
116. Dunlap BI, Connolly JWD, Sabin JR (1979) Some approximations in applications of X-alpha theory. *J Chem Phys* 71(8):3396–3402
117. Parr RG, Bartolotti LJ (1983) Some remarks on the density functional theory of few-electron systems. *J Phys Chem* 87:2810–2815
118. Ayers PW (2000) Density per particle as a descriptor of Coulombic systems. *Proc Natl Acad Sci* 97:1959–1964
119. Ayers PW, Cedillo A (2009) The shape function. In: Chattaraj PK (ed) *Chemical reactivity theory: a density functional view*. Taylor and Francis, Boca Raton, p 269
120. De Proft F, Ayers PW, Sen KD, Geerlings P (2004) On the importance of the “density per particle” (shape function) in the density functional theory. *J Chem Phys* 120:9969–9973
121. Borgoo A, Godefroid M, Indelicato P, De Proft F, Geerlings P (2007) Quantum similarity study of atomic density functions: insights from information theory and the role of relativistic effects. *J Chem Phys* 126(4):044102. doi: [10.1063/1.2428295](https://doi.org/10.1063/1.2428295)
122. Sen KD, De Proft F, Borgoo A, Geerlings P (2005) N-derivative of Shannon entropy of shape function for atoms. *Chem Phys Lett* 410(1–3):70–76
123. Borgoo A, Godefroid M, Sen KD, De Proft F, Geerlings P (2004) Quantum Similarity of atoms: a numerical Hartree–Fock and information theory approach. *Chem Phys Lett* 399(4–6):363–367