Constraining density functional approximations to yield self-interaction free potentials.

Nikitas I. Gidopoulos[1,2], Nektarios N. Lathiotakis[2]

[1] *ISIS, STFC, Rutherford Appleton Laboratory, Didcot, OX11 0QX, United Kingdom*

[2] *National Hellenic Research Foundation, 48 Vassileos Constantinou Ave. GR 11635 Athens, Greece*

(Dated: August 1, 2011)

Self-interactions (SIs) are a major problem in density functional approximations and the source of serious divergence from experimental results. Here, we propose to constrain the effective local potential to be SI free, even though it may correspond to a total energy that is contaminated with SIs. More specifically, we constrain the screening potential to be the electrostatic potential of a non-negative screening density of $N-1$ electrons. In this way, the optimal effective potentials exhibit the correct $1/r$ asymptotic decay resulting in significantly improved one-electron properties.

PACS numbers: 31.15.E-,31.15-p,31.15.xt,71.15.Mb

It is already thirty years since Perdew and Zunger in a seminal contribution [1] proposed to cure the self-interaction (SI) error in density functional approximations (DFAs). The SI error arises from the incomplete cancellation of the self-repulsion of the electron density ρ in the direct Coulomb or Hartree energy $U[\rho]$ by the approximate exchange energy functional $E_{\rm X}^{\rm DFA}[\rho]$. SI errors manifest in inaccuracies of DFAs [2] in many ways, e.g. in the calculation of binding energies[3], underestimation of activation energy barriers[4, 5], and in single-particle properties like ionization potentials (IPs) [6, 7], electron affinities (EAs) (unbound anions) [8] and band gaps of solids[9].

Perdew and Zunger [1] proposed a many-body SI correction energy term which, in the limit of a single electron, eliminates the SI error exactly. Their work initiated the field known as self-interaction corrected density functional theory (SIC-DFT). Unfortunately, the many-body generalization of the one-electron SI energy correction is not unique and to date an unambiguous definition is not available. Rigorously, we have a sufficient condition for an approximate exchange and correlation (XC) energy density functional $E_{\rm XC}[\rho]$ to be N-representable [10] and thus free from many-body SI errors. An important development in this area is the appreciation of the underlying relationship between the SI error with the fractional charge error [10]. The approximate treatment of many-body SI errors with SIC-DFT leads to single-particle equations with orbital dependent potentials, i.e. the minimization problem is significantly more complicated than an iterative diagonalization. For solids, SIC-DFT is expressed in terms of maximally localized Wannier states. An advantage of removing SI errors is that it improves orbital energies[11, 12]. These orbital energies, are commonly obtained as the eigenvalues of the non-diagonal Lagrange multiplier matrix employed to enforce orbital orthogonality, although the diagonal values of this matrix have also been proposed as appropriate [13]. Despite complications, SIC-DFT has been extensively developed and applied to a large variety of systems[14–19].

Probably the most serious flaw caused by SI errors lies in the asymptotic behavior of the Kohn-Sham (KS) potential [20]. If the cancellation of SI terms was complete then at infinity, the electron-electron contribution to the KS potential (Hartree and XC) should be $(N-1)/r$ where N is the number of electrons. The physical meaning is obvious, at infinity each electron feels the screening of the nuclear charge by the remaining $N-1$ electrons. The components of the Hartree potential $v_{\rm H}$ and of the exact XC potential $v_{\rm XC}$ to the asymptotic decay are N/r, and $-1/r$ respectively. However, the asymptotic decay of $v_{\rm XC}^{\rm DFA}$ in typical DFAs, like the local density approximation (LDA) or the generalized gradient approximation (GGA), does not follow a power law $(-c/r)$, but is exponentially fast $(c=0)$. Consequently an electron at infinity is over-screened by N rather than $N-1$ electrons. The incorrect asymptotic behavior has dramatic consequences on one-electron properties like the IPs, EAs and the fundamental gaps. It also impairs the optical spectrum through linear response in time dependent DFT.

Aiming to deal in an unambiguous manner with SIs in finite systems we decided, rather than focusing on the approximate Hartree (U) and $E_{\rm XC}$ energies (which remain contaminated with SI errors), to turn our attention instead to the effective local potential. On the latter physical constraints can be imposed in order to remove any effect of SI errors.

In KS theory, $v_{\rm HXC} = v_{\rm H} + v_{\rm XC}$ screens the nuclear attraction felt by a KS electron. By virtue of Poisson's equation,

$$\nabla^2 v_{\rm HXC}(\mathbf{r}) = -4\pi\rho_{\rm HXC}(\mathbf{r})\,. \qquad (1)$$

$\rho_{\rm HXC}$ is the density with electrostatic potential $v_{\rm HXC}$. From the asymptotic behavior of $v_{\rm HXC}^{\rm DFA}$ in LDA or GGA, we argued that $\int \rho_{\rm HXC}^{\rm DFA} = N$ and the nuclear charge is over-screened in these approximations.

In our KS equations, we introduce the screening potential $v_{\rm scr}$ in place of $v_{\rm HXC}^{\rm DFA}$. We constrain $v_{\rm scr}$ to be the electrostatic potential due to the density $\rho_{\rm scr}(\mathbf{r}) \geq 0$ of some $N-1$ electrons that screen the nuclear charge:

$$v_{\rm scr}(\mathbf{r}) = \int d\mathbf{r}'\, \frac{\rho_{\rm scr}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \qquad (2)$$

In contrast to $v_{\rm HXC} \doteq \delta(U[\rho]+E_{\rm XC}[\rho])/\delta\rho$, the potential $v_{\rm scr}$ is not the functional derivative of $U[\rho]+E_{\rm XC}[\rho]$.

Following the optimized potential method (OPM) [21–24], our single-particle KS equations have in place of $v_{\rm HXC}^{\rm DFA}$, the potential $v_{\rm scr}$ (2) which satisfies the constraints:

$$\int d\mathbf{r}\ \rho_{\rm scr}(\mathbf{r}) \ = \ N-1\,, \tag{3}$$

$$\rho_{\rm scr}(\mathbf{r}) \ \geq \ 0\,. \tag{4}$$

As in the OPM, our KS equations yield orbitals whose N-electron ground state density ρ minimizes the approximate KS total energy.

Eq. (3) is equivalent to constraining the XC density $\rho_{\rm XC}$ $(=\rho_{\rm HXC}-\rho)$ [25, 26] to integrate to -1 [25]. The normalization of the total screening charge (3) is necessary for the absence of SIs from the screening potential, since if each electron repelled itself partly, the total screening charge would exceed $N-1$. However, Eq. (3) is not sufficient for the absence of SI effects in $v_{\rm scr}$, as it would be energetically favorable to have SIs and over-screening of the nuclear charge locally near the system, plus a compensating negative charge far away from the system to satisfy (3).

Including condition (4), the non-negative $\rho_{\rm scr}(\mathbf{r})$ represents a physical density of $N-1$ electrons that screen the nuclear charge. There is no longer any freedom to have SIs and over-screening locally with a compensating negative charge at large distances. Hence, eqs. (3), (4) together become sufficient to ensure that $v_{\rm scr}$ is SI free. In a forthcoming publication we argue that the screening density of the exact KS potential (1) satisfies constraints (3), (4), and consequently the latter do not constitute an additional approximation.

We warn that if the screening density is expanded in a small finite basis set it is possible that employing (3) alone may result in a solution that appears physical. This is an artifact of the smallness of the basis set, and by increasing the size of the screening density basis set, the pathology of not having a sufficient condition will emerge.

The system of constraints (3), (4) is equivalent to the (numerically) simpler system (3), (5), with:

$$\int d\mathbf{r}\ |\rho_{\rm scr}(\mathbf{r})| = N-1\,. \tag{5}$$

Our search for the screening potential will be performed by expanding it in a basis

$$v_{\rm scr}(\mathbf{r}) = \sum_l v_l\ \xi_l(\mathbf{r}), \ \ \text{with}\ \ \xi_l(\mathbf{r}) = \int d\mathbf{r}'\ \frac{\chi_l(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{6}$$

where $\chi_l(\mathbf{r})$ is an auxiliary basis set, for example localized gaussians.

The minimization of the approximate total energy, $E_{\rm DFA}$, with respect to v_l under the conditions (3) and (5) leads to the variation equation

$$\frac{\delta E_{\rm DFA}}{\delta v_l} = \mu\, X_l + \lambda\, \bar{X}_l\,, \tag{7}$$

with $X_l = \int d\mathbf{r}\ \chi_l(\mathbf{r})$, $\bar{X}_l = \int d\mathbf{r}\ \chi_l(\mathbf{r})\,|\rho_{\rm scr}(\mathbf{r})|/\rho_{\rm scr}(\mathbf{r})$ and μ, λ are Lagrange multipliers to satisfy (3, 5).

The derivative on the left hand side is obtained, as in OPM, through a chain rule involving derivatives with respect to orbitals and the derivative with respect to the potential.

$$\frac{\delta E_{\rm DFA}}{\delta v_l} = 2\sum_{ia} \frac{v_{ia}^{\rm HXC} - v_{ia}^{\rm scr}}{\epsilon_i - \epsilon_a}\ S_{ia}^{(l)} \tag{8}$$

where i runs over occupied, a over unoccupied eigenorbitals of $v_{\rm scr}$, with ϵ_i and ϵ_a the corresponding eigenenergies. $v_{ia}^{\rm scr}$ and $v_{ia}^{\rm HXC}$ are the matrix elements of the potentials $v_{\rm scr}$ and $v_{\rm HXC}^{\rm DFA}(\doteq \delta(U[\rho]+E_{\rm XC}^{\rm DFA}[\rho])/\delta\rho)$ respectively. Finally,

$$S_{ia}^{(l)} = \int d\mathbf{r}\ \phi_a(\mathbf{r})\ \phi_i(\mathbf{r})\ \xi_l(\mathbf{r})\,. \tag{9}$$

Eqs. (7) and (8) define a non-linear system of equations with respect to v_l. This system can be solved, using an iterative scheme of two steps. In the first step, a linear system is solved by keeping the quantities ϕ_i, ϕ_a, ϵ_i, ϵ_a, $v_{ia}^{\rm HXC}$, and $S_{ia}^{(l)}$ frozen. In the second step, a single-electron Hamiltonian problem is solved with the potential obtained in the previous step and the quantities ϕ_i, ϕ_a, ϵ_i, ϵ_a, $v_{ia}^{\rm HXC}$, and $S_{ia}^{(l)}$ are updated. Our numerical implementation proved that this scheme is very efficient and usually only a few iterations are required to converge using a mixing scheme similar to Kohn-Sham iterative procedure. The potential obtained at the first step, i.e. when orbitals are frozen to the Kohn Sham orbitals is already a very good approximation to the local potential. The linear system that needs to be solved in each iteration has the form

$$\sum_l A_{kl} v_l = b_k + \mu\, X_k + \lambda\, \bar{X}_k\,, \tag{10}$$

with

$$A_{kl} = \sum_{ia} \frac{S_{ia}^{(k)}\ S_{ai}^{(l)}}{\epsilon_i - \epsilon_a}\,, \ \text{and} \ \ b_{kl} = \sum_{ia} \frac{S_{ia}^{(k)}\ v_{ia}^{\rm HXC}}{\epsilon_i - \epsilon_a} \tag{11}$$

The Lagrange multipliers μ, λ are given by the solution of a simple 2×2 linear system obtained from Eq. (10) by multiplying both sides by the inverse of A, then by X_k (or $\bar{X}_k$) and summing over k, and using Eqs. (3, 5).

Eqs. (10) and (11) constitute a simple problem in the OPM. The solution of Eq. (10) is complicated because often the matrix A is singular. This problem is well known in the OPM with finite basis sets [27–31], and the solution involves the inversion of A in the space of its non-singular eigenvectors, usually with a singular value decomposition (SVD). However, even after the SVD, and depending on the particular basis sets, the resulting effective potential may look unphysical. In a separate publication, we argue that behind the known mathematical problem lies a discontinuity of the optimal potential, when the orbital basis

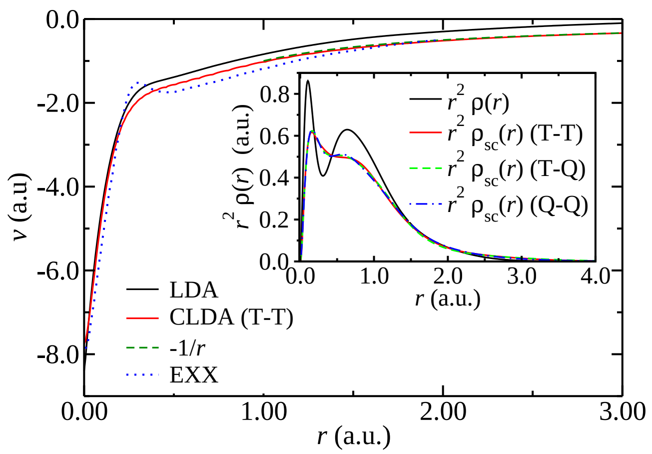

Figure 1: The effective potential and screening charge density (inset) for Ne atom. The effective potential EXX (from Ref. [29]) as well as $-1/r$ are shown. (X-Y) notation stands for cc-pVXZ and cc-pVYZ uncontracted for the orbitals and the auxiliary basis sets respectively.

set is truncated with a finite basis [32]. In the present work, this discontinuity is reduced by the restriction of the admissible potentials to satisfy conditions (3), (5).

To perform the SVD of A, we divide the space spanned by the eigenvectors of A in the null space (eigenvalues assumed zero) and the rest, using a small parameter θ as cut-off. In the non-singular subspace of eigenvectors we perform a usual inversion of the matrix. For the null subspace all eigenvalues of the inverse are set to zero. Sometimes this division is obvious, i.e. there is a clear gap in the eigenvalues of a few orders of magnitude. There are cases however, where the eigenvalues of A approach zero smoothly. Then, the potential changes slightly when the assumed non-singular space increases by lowering θ. There is a point however beyond which unphysical wiggles appear in the potential. Our strategy is to use a value of θ that is as small as possible without introducing unphysical features in the potential. For the systems and basis sets we considered, $\theta = 10^{-6}$ has proven a reasonable choice.

To illustrate our approach we chose the LDA functional[33] and we refer to the combined method as constrained LDA (CLDA). In Fig. 1, we show the LDA and CLDA potentials for Ne atom using finite basis sets as well as the screening density. Evidently, the effective potential obtained with CLDA has the correct asymptotic behavior. Also, the screening density with respect to the different basis sets is also converged. Potentials with the correct asymptotics are also obtained for larger systems like CO molecule as shown in Fig. 2. In the inset of Fig. 2 we show that the IP value for CO is essentially independent of θ. In the top of Table I, we show the IPs calculated with CLDA and with LDA, for various atomic and molecular systems, obtained as the negative of the one-electron energy, E_{H}, for the highest

	Basis	Q_n	ΔE	IP(LDA)	IP(CLDA)	Exp
He	T-Q	0	$1.5 \cdot 10^{-3}$	15.46	23.14	24.6
Be	T-T	$3.0 \cdot 10^{-4}$	$1.1 \cdot 10^{-4}$	5.59	8.62	9.32
Ne	T-T	0	$2.7 \cdot 10^{-5}$	13.16	18.94	21.6
H_2O	T-T	$6.0 \cdot 10^{-5}$	$1.1 \cdot 10^{-5}$	6.96	11.24	12.8
NH_3	T-T	$6.0 \cdot 10^{-5}$	$8.2 \cdot 10^{-6}$	6.00	9.81	10.8
CH_4	D-D	$1.5 \cdot 10^{-3}$	$2.7 \cdot 10^{-4}$	9.28	12.52	14.4
C_2H_2	D-D	$1.9 \cdot 10^{-4}$	$4.1 \cdot 10^{-5}$	7.02	10.63	11.5
C_2H_4	D-D	$3.9 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$	6.67	9.57	10.7
CO	D-D	$2.0 \cdot 10^{-5}$	$3.6 \cdot 10^{-4}$	8.75	12.73	14.1
NaCl	D-D	$1.2 \cdot 10^{-2}$	$6.8 \cdot 10^{-4}$	5.13	7.87	8.93
F^-	T^a-T	$1.0 \cdot 10^{-4}$	$2.7 \cdot 10^{-5}$	$E_{\mathrm{H}} > 0$	2.23	3.34
Cl^-	T^a-T	$1.0 \cdot 10^{-5}$	$1.6 \cdot 10^{-4}$	$E_{\mathrm{H}} > 0$	2.61	3.61
OH^-	T^a-T	$4.0 \cdot 10^{-5}$	$1.4 \cdot 10^{-4}$	$E_{\mathrm{H}} > 0$	0.99	1.83
NH_2^-	T^a-T	$4.0 \cdot 10^{-5}$	$8.2 \cdot 10^{-4}$	$E_{\mathrm{H}} > 0$	0.18	0.77
CN^-	T^a-T	$1.0 \cdot 10^{-5}$	$1.1 \cdot 10^{-4}$	0.13	2.87	3.77

[a]For negative ions, aug-cc-pVXZ basis sets were used for the orbital expansion.

Table I: The total energy difference ΔE of CLDA from plain LDA the total negative screening charge Q_n (in e) and the IPs for selected atoms, molecules (top) and negative ions (bottom). Basis set notation is explained in the caption of Fig. 1. For the neutral systems we compare with experimental values of the IP, while for the negative ions with experimental values of the EA of the corresponding neutral system. All energies are in eV.

occupied molecular orbital. The IPs from CLDA are on average roughly 10% underestimated. This divergence should be contrasted to the dramatic 40% errors of plain LDA. Given the severe underestimation of IPs and the fundamental gaps of solids by LDA and GGA, our approach offers a significant qualitative improvement. Contrary to LDA, negative ions are predicted to be bound by CLDA as shown in the bottom of Table I. Even though EAs of neutral systems (predicted as $-E_{\mathrm{H}}$ of the corresponding negative ions), are underestimated compared to experiment by about 40%, it is nevertheless encouraging that qualitatively correct EAs can be obtained with CLDA.

The differences of the total energies obtained with CLDA from those obtained with LDA are also shown in Table I. These differences are very small for all systems i.e. the additional constraint on the potential does not change the total energy. The total negative screening charge, Q_n, is shown in Table I as a measure of how well the positivity condition is fulfilled by the optimal potential. Although we did not manage to eliminate it completely, Q_n is very small compared to the total screening charge and does not affect the quality of our effective potential.

Until now, the main errors stemming from SIs were not in the total energy but resulted from deficiencies of the local potential. In passing we remark that splitting $E_{\mathrm{XC}} = E_{\mathrm{X}} + E_{\mathrm{C}}$ as a sum of exchange and correlation energies and then treating exchange exactly (EXX) is not necessarily the best strategy [34]. The main advantage of

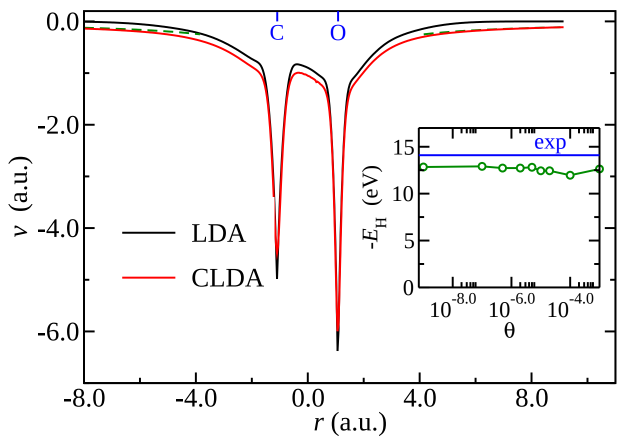

Figure 2: The effective potential for CO molecule along the molecular axis. Green dashed lines indicate the correct $\pm 1/r$ asymptotic behavior with r measured from the molecular center. In the inset, the $E_{\rm H}$ as a function of the SVD parameter θ is shown with the horizontal blue line indicating the experimental value of the IP.

employing EXX is the cancellation of SIs and the quality of the KS potential, however at a computational cost compared with LDA/GGA, and an ensuing complicated description of correlation through non-local orbital functionals [35]. In fact an appropriate non-local $E_{\rm C}[\rho]$ of cost no-higher than EXX is not available yet.

Attempting to address the SI problem in DFAs for finite systems, we noted the ambiguity in the quantification of the SI error in the Hartree and XC energies. Nevertheless, we proposed unambiguous constraints for the KS potential which keep the scaling of computational cost at the level of the corresponding DFA and are sufficient to eliminate any effects of SIs from the potential (where it matters) with a minimal increase of the total energy. For LDA, the constrained KS potentials have the correct asymptotic behavior and give significantly improved IPs over the unconstrained LDA results. At the same time, we have kept the description of XC together which has advantages, for example it exploits the cancellation of errors in E_{XC} and provides an improved description of electron-pair bonds [34].

Interesting questions regarding our approach are how it applies to extended systems (N, $N-1$ are the same) and the implication of our constraints to the dissociation of diatomic molecules and to size-consistency. These questions are related to the localization or not of the XC density $\rho_{\rm XC}$ in the solid or as the molecule dissociates at large distances between the fragments.

[1] J. P. Perdew, A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[2] S. Kümmel, L. Kronik, Rev. Mod. Phys. **80**, 3 (2008).
[3] S. Kurth, J. P. Perdew, P. Blaha, Int. J. Quant. Chem. **75**, 889 (1999).
[4] K. D. Dobbs, and D. A. Dixon, J. Chem. Phys. **98**, 12584 (1994)
[5] T. N. Truong and W. Duncan, J. Chem. Phys. **101**, 7408 (1994).
[6] J. P. Perdew, R. G. Parr, M. Levy, J. L. Balduz, Jr., Phys. Rev. Lett. **49**, 1691 (1982).
[7] S. Goedecker, C. J. Umrigar, Phys. Rev. A **55**, 1765 (1997).
[8] M.C. Kim, E. Sim, K. Burke, J. Chem. Phys. **134**, 171103, (2011).
[9] K. Terakura, T. Oguchi, A. R. Williams, and J. Kübler, Phys. Rev. B **30**, 4734 (1984).
[10] P. Mori-Sánchez, A.J. Cohen, W. Yang, J. Chem. Phys., **125**, 201102 (2006).
[11] J. G. Harrison, R. A. Heaton, and C. C. Lin, J. Phys. B, **16**, 2079 (1983).
[12] T. Fujiwara, M. Arai, Y. Ishii, in *Strong Coulomb correlations in electronic structure calculations*, V. I. Anisimov, ed. Gordon and Breach (2000), p. 167.
[13] O. A. Vydrov, G. E. Scuseria, and J. P. Perdew, J. Chem. Phys. **126** 154109 (2007).
[14] P. Strange, A. Svane, W.M. Temmerman, Z.Szotek and H. Winter, Nature **399** 756 (1999).
[15] I. D. Hughes, et al, Nature **446** 650 (2007)
[16] C. Toher, and S. Sanvito, Phys. Rev. Lett. **99**, 056801 (2007).
[17] A. Filippetti and N. Spaldin, Phys. Rev. B67, 125109 (2003).
[18] D. Vogel, P. Krüger, J. Pollman, Phys. Rev. B **54**, 5495 (1996).
[19] A. Ruzsinszky, J. P. Perdew, G. I. Csonka, G. E. Scuseria, and O. A. Vydrov, Phys. Rev. A **77**, 060502(R) (2008).
[20] J. P. Perdew, Adv. Quantum Chem. **21**, 113 (1990).
[21] R.T. Sharp, G.K. Horton, Phys. Rev. **90**, 317, (1953).
[22] J.D. Talman, W.F. Shadwick, Phys. Rev. A **14**, 36 (1976).
[23] S. Kümmel, L. Kronik, Rev. Mod. Phys. **80**, 3, (2008).
[24] F.A. Bulat, M. Levy, Phys. Rev. A **80**, 052510, (2009).
[25] A. Görling, Phys. Rev. Lett. **83**, 5459, (1999).
[26] S. Liu, P.W. Ayers, R.G. Parr, J. Chem. Phys., **111**, 6197 (1999).
[27] V. N. Staroverov, G. E. Scuseria, E. R. Davidson, J. Chem. Phys. **124**, 141103, (2006).
[28] S. Hirata et al, J. Chem. Phys. **115**, 1635, (2001).
[29] A. Heßelmann, A.W. Götz, F. Della Sala, A. Görling, J. Chem. Phys. **127**, 054102 (2007).
[30] C. Kollmar, M. Filatov, J. Chem. Phys. **127**, 114104, (2007).
[31] V.N. Glushkov, S.I. Fesenko, H.M. Polatoglou, Theor. Chem. Acc. **124**, 365, (2009)
[32] N. I. Gidopoulos, N. N. Lathiotakis, to be published.
[33] J. C. Slater, The Self-Consistent Field for Molecules and Solids: Quantum Theory of Molecules and Solids, vol. 4, McGraw-Hill, New York (1974); S.H. Vosko, L. Wilk and M. Nusair. Can. J. Phys. **58**, 1200, (1980).
[34] E.J. Baerends, O.V. Gritsenko, J. Chem. Phys., **123**, 062202 (2005).
[35] N.I. Gidopoulos, Phys Rev A, **83**, 040502(R) (2011).