

Approximate Functionals for Multistate Density Functional Theory

Alexander Humeniuk*

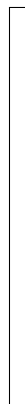
*current address: Department of Chemistry, Graduate School of Science, Kyoto University,
Kitashirakawa-Oiwakecho, Sakyo-Ku, Kyoto 606-8502, Japan*

E-mail: alexander.humeniuk@gmail.com

Abstract

Recently Lu and Gao [J. Phys. Chem. Lett. 13, 33, 7762-7769 (2022)] published a new, rigorous density functional theory for excited states and proved that the projection of the kinetic and electron repulsion operators into the subspace of the lowest electronic states are a universal functional of the matrix density $\mathbf{D}(\mathbf{r})$. This is the first attempt to find an approximation to the multistate universal functional $\mathcal{F}[\mathbf{D}(\mathbf{r})]$. It is shown that \mathcal{F} (i) does not explicitly depend on the number of electronic states and (ii) is an analytic matrix functional. The Thomas-Fermi-Dirac-von Weizsäcker model and the correlation energy of the homogeneous electron gas are turned into matrix functionals guided by two principles: That each matrix functional should transform properly under basis set transformations, and that the ground state functional should be recovered for a single electronic state. Lieb-Oxford-like bounds on the average kinetic and electron repulsion energies in the subspace are given. When evaluated on the numerically exact matrix density of LiF, this simple approximation reproduces the matrix elements of the electron repulsion operator in the basis of the exact eigenstates accurately for all bond lengths. In particular the off-diagonal elements of the effective Hamiltonian that come from the interactions of different electronic states can be calculated with the same or better accuracy than the diagonal elements. Unsurprisingly, the largest error comes from the kinetic energy functional. More exact conditions that constrain the functional form of \mathcal{F} are needed to go beyond the local density approximation.

Graphical TOC Entry



Keywords

excited states, time-independent density functional theory, multiconfigurational DFT, static correlation

1 Introduction

Kohn-Sham (KS) density functional theory (DFT)^{1,2} is probably the single most used electronic structure method in academia and industry, because it can be operated as a black box and often gives reasonable answers without requiring excessive computational time. For molecules, linear-response³ time-dependent⁴ DFT (LR-TDDFT) provides excited state properties very efficiently, but it has problems with core-excitations,^{5,6} doubly excited states,⁷ charge transfer states,⁸ Rydberg states⁸ and conical intersections to the ground state.⁷ Although some of these issues can be fixed with tuned⁹ range-separated¹⁰ or projected⁶ hybrid functionals or spin-flip DFT,¹¹ whenever the electronic structure becomes more complicated, multi-reference wavefunction-based methods have to be employed.

In wavefunction theory (WFT) near-degenerate states can be included in a multi-state treatment through configuration interaction (CI) or multi-configurational self-consistent field (MC-SCF). A few important configurations can capture the "static correlation" and give a qualitatively correct electronic structure. The remaining configurations each have a very small weight in the wavefunction, but since there are many of them, huge CI expansions are needed to get accurate, quantitative energies. It is the treatment of this "dynamic correlation" by perturbation theory or configuration interaction that makes wavefunction-based methods so expensive. In density functional theory, on the other hand, the exchange-correlation functional already accounts for the "dynamic correlation", while the "static correlation" is problematic. Since DFT is formulated in terms of the electron density rather than wavefunctions, it is not clear how to compute the interaction of states, which gives the off-diagonal elements in some effective Hamiltonian. The central question in this article therefore is the following: How does one get the off-diagonal matrix elements for treating static correlation with DFT in a configuration-interaction-like manner?

The development of a multi-configurational and time-independent, variational DFT is an active research area, but a far cry from the established TD-DFT. A number of variational density-functional theories for excited states have been formulated, most notably

Theophilou's subspace DFT,¹² Gross-Oliveira-Kohn ensemble DFT¹³ and Levy & Nagy's variational DFT for individual excited states.¹⁴ In those theories, the universal excited-state functional depends on additional parameters apart from an electronic density, which allow to distinguish between the excited state and the ground state: In ensemble density functional theory^{13,15} there is a dependence on a vector of ensemble weights and Levy & Nagy's excited state functional¹⁴ depends on both the excited and the ground state densities. In addition the functionals seem to depend on the number of electronic states, so that a different functional would be needed for 2,3 etc. states: In ensemble DFT the length of the weight vector changes with the number of states N ; in Levy and Nagy's theory the constrained search in Eqn.(2) of Ref.¹⁴ is over all wavefunctions that are orthogonal to the first $N - 1$ states, which introduces an implicit dependence on N . In Görling's Kohn-Sham formalism for excited states the functionals depends on the ground state density and the index of the excited state.¹⁶ Ayers showed that the Coulomb interaction is special, because the density of an excited state of a Coulombic system also encodes the excitation level, so that no additional parameters are needed.¹⁷ The universal functional for any electronic state depends only on the electronic density, provided that the external potential is Coulombic.¹⁷ However, although the existence of those universal functionals can be proven, it is not clear how to approximate them for practical calculations and if or how they differ from the ground state functional.

Most publications on time-independent DFT are formal. They derive their value from providing justification for the use of some simpler theory such as Δ SCF which employs a ground state functional. Numerical calculations using variational DFT methods are rare and usually reuse the ground state functional. Pastorczak et al. tested ensemble DFT methods on small molecules using ground state functionals or exact exchange with self- and ghost-interaction corrections.¹⁸ They found that the corrections do not necessarily improve the energies and that ensemble DFT cannot compete with TD-DFT in terms of accuracy.¹⁸

Multistate density functional theory (MSDFT)¹⁹ predates the rigorous theorems of Lu

and Gao²⁰ that establish a existence of a universal functional of the matrix density. Originally MSDFT was developed as a hybrid between DFT and wavefunction theory (WFT) for treating multiconfigurational problems that combines the best of both worlds. Dynamic correlation is incorporated into the configurations by DFT, while static correlation is accounted for by letting the configurations interact. Diagonalizing an effective Hamiltonian leads to the adiabatic eigenstates. MSDFT is akin to multiconfigurational wavefunction-based methods such MCSCF or multireference configuration interaction (MRCI), but has the advantage that only a minimal number of configurations is needed²¹ since DFT is very good at handling the dynamic correlation.²² The coupling between the configurations is approximated by the matrix elements of the configurations (the WFT part) and a *transition density functional* (TDF),²⁴ has been approximated as the overlap-weighted average correlation energy of the interacting states,²⁵ $\mathcal{E}_c^{\text{TDF}}[\rho_{AB}] = \frac{1}{2}S_{AB} (E_c^{\text{KS}}[\rho_A] + E_c^{\text{KS}}[\rho_B])$, or by the multiplet-sum method^{23,26} that allows to determine the coupling rigorously from the requirement that states in a multiplet (same total spin but different spin projections) are degenerate in the absence of a magnetic field. These approximations make use of existing, Kohn-Sham density functional approximations (DFA) for the ground state and are born out by encouraging benchmark calculations.²⁷ But in view of Lu and Gao's multistate equivalent of the Hohenberg-Kohn theorem, which states that there is a one-to-one correspondence between the matrix density of the lowest few states and the Hamiltonian matrix, it is clear that $\mathcal{E}_c^{\text{TDF}}$ cannot be a functional of a individual state densities ρ_A , ρ_B or the transition density ρ_{AB} . Therefore Lu and Gao replaced the transition density functional by a *multistate correlation matrix functional* $\mathcal{E}_c[\mathbf{D}]$, which depends on the whole matrix density $\mathbf{D}(\mathbf{r})$ rather than on the individual state or transition densities.²¹ However, because all the complexity is packed into one functional, it is difficult to reason about the properties of \mathcal{E}_c .

Another approach for combining WFT and DFT is multiconfiguration pair-density functional theory (MC-PDFT).²⁸ It uses CASSCF wavefunctions to calculate the kinetic energy, nuclear attraction and direct part of the electron-electron repulsion energy, while the in-

direct part is computed from the one-particle density and the on-top pair density using a new type of density functional. In this way double counting of the correlation energy is avoided. Double counting can also be avoided by splitting the Coulomb interaction into a short- and a long-range part and treating the first with DFT and the latter with WFT.²² A variational version of pair-density functional theory has been published recently.²⁹ Although the exact pair-density functional, if it even exists, is unknown, ground state functionals of the total density and spin polarization can be "translated" into functionals of the total density and on-top pair density.²⁸ When states interact strongly around conical intersections or avoided crossings, they cannot be calculated separately as this can lead to spurious crossings of potential energy surfaces.³⁰ State-interaction pair-density functional theory (SI-PDFT)³⁰ remedies this problem by diagonalizing the effective Hamiltonian in a final step. The diagonal elements of the Hamiltonian come from MC-PDFT calculations for each state using the translated functional. For lack of a prescription for the off-diagonal Hamiltonian matrix elements in the DFT framework, a kind of diabaticization scheme is employed to get non-zero couplings between the states.³⁰ A later multistate (MS) version of PDFT, called VMS-PDFT (V for variational), determines the unitary transformation from the SA-CASSCF wavefunctions to a set of intermediate states, which give rise to off-diagonal couplings, by maximizing the trace of the effective Hamiltonian.³¹ The dependence of the trace on the basis transformation is an artifact of making different approximations to the diagonal and off-diagonal elements. Sand et al. state that "the evaluation of off-diagonal elements in an effective Hamiltonian is not well-defined, as no density functional that couples two interacting states is available".³⁰ The purpose of this article is to show that it is indeed possible to obtain off-diagonal elements with the help of matrix density functional theory, and that it is not necessary to treat diagonal and off-diagonal elements differently.

In general, existing variational DFT methods suffer from the lack of a clear route to designing new functionals that go beyond repurposing existing KS-DFT functionals. The other open question is how to compute matrix elements for state interactions consistently

with density functional theory.

The rigorous, multistate DFT (MSDFT) theory by Lu and Gao²⁰ might provide answers to those questions. First of all, the universal functional only depends on a matrix representation of the one-particle density operator, the *matrix density*. The matrix density contains the densities of the electronic states on the diagonal and the transition densities between them in the off-diagonal elements. It represents the electron density operator in a subspace of electronic states. It should not be confused with the density matrix of a single electronic state. Each matrix element is a function of space. And although the number of rows and columns of the matrix density is equal to the number of electronic states that are considered, I will argue that the functional of Lu and Gao does not depend on the number of states, because it is an *analytic matrix functional*. This property suggests a simple path to an approximate multistate functional: by retaining the functional form of an approximate ground state functional and replacing the scalar density with the matrix density.³² The universal matrix functional provides not only the state energies but also the off-diagonal elements of the Hamiltonian. MSDFT resembles the state-averaged MCSCF of wavefunction theory, however, it is formulated in terms of electron densities and transition densities rather than wavefunctions. The interaction between states is more transparent than in ensemble DFT, where the densities of the individual states are combined into a single ensemble density, whose physical meaning is not so obvious.

Outline of the article: Section 1.1 summarizes Lu and Gao's theorems and highlights the importance of how the universal functional transforms under a change of basis. The basic recipe for translating ground state Kohn-Sham functionals into matrix density functionals by repurposing the coefficients of their Taylor series and the limitations of this approach are outlined in section 1.2. This idea only works if the universal functional $\mathcal{F}[\mathbf{D}(\mathbf{r})]$ has certain mathematical properties: Section 2.1 introduces the term *analytic matrix functional* to characterize \mathcal{F} and sketches a proof that $\mathcal{F}[\mathbf{D}(\mathbf{r})]$ does not explicitly depend on the dimension of the subspace, i.e. the dimension of the matrix $\mathbf{D}(\mathbf{r})$. In section 2.2 a local

density approximation to the universal functional is presented. The approximate functional is tested on the avoided crossing in LiF (section 3) and section 4 concludes.

Proofs of the lower bounds on the average kinetic and electron-repulsion energies in terms of the subspace density are relegated to section 5 of the supporting information.

1.1 Rigorous MSDFT: Lu & Gao's Theorems

Recently Lu and Gao²⁰ published a new, rigorous density functional theory for excited states, in which the ground and excited states are treated on an equal footing. The fundamental variable is the projection of the one-particle density operator $\hat{\rho}(\mathbf{r}) = \sum_{a=1}^n \delta(\hat{\mathbf{r}}_a - \mathbf{r})$ on the subspace \mathbb{V}^N of the lowest N eigenstates, where $\hat{\mathbf{r}}_a$ is the position operator of electron a and the sum goes over all $a = 1, \dots, n$ electrons. The role of the scalar density $\rho(\mathbf{r})$ from the Hohenberg-Kohn theorem is taken by the multistate matrix density $\mathbf{D}(\mathbf{r})$, which is a matrix function of position \mathbf{r} having the state densities $D_{II}(\mathbf{r}) = \rho_I(\mathbf{r})$ on its diagonal and the transition densities $D_{IJ}(\mathbf{r}) = \langle \Psi_I | \sum_{a=1}^n \delta(\hat{\mathbf{r}}_a - \mathbf{r}) | \Psi_J \rangle$ on the off-diagonal. Building on Theophilou's theorem,¹² which states that the subspace of the lowest N electronic states is uniquely determined by the average density of those states, Lu and Gao prove two theorems that give access to the individual eigenenergies:

(1) Firstly they showed that the projection of the Hamiltonian operator onto the subspace \mathbb{V}^N is a functional $\mathbf{H}[\mathbf{D}(\mathbf{r})]$ of the matrix density $\mathbf{D}(\mathbf{r})$,³³

$$\mathbf{H}[\mathbf{D}(\mathbf{r})] = \mathcal{F}[\mathbf{D}(\mathbf{r})] + \int d^3r v(\mathbf{r})\mathbf{D}(\mathbf{r}). \quad (1)$$

This multi-state analogue of the Hohenberg-Kohn theorem states that the Hamiltonian can be split into an external potential (the attractive nuclear potential $v(\mathbf{r})$ which differs for each molecule) and a universal functional $\mathcal{F}[\mathbf{D}(\mathbf{r})]$ for the electronic kinetic energy and electron-electron repulsion (that is the same for all systems).

An important consequence of the fact that both $\mathbf{D}(\mathbf{r})$ and $\mathbf{H}[\mathbf{D}(\mathbf{r})]$ are projections of op-

erators onto the same subspace, is that the universal matrix functional transforms bilinearly under a change of basis within the subspace.³⁴ If the states in the subspace are transformed as $|\Psi'_I\rangle = \sum_{J=1}^N L_{IJ}^* |\Psi_J\rangle$, the matrix density in the new basis becomes $\mathbf{D}'(\mathbf{r}) = \mathbf{L}\mathbf{D}(\mathbf{r})\mathbf{L}^\dagger$, and the Hamiltonian becomes $\mathbf{H}' = \mathbf{L}\mathbf{H}\mathbf{L}^\dagger$. Consequently the universal functional has to transform as²⁰

$$\mathcal{F}[\mathbf{L}\mathbf{D}(\mathbf{r})\mathbf{L}^\dagger] = \mathbf{L}\mathcal{F}[\mathbf{D}(\mathbf{r})]\mathbf{L}^\dagger \quad (2)$$

under a basis transformation \mathbf{L} with $\mathbf{L}^{-1} = \mathbf{L}^\dagger$. Note that $\mathbf{D}(\mathbf{r})$ is an $N \times N$ matrix function of \mathbf{r} , but the basis transformation \mathbf{L} is just an $N \times N$ matrix.

(2) Secondly Lu and Gao formulated a variational principle for finding the exact $\mathbf{D}(\mathbf{r})$, which is a consequence of theorem 2 of Theophilou's subspace DFT.¹² Let the N orthonormal trial vectors Ψ'_I span the subspace $\mathbb{V}^{N'}$. Then the average subspace energy, the multistate (MS) energy in the notation of Lu and Gao,

$$E_{\text{MS}}[\{\Psi'_I\}] = \frac{1}{N} \sum_{I=1}^N \langle \Psi'_I | \hat{H} | \Psi'_I \rangle, \quad (3)$$

of the trial space is larger or equal than the subspace energy of the exact subspace (which is spanned by the exact lowest N eigenstates Ψ_I), $E_{\text{MS}}[\mathbb{V}^N] \leq E_{\text{MS}}[\mathbb{V}^{N'}]$. Since according to Theophilou's theorems 3 to 5 the subspace is uniquely determined by the subspace density, $\rho_{\mathbb{V}^N}(\mathbf{r}) = \frac{1}{N} \sum_{I=1}^N D_{II}(\mathbf{r})$, it follows (theorem 2 of Lu and Gao²⁰) that the multistate energy of any trial matrix density $\mathbf{D}'(\mathbf{r})$ is larger or equal to the exact multistate energy, $E_{\text{MS}}[\mathbf{D}(\mathbf{r})] \leq E_{\text{MS}}[\mathbf{D}'(\mathbf{r})]$.

The equality holds when the subspaces are the same $\mathbb{V}^N = \mathbb{V}^{N'}$, in which case $\mathbf{D}'(\mathbf{r})$ and the exact $\mathbf{D}(\mathbf{r})$ as well as the Hamiltonians $\mathbf{H}' = \mathbf{H}[\mathbf{D}'(\mathbf{r})]$ and the exact $\mathbf{H}[\mathbf{D}(\mathbf{r})]$, respectively, differ only by a basis transformation within the subspace. Therefore the subspace of the lowest eigenstates can be found by minimizing $E_{\text{MS}}[\mathbf{D}(\mathbf{r})]$ as a functional of $\mathbf{D}(\mathbf{r})$. Since the Hamiltonian in the basis of the eigenstates should be diagonal, the eigenenergies can be obtained by diagonalizing the $N \times N$ matrix $\mathbf{H}[\mathbf{D}(\mathbf{r})]$.

The significance of the theorems by Lu and Gao is not so much the existence of the mapping $\mathbf{D}(\mathbf{r}) \rightarrow \mathbf{H}$ (which is already guaranteed by the HK theorem¹⁶) but that their universal matrix functional has the transformation property 2. This restricts the structure of the universal functional and gives some hints on how it can be approximated. It constrains the search for matrix functionals to the class of *analytic matrix functionals*, as will be explained later on. It also implies that its functional form is independent of the number of states in the subspace, so that the same functional holds for N states or for a single state 1.

1.2 Basic Idea of Translating KS Functionals for MSDFT

An important open question is how to find approximate multistate density functionals. In a preprint on Chemrxiv³² Lu and Gao mention that a local ground state functional can be turned into a matrix functional by applying it to the eigenvalues of the matrix density. I wish to investigate this idea further.

In the local density approximation (LDA) the exchange-correlation and kinetic energy *functionals* of the density $\rho(\mathbf{r})$, $E_{xc}[\rho]$ and $T[\rho]$, respectively, are approximated by integrals over *functions*,

$$E_{xc}[\rho(\mathbf{r})] = \int d^3r \epsilon_{xc}(\rho(\mathbf{r})), \quad (4)$$

$$T[\rho(\mathbf{r})] = \int d^3r t(\rho(\mathbf{r})). \quad (5)$$

The exchange-correlation energy density $\epsilon_{xc}(\rho)$ and the kinetic energy density $t(\rho)$ are functions of a scalar variable. Let us call such a function $f : \mathbb{R} \mapsto \mathbb{R}$. Assuming that f is an analytic function, so that all the derivatives $f^{(k)}(\rho) = \frac{d^k f}{d\rho^k}$ exist, f can be expanded in a Taylor series around $\rho = 0$,

$$f(\rho) = \sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(0) \rho^k. \quad (6)$$

Now it is easy to turn the function $f(\rho)$ into a matrix function $\mathbf{f}(\mathbf{D}) : \mathbb{R}^{N \times N} \mapsto \mathbb{R}^{N \times N}$ of the

matrix density \mathbf{D} , by taking the coefficients of the Taylor series and replacing ρ by \mathbf{D} ,

$$\mathbf{f}(\mathbf{D}) = \sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(0) \mathbf{D}^k, \quad (7)$$

where $\mathbf{D}^k = \underbrace{\mathbf{D} \cdot \mathbf{D} \cdot \dots \cdot \mathbf{D}}_{k \text{ times}}$ is the k -th matrix power. The corresponding LDA matrix functional $\mathcal{F}[\mathbf{D}]$ becomes

$$\mathcal{F}[\mathbf{D}(\mathbf{r})] = \int d^3r \mathbf{f}(\mathbf{D}(\mathbf{r})). \quad (8)$$

A matrix function can simply be computed by applying the scalar function to each of its eigenvalues, but the definition via the Taylor series generalizes more easily to the multivariate case.

If the scalar functional $\mathcal{F}[\rho]$ only depends on the total electronic densities (summed over spins), as in the LDA, this procedure determines a unique matrix functional $\mathcal{F}[\mathbf{D}]$. However, to account for inhomogeneity or spin-polarization, the energy density function has to depend on multiple arguments. For example, the exchange-correlation energy density in the generalized gradient approximation (GGA), $\epsilon_{xc}(\rho(\mathbf{r}), \nabla\rho(\mathbf{r}))$, or the von-Weizsäcker (vW) kinetic energy density $t_{vW}(\rho(\mathbf{r}), \nabla\rho(\mathbf{r}))$ are functions of both the density and its gradient. To translate a function $f(\rho, \nabla\rho)$ into a matrix function $\mathbf{f}(\mathbf{D}, \nabla\mathbf{D})$, we follow the same recipe as for the single variable case, and expand f into a multivariate Taylor series. Denoting the partial derivatives by $f^{(m,n)} = \frac{\partial^{m+n} f}{\partial \rho^m (\nabla\rho)^n}(0,0)$, the first few terms of the series around $(\rho, \nabla\rho) = (0,0)$ are,

$$f(\rho, \nabla\rho) = f(0,0) + f^{(1,0)} \rho + f^{(0,1)} \nabla\rho + \frac{1}{2} f^{(1,1)} \rho(\nabla\rho) + \dots \quad (9)$$

When substituting the matrices $\mathbf{D}, \nabla\mathbf{D}$ for the scalar variables $\rho, \nabla\rho$, ambiguities arise, because in matrix multiplication the order of the factors matters. It is not clear, for instance, if the term $\rho(\nabla\rho)$ should be replaced by $\mathbf{D} \cdot \nabla\mathbf{D}$ or by $\nabla\mathbf{D} \cdot \mathbf{D}$ or the average of the two. Since $\mathbf{D}(\mathbf{r})$ and $\nabla\mathbf{D}(\mathbf{r})$ do not commute, $[\mathbf{D}(\mathbf{r}), \nabla\mathbf{D}(\mathbf{r})] \neq 0$, any order of factors will give a

different matrix functional.

Similar ambiguities arise with the local spin-density approximation (LSDA), where $\epsilon_{xc}(\rho^\uparrow(\mathbf{r}), \rho^\downarrow(\mathbf{r}))$ depends on the spin-up and spin-down densities, since the matrix densities for different electron spins in general do not commute, $[\mathbf{D}^\uparrow(\mathbf{r}), \mathbf{D}^\downarrow(\mathbf{r})] \neq 0$.

One sees that from one multivariate ground state functional an infinite number of matrix-density functionals can be derived. The indeterminacy can be reduced to some extent by enforcing known properties of the exact multistate functional. For example, the multistate-equivalent of the von-Weizsäcker kinetic energy functional should yield the exact kinetic energy matrix for one-electron systems.

2 Theory

In the following the subscripts $a, b = 1, \dots, n$ enumerate electrons, Latin capital letters $I, J, K, L = 1, \dots, N$ enumerate electronic states in the subspace. Superscripts α, β denote the electronic spin. Atomic units are used throughout.

2.1 Properties of the Matrix Density Functional.

The transformation property 2 implies that \mathcal{F} is an *analytic matrix functional*. As we know, an analytic function of a scalar variable, $f(z) : \mathbb{C} \rightarrow \mathbb{C}$ is defined by its Taylor series around some point z_0 :

$$f(z) = \sum_{k=0}^{\infty} \frac{f^{(k)}}{k!} (z - z_0)^k \quad (10)$$

With the help of the Taylor series the scalar function can be turned into a analytic matrix function, $f(\mathbf{Z}) : \mathbb{C}^{N \times N} \rightarrow \mathbb{C}^{N \times N}$, where the product of complex numbers is replaced by the matrix product

$$f(\mathbf{Z}) = \sum_{k=0}^{\infty} \frac{f^{(k)}}{k!} (\mathbf{Z} - z_0 \mathbf{I})^k \quad (11)$$

where \mathbf{I} is the identity matrix. If the Taylor series converges everywhere, we can expand around $z_0 = 0$ without loss of generality. The coefficients $f^{(k)}$ are independent of the matrix dimension. Under a similarity transformation, an analytic matrix function behaves as

$$f(\mathbf{LZL}^{-1}) = \sum_{k=0}^{\infty} \frac{f^{(k)}}{k!} (\mathbf{LZL}^{-1})^k = \mathbf{L}f(\mathbf{Z})\mathbf{L}^{-1}, \quad (12)$$

since $(\mathbf{LZL}^{-1})^k = \mathbf{LZ}(\mathbf{L}^{-1}\mathbf{L})\mathbf{ZL}^{-1} \dots = \mathbf{LZ}^k\mathbf{L}^{-1}$. The definitions of analytic matrix functions in terms of a Taylor series or the transformation property are equivalent.³⁵ An analytic matrix *functional* $\mathcal{F}[\mathbf{D}(\mathbf{r})] : (L^1)^{N \times N} \rightarrow \mathbb{C}^{N \times N}$ takes a matrix of L^1 -integrable functions, which depend on the 3-dimensional spatial coordinate vector \mathbf{r} , as an argument. Because of the transformation property 2 the series expansion of the matrix density functional $\mathcal{F}[\mathbf{D}(\mathbf{r})]$ is constrained to have the form

$$\mathcal{F}[\mathbf{D}(\mathbf{r})] = \sum_{k=0}^{\infty} \frac{1}{k!} \int d^3r_1 d^3r_2 \dots d^3r_k \quad (13)$$

$$f^{(k)}(\mathbf{r}_1, \dots, \mathbf{r}_k; \mathcal{N}(\mathbf{r}), \text{tr}(\mathbf{D}(\mathbf{r})), \dots) \mathbf{D}(\mathbf{r}_1) \mathbf{D}(\mathbf{r}_2) \dots \mathbf{D}(\mathbf{r}_k)$$

The coefficients $f^{(k)}$ are scalar functions and the matrix density occurs either as a matrix product or in scalar quantities that are invariant under basis transformations, such as the trace $\text{tr}(\mathbf{D}(\mathbf{r}))$. The transformation property does not exclude the possibility that the functional depends implicitly on the dimension of the subspace, which can be expressed as $\mathcal{N}(\mathbf{r}) = \text{tr}\{\mathbf{D}(\mathbf{r})\mathbf{D}^{-1}(\mathbf{r})\}$ and therefore is also a functional of the matrix density.

$\mathcal{F}[\mathbf{D}(\mathbf{r})]$ **does not depend on the subspace dimension.** Although an analytic matrix functional can accept a matrix density of any dimension, it is not immediately obvious that the functional has to be the same independently of the dimension of the subspace. The ensemble density functional theory of Gross, Oliveira and Kohn^{13,15} certainly depends on the weights of the states, so that there is a different functional for an ensemble of 2,3,4,... states. It is conceivable that a similar dependence on the subspace dimension exists for the matrix density functional $\mathcal{F}[\mathbf{D}(\mathbf{r})]$. However, the following considerations (which falls short

of a rigorous proof) suggests that this is not the case and that the same matrix functional applies to all subspace dimensions: Let us suppose that the matrix functional could be different and denote the functional for a subspace with dimension N as $\mathcal{F}^{(N)}$. Let $v(r)$ be the external potential (for instance the Coulomb attraction to the nuclei in a molecule) that is localized around the origin \mathbf{O} . $v(r)$ decays like $\propto -\frac{1}{r}$ or faster with distance. A copy of the same nuclei is put at a distant point \mathbf{R} , so that the total external potential is $v(\mathbf{r}) + v(\mathbf{r} - \mathbf{R})$ as shown in Fig. S1 of the SI.

$\mathbb{V}_{\mathbf{O}}^N$ is the subspace spanned by N bound, electronic wavefunctions (not necessarily eigenstates) that are localized around the origin \mathbf{O} . These wavefunctions decay exponentially with distance.³⁶ $\mathbb{V}_{\mathbf{R}}^N$ is spanned by the same wavefunctions rigidly translated to the center \mathbf{R} . A subspace of twice the dimension is formed by combining the subspaces of the left and right molecules, $\mathbb{V}^{2N} = \mathbb{V}_{\mathbf{O}}^N \cup \mathbb{V}_{\mathbf{R}}^N$. Because of the large distance R , the transition densities between states localized at \mathbf{O} and those at \mathbf{R} vanish. The state densities and transition densities among the states localized on the same molecule are the same for both molecules except for a rigid shift. The matrix density for the combined subspace therefore consists of two blocks of size N , each, $\mathbf{D}^{(2N)}(\mathbf{r}) = \mathbf{D}^{(N)}(\mathbf{r}) \oplus \mathbf{D}^{(N)}(\mathbf{r} - \mathbf{R})$. As the distance between the two molecules is increased ($R \rightarrow \infty$), the matrix elements of the Hamiltonian vanishes, $\langle \Psi_A | \hat{H} | \Psi_B \rangle \rightarrow 0$, if the wavefunctions Ψ_A and Ψ_B are localized on different molecules. The matrix elements of the kinetic energy decrease exponentially, $\langle \Psi_A | \hat{T} | \Psi_B \rangle \propto \exp(-\alpha R)$, for the Coulomb operator the decrease is like $\langle \Psi_A | \hat{W} | \Psi_B \rangle \propto \frac{1}{R}$. Therefore, in the limit $R \rightarrow \infty$, the universal part of the Hamiltonian in the subspace $\mathbb{V}^{(2N)}$ is just block-diagonal, $\hat{H}_0^{(2N)} = \hat{H}^{(N)} \oplus \hat{H}^{(N)}$. The projection of \hat{H}_0 into the combined subspace $\mathbb{V}^{(2N)}$ is a functional of the matrix density $\mathbf{D}^{(2N)}$, and the Hamiltonians in the individual subspaces $\mathbb{V}_{\mathbf{O}}^{(N)}$ and $\mathbb{V}_{\mathbf{R}}^{(N)}$ are functionals of the matrix densities $\mathbf{D}^{(N)}(\mathbf{r})$ and $\mathbf{D}^{(N)}(\mathbf{r} - \mathbf{R})$, respectively. Furthermore the universal functional should be translationally invariant, so that $\mathcal{F}^{(N)}[\mathbf{D}^{(N)}(\mathbf{r} - \mathbf{R})] = \mathcal{F}^{(N)}[\mathbf{D}^{(N)}(\mathbf{r})] = \mathbf{H}_0^{(N)}$. Therefore the following relation holds between the functionals $\mathcal{F}^{(2N)}$ and $\mathcal{F}^{(N)}$:

$$\mathcal{F}^{(2N)} \left[\begin{pmatrix} \mathbf{D}^{(N)}(\mathbf{r}) & 0 \\ 0 & \mathbf{D}^{(N)}(\mathbf{r} - \mathbf{R}) \end{pmatrix} \right] = \begin{pmatrix} \mathbf{H}_0^{(N)} & 0 \\ 0 & \mathbf{H}_0^{(N)} \end{pmatrix} = \begin{pmatrix} \mathcal{F}^{(N)}[\mathbf{D}(\mathbf{r})] & 0 \\ 0 & \mathcal{F}^{(N)}[\mathbf{D}(\mathbf{r})] \end{pmatrix} \quad (14)$$

The plan is to plug the functional series expansion of Eqn. 13 into Eqn. 14 in order to work out the relation between the functional $\mathcal{F}^{(2N)}$ and $\mathcal{F}^{(N)}$. A priori it is assumed that the functional could be different for different subspace dimensions. It will be shown that this is not the case. However, Eqn. 14 is not yet in a suitable form to compare the two functionals, since on its LHS the functional $\mathcal{F}^{(2N)}$ operates on a matrix density of dimension $2N$, while on the RHS the arguments of the functional $\mathcal{F}^{(N)}$ have dimension N . So we have to transform the expression in such a way, that $\mathcal{F}^{(2N)}$ operates on a matrix density of dimension N .

First some preliminaries. The functional series expansion involves integrations over three-dimensional space. The three-dimensional space is partitioned into Voronoi polyhedra. Each point \mathbf{r} is assigned to the closest molecule,

$$V_{\mathbf{O}} = \{\mathbf{r} : |\mathbf{r}| < |\mathbf{r} - \mathbf{R}|\} \quad (15)$$

$$V_{\mathbf{R}} = \{\mathbf{r} : |\mathbf{r} - \mathbf{R}| \leq |\mathbf{r}|\} \quad (16)$$

$$\mathbb{R}^3 = V_{\mathbf{O}} \cup V_{\mathbf{R}}. \quad (17)$$

Integration over space is split into separate integrations over the Voronoi polyhedra,

$$\int_{\mathbb{R}^3} \dots d^3r = \int_{V_{\mathbf{O}}} \dots d^3r + \int_{V_{\mathbf{R}}} \dots d^3r. \quad (18)$$

Since the wavefunctions are localized around the potential at either \mathbf{O} or \mathbf{R} , there is no point in space where $\mathbf{D}(\mathbf{r})$ and $\mathbf{D}(\mathbf{r} - \mathbf{R})$ are simultaneously non-zero. So depending on where the point \mathbf{r} lies, the projection of the density operator in the subspace \mathbb{V}^{2N} has only

one non-zero block at the top or at the bottom:

$$\mathbf{D}^{(2N)}(\mathbf{r}) = \mathbf{D}^{(N)}(\mathbf{r}) \oplus \mathbf{D}^{(N)}(\mathbf{r} - \mathbf{R}) = \begin{cases} \mathbf{D}^{(N)}(\mathbf{r}) \oplus 0 & \text{if } \mathbf{r} \in V_{\mathbf{O}} \\ 0 \oplus \mathbf{D}^{(N)}(\mathbf{r} - \mathbf{R}) & \text{if } \mathbf{r} \in V_{\mathbf{R}} \end{cases} \quad (19)$$

If the functional should depend implicitly on the subspace dimension, the number of electronic states has to be expressed as a functional of the matrix density. One way to do this is

$$\mathcal{N}(\mathbf{r}) = \text{tr} \{ \mathbf{D}(\mathbf{r}) \mathbf{D}^{-1}(\mathbf{r}) \}. \quad (20)$$

If at some point the matrix density does not have full rank and therefore is not invertible, the pseudoinverse is used instead of the inverse. $\mathcal{N}(\mathbf{r})$ always equals the rank of the matrix density. Since $\mathbf{D}^{(2N)}(\mathbf{r})$ has at most rank N (the number of wavefunctions that are simultaneously non-zero is at most N),

$$\mathcal{N}^{(2N)}(\mathbf{r}) = \begin{pmatrix} \mathcal{N}^{(N)}(\mathbf{r}) & \text{if } \mathbf{r} \in V_{\mathbf{O}} \\ \mathcal{N}^{(N)}(\mathbf{r} - \mathbf{R}) & \text{if } \mathbf{r} \in V_{\mathbf{R}}. \end{pmatrix} \quad (21)$$

By using the series expansion 13 for $\mathcal{F}^{(2N)}[\mathbf{D}^{(2N)}(\mathbf{r})]$, splitting the integration volume into the Voronoi polyhedra, using relations 19 and 21 and the translational invariance, it can be shown (see Eqn. SI-21 in the SI for details) that $\mathcal{F}^{(2N)}[\mathbf{D}^{(2N)}(\mathbf{r})] = \mathcal{F}^{(2N)}[\mathbf{D}^{(N)}(\mathbf{r})] \oplus \mathcal{F}^{(2N)}[\mathbf{D}^{(N)}(\mathbf{r})]$. From comparison with Eqn. 14 it then follows that

$$\mathcal{F}^{(2N)}[\mathbf{D}^N(\mathbf{r})] = \mathcal{F}^{(N)}[\mathbf{D}^N(\mathbf{r})]. \quad (22)$$

for any matrix density $\mathbf{D}^{(N)}(\mathbf{r})$ that decays exponentially with distance. Now the dimension of the argument is the same on both sides. The above equality suggests that the same universal functional applies to all subspace dimensions and one can drop the superscript and simply write $\mathcal{F}[\mathbf{D}(\mathbf{r})]$. For $N = 1$ it is equal to the ground-state kinetic and Hartree-

exchange-correlation (Hxc) functionals,

$$\mathcal{F}[\rho(\mathbf{r})] = T[\rho(\mathbf{r})] + E_{\text{Hxc}}[\rho(\mathbf{r})] \quad (23)$$

However, deducing the exact multi-state functional from known ground-state kinetic-energy and Hartree-exchange-correlation functionals is not readily possible. Even if one has a closed expression of a local functionals in term of $\rho(\mathbf{r})$, it is not clear where $\rho(\mathbf{r})$ has to be replaced by the scalar $\text{tr}(\mathbf{D}(\mathbf{r}))$ and where by the matrix $\mathbf{D}(\mathbf{r})$. For $N = 1$, all three are the same, i.e. $\rho(\mathbf{r}) = \text{tr}(\mathbf{D}(\mathbf{r})) = \mathbf{D}(\mathbf{r})$. Nevertheless ground-state functionals can give some guidance.

The next step is to find a multistate matrix density functional, that has the required transformation properties and reduces to a known ground state functional for $N = 1$.

2.2 An Approximate Matrix Functional

The electronic Hamiltonian $\hat{H} = \hat{T} + \hat{W} + \hat{V}$ consists of the kinetic energy $\hat{T} = \sum_{a=1}^n -\frac{1}{2}\nabla_a^2$, the electron-electron repulsion $\hat{W} = \sum_{a<b}^n \frac{1}{|\mathbf{r}_a - \mathbf{r}_b|}$ and the external potential $\hat{V} = \sum_{a=1}^n v(\mathbf{r}_a)$, where $v(\mathbf{r})$ usually is the electron-nuclear attraction. While the projection of the external potential into any subspace is trivially a functional of the matrix density of that subspace,

$$V[\mathbf{D}(\mathbf{r})]_{IJ} = \langle \Psi_I | \hat{V} | \Psi_J \rangle = \int d^3r v(\mathbf{r}) D_{IJ}(\mathbf{r}), \quad (24)$$

the functionals for \hat{T} and \hat{W} have to be approximated. The kinetic energy multistate functional $\mathbf{T}[\mathbf{D}]$ will be derived from the Thomas-Fermi and von-Weizsäcker ground state functionals. The electron-electron repulsion is split into a Hartree-like direct term $\mathbf{J}[\mathbf{D}]$, a Dirac-like exchange term $-\mathbf{K}^{\text{LDA}}[\mathbf{D}]$, a correlation-like term $\mathbf{C}^{\text{LDA}}[\mathbf{D}]$ based on Chachiyo's fit to the correlation energy of the homogeneous electron gas and a self-interaction correction for the core electrons. Each functional of the matrix density behaves properly under basis transformations of the state vectors. Auxiliary basis states are not used to compute any parts of the Hamiltonian. Let us start with the electron-electron repulsion.

2.2.1 Electron Repulsion Functional

The matrix element of the electron repulsion operator, $\sum_{a<b}^n 1/|\mathbf{r}_a - \mathbf{r}_b|$, between two electronic states Ψ_I and Ψ_J with n electrons is

$$W_{IJ} = \int dx_1 \int dx_2 \dots \int dx_n \Psi_I^*(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \sum_{a<b}^n \frac{1}{|\mathbf{r}_a - \mathbf{r}_b|} \Psi_J(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n), \quad (25)$$

where $\mathbf{x}_a = (\mathbf{r}_a, \sigma_a)$ are the spatial and spin coordinates of electron a . Because of the requirement that $W[\mathbf{D}]_{IJ}$ is an analytic matrix functional, the Hartree-like term

$$\mathbf{J}[\mathbf{D}] = \frac{1}{2} \int \int' \frac{\mathbf{D}(\mathbf{r})\mathbf{D}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (26)$$

contains a matrix product that mixes the electron density of a single state, $D_{II}(\mathbf{r})$, with the transition densities $D_{IJ}(\mathbf{r}')$ ($I \neq J$). The electrostatic energy of a single excited state, which has a classical interpretation, cannot be calculated in isolation. It depends on the transition densities, which have no classical analogue. Only if the transition densities to other states in the selected subspace vanish everywhere, can the state be treated in isolation.

The exchange-like term in the local density approximation (LDA)^{38, 39}

$$\mathbf{K}^{\text{LDA}}[\mathbf{D}] = C_x \int (\mathbf{D}(\mathbf{r}))^{4/3} d^3r, \quad C_x = 0.7386 \quad (27)$$

or in the local spin density approximation (LSDA),⁴⁰

$$\mathbf{K}^{\text{LSDA}}[\mathbf{D}^\alpha, \mathbf{D}^\beta] = 2^{1/3} C_x \int (\mathbf{D}^\alpha(\mathbf{r}))^{4/3} + (\mathbf{D}^\beta(\mathbf{r}))^{4/3} d^3r, \quad (28)$$

involves a fractional matrix power of the matrix density, $(\mathbf{D}^\alpha)^{4/3}$. It is important to stress that this is different from applying the function $\bullet^{4/3}$ to each element of the matrix. The best way to actually compute a matrix function such as the fractional power, is to first do a spectral decomposition of the matrix density at each point in space. Let $\mathbf{\Lambda}(\mathbf{r}) =$

$\text{diag}(\lambda_1(\mathbf{r}), \lambda_2(\mathbf{r}), \dots, \lambda_N(\mathbf{r}))$ be the diagonal matrix of eigenvalues of $\mathbf{D}(\mathbf{r})$, which is a function of the point in space \mathbf{r} . The corresponding orthonormalized eigenvectors are the columns of the unitary matrix $\mathbf{U}(\mathbf{r})$ satisfying $\mathbf{U}^\dagger(\mathbf{r})\mathbf{U}(\mathbf{r}) = \mathbf{I}_{N \times N}$, $\forall \mathbf{r}$. Then the eigendecomposition of $\mathbf{D}(\mathbf{r})$ is

$$\mathbf{D}(\mathbf{r}) = \mathbf{U}(\mathbf{r})\mathbf{\Lambda}(\mathbf{r})\mathbf{U}^\dagger(\mathbf{r}) \quad (29)$$

Note that eigenvalues and eigenvectors are functions of space. And the fractional matrix power is

$$\mathbf{D}^{4/3}(\mathbf{r}) = \mathbf{U}(\mathbf{r})\mathbf{\Lambda}^{4/3}(\mathbf{r})\mathbf{U}^\dagger(\mathbf{r}). \quad (30)$$

In general any analytic scalar function $f : \mathbb{C} \mapsto \mathbb{C}$ can be converted into an analytic matrix function $\mathbf{F}[\mathbf{D}] : \mathbb{C}^{N \times N} \mapsto \mathbb{C}^{N \times N}$ by applying f to the eigenvalues of \mathbf{D} . Actually applying the ground state exchange functional element-wise is not even possible, since the transition density $D_{IJ}(\mathbf{r})$ can be negative. For negative arguments the fractional power is not uniquely defined. This problem does not occur for the fractional matrix power, since the matrix density is positive definite, so that all eigenvalues are positive everywhere.

Let us move on to the correlation energy. The Chachiyo functional⁴¹ is a simple and elegant parametrization of the correlation energy per electron. It recovers the high-density limit and fits the quantum Monte-Carlo results of Ceperley and Alder⁴² rather well. In terms of the Wigner-Seitz radius $r_s = (4\pi/3\rho)^{-1/3}$ the correlation energy per particle is given by $\epsilon_{\text{corr}}(\rho) = a \log(1 + br_s^{-1} + br_s^{-2})$. Since ϵ_{corr} is an analytic function of the electron density it can be turned into a multistate functional by retaining the functional form and replacing the scalar density argument $\rho = \rho^\alpha + \rho^\beta$ with the matrix density $\mathbf{D} = \mathbf{D}^\alpha + \mathbf{D}^\beta$. Considering only the paramagnetic part of the correlation energy, the multistate extension of the Chachiyo correlation functional (Eqn. 8 in Ref.⁴¹) can be written in the following form:

$$\mathbf{C}^{\text{LDA}}[\mathbf{D}] = \int a \log(\mathbf{1} + b_1 \mathbf{D}^{1/3}(\mathbf{r}) + b_2 \mathbf{D}^{2/3}(\mathbf{r})) \mathbf{D}(\mathbf{r}) d^3r \quad (31)$$

with the parameters

$$a = \frac{\log(2) - 1}{2\pi^2} = -0.01554534543482745 \quad (32)$$

$$b = 20.4562557 \quad (\text{paramagnetic case}) \quad (33)$$

$$b_1 = (4\pi/3)^{1/3} b = 32.975319597703546 \quad (34)$$

$$b_2 = (4\pi/3)^{2/3} b = 53.155949872619715. \quad (35)$$

The integrand of Eqn. 31, the correlation energy density, is calculated by diagonalizing \mathbf{D} , applying the function $f(D) = a \log(1 + b_1 D^{1/3} + b_2 D^{2/3})D$ to the eigenvalues and transforming the result back from the eigenbasis to the original basis. For a single electronic state Eqn. 31 reduces to the correlation energy of the local density approximation. This functional only evaluates the paramagnetic part of the correlation energy, the spin polarization is assumed to be zero. For closed-shell molecules the diagonal elements of the matrix density are not spin-polarized, however the transition densities usually have a large spin-polarization. Unfortunately, it is not obvious how to turn the spin polarization $\zeta = (\rho^\alpha - \rho^\beta)/\rho$ into a matrix functional. With spin polarization the matrix functional would have to depend on two matrix densities, $\mathbf{D}^\alpha + \mathbf{D}^\beta$ and $\mathbf{D}^\alpha - \mathbf{D}^\beta$, which do not commute. The correlation energy of the uniform electron gas is always negative. For medium and low densities ($0.5 \leq r_s \leq 10.0$), the ratio of ferromagnetic to paramagnetic correlation energy is approximately 1.1, see Fig. S2 in the Supporting Information (SI). So by neglecting the difference between paramagnetic and ferromagnetic correlation we incur an error of approximately 10%.

The functional form of the Hartree term (Eqn. 26) for the off-diagonal elements can be justified qualitatively as follows: Suppose for simplicity that the wavefunctions Ψ_I, Ψ_J, \dots in the electronic subspace are Hartree products built from a set of orthonormal orbitals $\phi_a(\mathbf{r})$. The electrons are assumed to be spinless and distinguishable. All wavefunctions have $n - 1$

orbitals ($\phi_1, \dots, \phi_{n-1}$) in common and differ only in the last orbital (ϕ_I or ϕ_J etc.),

$$\Psi_I = \phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2)\cdots\phi_{n-1}(\mathbf{r}_{n-1})\phi_I(\mathbf{r}) \quad (36)$$

$$\Psi_J = \underbrace{\phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2)\cdots\phi_{n-1}(\mathbf{r}_{n-1})}_{\text{"inactive"}}\phi_J(\mathbf{r}). \quad (37)$$

The diagonal and off-diagonal parts of the matrix density for these states are

$$D_{II}(\mathbf{r}) = \sum_{1 \leq a \leq n-1} |\phi_a(\mathbf{r})|^2 + |\phi_I(\mathbf{r})|^2 \quad (38)$$

and

$$D_{IJ}(\mathbf{r}) = \phi_I^*(\mathbf{r})\phi_J(\mathbf{r}) \quad I \neq J, \quad (39)$$

respectively. The matrix element of the electron repulsion operator between two different states is

$$\begin{aligned} W_{IJ} &= \int d^3r_1 \dots \int d^3r_n \Psi_I^*(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) \sum_{1 \leq a < b \leq n} \frac{1}{|\mathbf{r}_a - \mathbf{r}_b|} \Psi_J(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) \\ &= \int d^3r_1 \dots \int d^3r_n \sum_{1 \leq a < b \leq n} \frac{|\phi_1(\mathbf{r}_1)|^2 \dots |\phi_{n-1}(\mathbf{r}_{n-1})|^2 \phi_I^*(\mathbf{r}_n)\phi_J(\mathbf{r}_n)}{|\mathbf{r}_a - \mathbf{r}_b|} \\ &= \int d^3r_1 \int d^3r_2 \left(\sum_{1 \leq a \leq n-1} |\phi_a(\mathbf{r}_1)|^2 \right) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} D_{IJ}(\mathbf{r}_2) \end{aligned} \quad (40)$$

The last equality follows because all terms that involve an integral over \mathbf{r}_n vanish since ϕ_I and ϕ_J are orthogonal. So W_{IJ} is the electrostatic interaction of the transition density with the density of the "inactive" orbitals that are shared by all electronic states. On the other

hand, the Hartree-term in Eqn. 26 gives

$$\begin{aligned}
 J[\mathbf{D}]_{IJ} &= \frac{1}{2} \int d^3r_1 \int d^3r_2 \sum_K \frac{D_{IK}(\mathbf{r}_1)D_{KJ}(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \\
 &= \int \int \frac{1}{2} (D_{II}(\mathbf{r}_1) + D_{JJ}(\mathbf{r}_1)) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} D_{IJ}(\mathbf{r}_2) + \frac{1}{2} \sum_{K \neq I, J} \int \int D_{IK}(\mathbf{r}_1) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} D_{KJ}(\mathbf{r}_2)
 \end{aligned}
 \tag{41}$$

In going from the first to the second line, the terms where K is equal to I or J were separated out. The first term in Eqn. 41 differs from the exact electron repulsion (Eqn. 40) only by the spurious electrostatic interaction of D_{IJ} with $\frac{1}{2}(|\phi_I|^2 + |\phi_J|^2)$. The additional second term in Eqn. 41 involves the electrostatic interaction between transition densities of I and K . It is much smaller than the first, since densities have magnitudes on the order of n (number of electron), while transition densities are on the order of < 1 . The additional term is needed to ensure that $\mathbf{J}[\mathbf{D}(\mathbf{r})]$ is a matrix functional of \mathbf{D} and transforms as $\mathbf{J}[\mathbf{L}\mathbf{D}(\mathbf{r})\mathbf{L}^{-1}] = \mathbf{L}\mathbf{J}[\mathbf{D}(\mathbf{r})]\mathbf{L}^{-1}$ under a basis transformation of the electronic states.

For $I = J$ the second term in Eqn. 41 is positive, so that $J[\mathbf{D}]_{II}$ is larger than the electrostatic repulsion of the state density D_{II} alone. The trace of the Hartree-like matrix functional is larger than the sum of the Hartree energies of the individual states,

$$\text{tr}(\mathbf{J}[\mathbf{D}]) \geq \sum_{I=1}^N \frac{1}{2} \int \int' \frac{D_{II}(\mathbf{r})D_{II}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}.
 \tag{42}$$

However, in the electron repulsion without correlation, $\mathbf{J}[\mathbf{D}] - \mathbf{K}^{\text{LDA}}[\mathbf{D}]$, this larger direct electron repulsion is partly compensated for by the exchange part of the indirect electron repulsion \mathbf{K}^{LDA} , which on average is larger than the exchange energies of the individual electronic states,

$$\text{tr}(\mathbf{K}^{\text{LDA}}[\mathbf{D}]) = C_x \int d^3r \text{tr}(\mathbf{D}^{4/3}(\mathbf{r})) \geq \sum_{I=1}^N C_x \int d^3r D_{II}(\mathbf{r})^{4/3}.
 \tag{43}$$

This is a result of Klein’s inequality^{43,44} and is proven in section 2.2 of the SI. The correlation energy is left out in the comparison since it is usually an order of magnitude smaller than the exchange energy.

2.2.2 Self-Interaction Error of Core Electrons

Many failures of the local spin density approximation can be attributed to the self-interaction error.⁴⁵ The largest part of the deviation between the exact electron repulsion \mathbf{W} and the MSDFM approximation $\mathbf{J}[\mathbf{D}] - \mathbf{K}^{\text{LDA}}[\mathbf{D}] + \mathbf{C}^{\text{LDA}}[\mathbf{D}]$ does not depend on the geometry. This suggests that it is due to the core electrons, which do not participate in the bonding. It is known that in the core regions the self-interaction errors are particularly large (see section D in Ref.⁴⁶). We have to subtract the unphysical interaction of a core electron with itself. An electron in a 1s spin-up (α) core orbital $\phi_{1s}^\alpha(\mathbf{r})$ contributes to the DFT electron repulsion the spurious interaction of the density $\rho_{1s}^\alpha(\mathbf{r}) = |\phi_{1s}^\alpha(\mathbf{r})|^2$ with itself. There is a Hartree-like contribution,

$$J[\rho_{1s}^\alpha] = \frac{1}{2} \int \int' \frac{\rho_{1s}^\alpha(\mathbf{r})\rho_{1s}^\alpha(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \quad (44)$$

an exchange-like contribution,

$$-K^{\text{LSDA}}[\rho_{1s}^\alpha] = -2^{1/3}C_x \int \rho_{1s}^\alpha(\mathbf{r})^{4/3}, \quad (45)$$

and a correlation-like contribution,

$$C[\rho_{1s}^\alpha] = \int \rho_{1s}^\alpha(\mathbf{r})\epsilon_{\text{corr}}(\rho_{1s}^\alpha(\mathbf{r})), \quad (46)$$

that have to be subtracted from the diagonal elements $W_{II}[\mathbf{D}]$. This completely removes a single core electron’s repulsion with itself. The off-diagonal elements $W_{IJ}[\mathbf{D}]$, $I \neq J$, do not suffer from the self-intersection problem. The self-interaction energies from different elements (except the very light ones like hydrogen, lithium etc.) are added. Since core

orbitals are doubly occupied, a factor of two appears. The self-interaction correction (SIC) is supposed to cancel the self-interaction error (SIE), so there is a minus sign in front:

$$\text{SIC} = - \sum_c^{\text{core orbs.}} 2 \left(\frac{1}{2} \int \int' \frac{\rho_c(\mathbf{r})\rho_c(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} - 2^{1/3} C_x \int \rho_c(\mathbf{r})^{4/3} + \int \rho_c(\mathbf{r}) \epsilon_{\text{corr}}(\rho_c(\mathbf{r})) \right) \quad (47)$$

This self-interaction correction is not a functional of the matrix density. Like its ground state counterpart, the multistate LDA-like functional is too simplistic to cancel the self-interaction, so that a correction outside of the MSDFT formalism has to be employed.

The self-interaction correction is a constant that only depends on the elemental composition of the molecule, but not on the geometry and is the same for all electronic states. So the final expression for the matrix elements of the electron repulsion in the subspace of the lowest few electronic states is

$$W[\mathbf{D}]_{IJ} = J[\mathbf{D}]_{IJ} - K^{\text{LDA}}[\mathbf{D}]_{IJ} + C^{\text{LDA}}[\mathbf{D}]_{IJ} + \text{SIC} \delta_{IJ}. \quad (48)$$

2.3 Kinetic Energy Functional

The kinetic energy of a nearly-homogeneous electron gas ($\rho \approx \text{const}$, $\nabla\rho \approx 0$) is

$$T[\rho] = T_{\text{TF}}[\rho] + \frac{1}{9}T_{\text{vW}}[\rho], \quad (49)$$

with the Thomas-Fermi (TF) functional,

$$T_{\text{TF}}[\rho(\mathbf{r})] = \frac{3}{10}(3\pi^2)^{2/3} \int \rho(\mathbf{r})^{5/3} d^3r, \quad (50)$$

which is exact for the free, non-interacting electron gas, and the von Weizsäcker (vW) correction,

$$T_{\text{vW}}[\rho(\mathbf{r})] = \frac{1}{8} \int \frac{|\nabla\rho(\mathbf{r})|^2}{\rho(\mathbf{r})} d^3r, \quad (51)$$

which yields the exact kinetic energy for a single electron. The factor $\frac{1}{9}$ in the Thomas-Fermi-von-Weizsäcker functional of Eqn. 49 stems from the conventional gradient expansion to second order around the homogeneous electron gas limit.⁴⁷

The generalization of the Thomas-Fermi functional to multiple electronic states is obtained by replacing the density $\rho(\mathbf{r})$ by the matrix density $\mathbf{D}(\mathbf{r})$ and taking the fractional matrix power $\mathbf{D}^{5/3}(\mathbf{r})$,

$$T_{\text{TF}}[\mathbf{D}(\mathbf{r})]_{IJ} = \frac{3}{10}(3\pi^2)^{2/3} \int [\mathbf{D}(\mathbf{r})^{5/3}]_{IJ} d^3r. \quad (52)$$

The von-Weizsäcker functional depends on both the density and its gradient, which turn into non-commuting matrices \mathbf{D} and $\nabla\mathbf{D}$ in the multistate case. The ground state functional of Eqn. 51 does not determine the order of the factors uniquely. If the matrix inverse is placed symmetrically between the two gradients, which form a scalar product, the multistate von-Weizsäcker functional becomes

$$T_{\text{vW}}[\mathbf{D}(\mathbf{r})]_{IJ} = \int d^3r \frac{1}{8} \sum_K \sum_L \nabla D_{IK}(\mathbf{r}) (\mathbf{D}^{-1}(\mathbf{r}))_{KL} \nabla D_{KL}(\mathbf{r}). \quad (53)$$

This is not the only conceivable extension. One could also use the gradient of the matrix square root of \mathbf{D} ,

$$T_{\text{vWsqrt}}[\mathbf{D}(\mathbf{r})]_{IJ} = \frac{1}{8} \int d^3r \sum_K^N \nabla(D^{1/2})_{IK}(\mathbf{r}) \nabla(D^{1/2})_{KJ}(\mathbf{r}). \quad (54)$$

Since $\nabla\mathbf{D}$ and \mathbf{D} do not commute, $\nabla\mathbf{D}^{1/2} \cdot \nabla\mathbf{D}^{1/2} \neq \nabla\mathbf{D}\mathbf{D}^{-1}\nabla\mathbf{D}$, unlike in the single-state case, where $(\nabla\sqrt{\rho})^2 = \nabla\rho^2/\rho$. For the small systems studied, equations 53 and 54 give slightly different results, but it is not clear which one is superior. In particular neither functional is exact for 1-electron systems. A multistate von-Weizsäcker functional that is exact for 1-electron matrix densities is presented in section 3 of the SI.

The kinetic functional that will be used in the Results section is taken as

$$\mathbf{T}[\mathbf{D}(\mathbf{r})] = \mathbf{T}_{\text{TF}}[\mathbf{D}(\mathbf{r})] + \frac{1}{9}\mathbf{T}_{\text{vW}}[\mathbf{D}(\mathbf{r})]. \quad (55)$$

3 Results

3.1 Dissociation Curve of LiF

In Ref.²⁰ lithium-fluoride was chosen as a test case, because it has an avoided crossing between the lowest ionic and covalent $^1\Sigma^+$ states, the position of which depends sensitively on the balance between static and dynamic correlation. The position of the avoided crossing (7.24 Å) can be estimated from the experimental ionization energy of lithium, the electron affinity of fluorine and the polarizabilities of the two atoms.⁴⁸ However, for MRCI or MCSCF the prediction of the crossing is a challenge, since a large number of configurations have to be included.⁴⁸ Lu and Gao showed that MSDFT requires much less configurations than CASSCF since the density functional incorporates most of the dynamic correlation into the diagonal parts of the Hamiltonian.²⁰ The dissociation curve of LiF was also studied with state-interaction pair-density functional theory³⁰ and constrained DFT configuration interaction.⁴⁹

The same system is used to test the LDA-like matrix density functionals described above. Since the implementation does not yet allow to find the density that minimizes the subspace energy, a simplified check is performed: The exact matrix density is calculated from the exact wavefunctions and is fed into the matrix functionals for the electron-repulsion and kinetic energy. The resulting projections of the electron-repulsion $\mathbf{W}[\mathbf{D}]$ (Eqn. 48) and the kinetic energy $\mathbf{T}[\mathbf{D}]$ (Eqn. 55) into the subspace of the lowest eigenstates are then compared with the exact values $\mathbf{W}^{\text{exact}}$ and $\mathbf{T}^{\text{exact}}$ that are evaluated from the exact wavefunctions. The exact matrix elements of the electron-repulsion are obtained by subtracting the kinetic and nuclear-attraction matrices from the diagonal matrix containing the exact adiabatic

eigenenergies:

$$W_{IJ}^{\text{exact}} = E_I^{\text{exact}} \delta_{IJ} - T_{IJ} - \int v(\mathbf{r}) D_{IJ}(\mathbf{r}) \quad (56)$$

The exact wavefunctions are calculated with PySCF⁵⁰ at the SA(2)-CASSCF level with the aug-cc-pVQZ basis set,⁵¹⁻⁵³ exploiting the $C_{\text{inf}v}$ symmetry of the linear molecule and solving only for the lowest two singlet states with $^1\Sigma^+$ symmetry (irrep A_1). The complete active space consists of 6 electrons in 21 orbitals. The initial guesses for 9 a_1 orbitals, 6 e_{1x} orbitals and 6 e_{1y} orbitals are taken from a Hartree-Fock calculation. Although the large active space captures both static and some dynamic correlation, the avoided crossing is underestimated at 6.8 Å. An even larger active space should shift the avoided crossing closer to the experimental value, but the calculation would become too expensive for a desktop computer. The exact position is of little importance. The goal is to validate the matrix functional at the (almost) exact matrix density.

All integrals and partial derivatives of the matrix density are computed numerically on multicenter spherical grids,^{54,55} the Hartree matrix is obtained by solving the Poisson equation.⁵⁶

Figure 1 shows the CASSCF potential energy curves and the matrix densities at the experimental minimum and at the avoided crossing. With the naked eye it is difficult to discern the small differences between the densities in different states. The matrix density is the only input to the matrix functional. The electron-repulsion and kinetic part of the effective Hamiltonian predicted by the LDA matrix functional are shown in Fig. 2 for different bond lengths. Figures 2a) and b) suggest that both the diagonal and off-diagonal matrix elements of the electron-repulsion operator are perfectly reproduced. After subtracting the large Hartree term, the exchange-correlation matrix still agrees well with the exact one (Fig. 2c and d). Table 1 shows the individual components of the functionals and the exact values at the avoided crossing. The exchange part accounts for most of the exchange-correlation and is approximately a factor of 10 larger than the correlation energy. This statement applies both for the diagonal and the off-diagonal elements.

The relative errors of the diagonal elements and the off-diagonal elements of the electron repulsion are less than 1% and less than 5%, respectively. It is remarkable that the off-diagonal elements W_{IJ} , which have no equivalent in single-state density functional theory, also match well.

Diagonal and off-diagonal matrix elements are treated on the same footing. The off-diagonal elements emerge naturally when turning the scalar LDA functional into a matrix functional by keeping its functional form (same coefficients in Taylor expansion) and replacing the scalar density $\rho(\mathbf{r})$ with the matrix density $\mathbf{D}(\mathbf{r})$. Diagonal and off-diagonal elements are not independent of each other. A unitary rotation in the subspace of the lowest eigenstates would mix them. Because the matrix density is represented in the adiabatic basis, the Hamiltonian $\mathbf{H}[\mathbf{D}(\mathbf{r})]$ is diagonal. If instead the adiabatic eigenstates $1^1\Sigma^+$ and $2^1\Sigma^+$ were to be mixed to form diabatic states, which have ionic (Li^+F^-) or covalent ($\text{Li}\cdot\text{F}\cdot$) character, the matrix functional for the Hamiltonian $\mathbf{H}[\mathbf{D}(\mathbf{r})]$ would also give the correct diabatic state energies and diabatic coupling H_{01} (to the extent possible for a simple LDA functional). By construction, the functionals of the matrix density transform like the Hamiltonian under rotations within the subspace.

The kinetic energy functional on the other hand disappoints (Figs. 2e and f). While the curve of the off-diagonal element $T_{0,1}$ follows the exact matrix element at least qualitatively, the diagonal elements are wrong by more than 1 Hartree. More importantly, the kinetic energy difference between the two states is overestimated by a factor of two. These failures are expected in view of the known difficulties in designing kinetic energy functionals for orbital-free DFT.⁵⁷

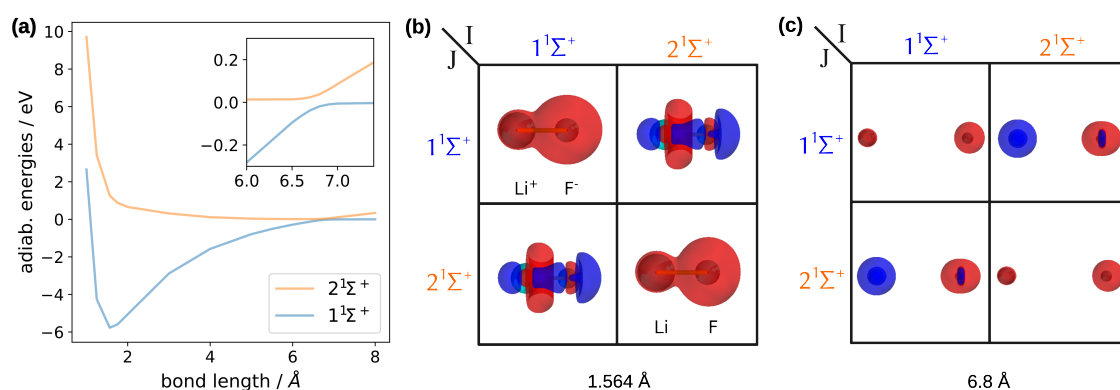


Figure 1: **Lithium-fluoride.** (a) SA-2-CASSCF(6,21)/aug-cc-pVQZ potential energy curves for the lowest 2 singlet states with Σ^+ symmetry, inset: zoom around avoided crossing. (b,c) Visual representation of the matrix density $D_{IJ}(\mathbf{r}) = D_{IJ}^\alpha(\mathbf{r}) + D_{IJ}^\beta(\mathbf{r})$ of lithium-fluoride at the experimental bond length of 1.564 Å (b) and at the avoided curve crossing (c). Diagonal blocks show the electronic state densities for the states $1^1\Sigma^+$ and $2^1\Sigma^+$, off-diagonal blocks the transition density. Isovalues of 0.025 and ≈ 0.001 were used for the state densities and the transition densities, respectively.

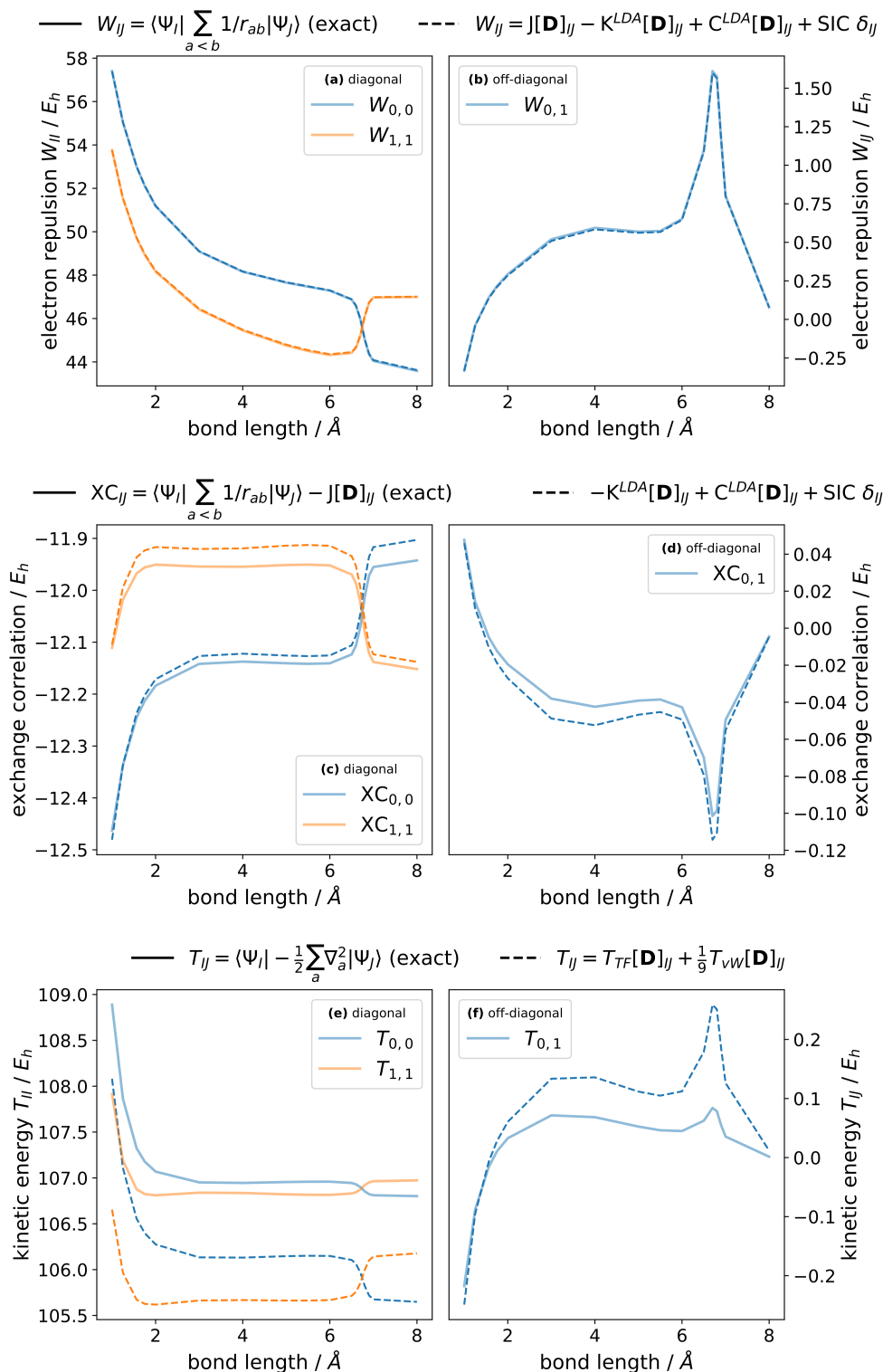


Figure 2: **Lithium-fluoride.** Comparison between exact WFT (solid lines) and approximate MSDFT matrix elements (dashed lines) of the total electron-electron repulsion (a,b), the exchange-correlation (c,d) and the kinetic energy (e,f) in the basis of the adiabatic eigenstates $1^1\Sigma^+$ ($I = 0$) and $2^1\Sigma^+$ ($I = 1$). Left column: diagonal elements, Right column: off-diagonal element.

Table 1: Electron repulsion and kinetic energy matrices in the basis of the exact eigenstates $1^1\Sigma^+$ ($I=0$) and $2^1\Sigma^+$ ($I=1$) of LiF at the avoided crossing (6.8 Å). (a) Exact electron repulsion matrix W_{IJ}^{exact} , multistate DFT approximation $W[\mathbf{D}]_{IJ} = J[\mathbf{D}]_{IJ} - K^{\text{LDA}}[\mathbf{D}]_{IJ} + C^{\text{LDA}}[\mathbf{D}]_{IJ} + \text{SIC}\delta_{IJ}$, Hartree ($J[\mathbf{D}]_{IJ}$), exchange ($-K^{\text{LDA}}[\mathbf{D}]_{IJ}$) and correlation ($C^{\text{LDA}}[\mathbf{D}]_{IJ}$) matrices and self-interaction correction for core electrons ($\text{SIC}\delta_{IJ}$). (b) Exact kinetic energy matrix T_{IJ}^{exact} and multistate DFT approximation $T[\mathbf{D}]_{IJ} = T_{\text{TF}}[\mathbf{D}]_{IJ} + \frac{1}{9}T_{\text{vW}}[\mathbf{D}]_{IJ}$, Thomas-Fermi ($T_{\text{TF}}[\mathbf{D}]_{IJ}$) and von-Weizsäcker ($T_{\text{vW}}[\mathbf{D}]_{IJ}$) kinetic energy matrices. All energies are in Hartree.

(a) electron repulsion, matrix elements (I, J)			
	(0,0)	(1,1)	(0,1)
W_{IJ}^{exact}	45.0339	46.0865	1.5754
$W[\mathbf{D}]_{IJ}$	45.0655	46.1082	1.5633
$J[\mathbf{D}]_{IJ}$	57.0469	58.1669	1.6740
$-K^{\text{LDA}}[\mathbf{D}]_{IJ}$	-10.6125	-10.6832	-0.1014
$C^{\text{LDA}}[\mathbf{D}]_{IJ}$	-0.7924	-0.7989	-0.0092
$\text{SIC}\delta_{IJ}$	-0.5766	-0.5766	0

(b) kinetic energy, matrix elements (I, J)			
	(0,0)	(1,1)	(0,1)
T_{IJ}^{exact}	106.8541	106.9205	0.0785
$T[\mathbf{D}]_{IJ}$	105.8242	105.9969	0.2509
$T_{\text{TF}}[\mathbf{D}]_{IJ}$	96.8843	97.0829	0.2829
$T_{\text{vW}}[\mathbf{D}]_{IJ}$	80.4596	80.2251	-0.2874

4 Conclusion

The effective Hamiltonian in the subspace of the lowest electronic states is an analytic matrix functional of the matrix density and does not depend on the number of electronic states. In contrast to the Δ SCF method, where a ground state functional is applied to the density of a single excited state configuration, the multistate functional yields both the energies of the configurations and the interactions between them.

By construction the average energy of the lowest states is invariant under basis transformations within the subspace. When turning ground state functionals into multistate functionals, ambiguities arise because scalar descriptors of the density become non-commuting matrices. As illustrated for the kinetic energy (section 3 of the SI), exact conditions of the unknown functional can help to constrain its functional form.

A simple LDA-like multistate functional was tested on the dissociation curve of LiF. Both diagonal and off-diagonal matrix elements of the electron-electron repulsion operator are reproduced well when the exact matrix density is used as input. The exchange-correlation part of the state interaction is dominated by the exchange term, while correlation is an order of magnitude smaller. The tested functional is not yet suitable for reliable calculations, but mostly so because the errors of the diagonal kinetic energy are so huge, while the errors of the off-diagonal elements are acceptable.

A limitation of this work is that the matrix density was not determined by minimizing the subspace energy. Using an efficient parameterization of the matrix density^{58, 21} multi-configurational self-consistent field calculations with an effective Hamiltonian computed from the multistate functional should be possible. This is left for future work.

If a good approximation to the universal multistate functional were known and it were somehow possible to restrict the minimization to valid Fermionic matrix densities, the multistate DFT of Lu and Gao would open the way for studying neutral excitations in very large systems in the spirit of orbital-free DFT.

Acknowledgement

There is no funding to declare.

Supporting Information Available

properties of matrix density functionals, comparison of paramagnetic and ferromagnetic correlation energy, proof of Klein's trace inequality, incorporation of exact constraints into von-Weizsäcker functional, Lieb-Oxford-like lower bounds for kinetic and electron repulsion energy of subspace.

Code and Data Availability: All calculations can be repeated on a desktop computer using the python scripts in the github repository https://github.com/humeniuka/multistate_density_functionals-v1/. The program package and data are archived at https://zenodo.org/TODO_ADD_URL.

References

- (1) Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Physical review* **1964**, *136*, B864.
- (2) Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical review* **1965**, *140*, A1133.
- (3) Casida, M. E. *Recent Advances In Density Functional Methods: (Part I)*; World Scientific, 1995; pp 155–192.
- (4) Runge, E.; Gross, E. K. Density-functional theory for time-dependent systems. *Phys. Rev. Lett.* **1984**, *52*, 997.
- (5) Imamura, Y.; Nakai, H. Time-dependent density functional theory (TDDFT) calculations for core-excited states: Assessment of an exchange functional combining the

- Becke88 and van Leeuwen–Baerends-type functionals. *Chemical Physics Letters* **2006**, *419*, 297–303.
- (6) Janesko, B. G. Projected Hybrid Density Functionals: Method and Application to Core Electron Ionization. *Journal of Chemical Theory and Computation* **2023**, *19*, 837–847.
- (7) Levine, B. G.; Ko, C.; Quenneville, J.; Martínez, T. J. Conical intersections and double excitations in time-dependent density functional theory. *Molecular Physics* **2006**, *104*, 1039–1051.
- (8) Dreuw, A.; Weisman, J. L.; Head-Gordon, M. Long-range charge-transfer excited states in time-dependent density functional theory require non-local exchange. *The Journal of chemical physics* **2003**, *119*, 2943–2946.
- (9) Stein, T.; Kronik, L.; Baer, R. Reliable prediction of charge transfer excitations in molecular complexes using time-dependent density functional theory. *Journal of the American Chemical Society* **2009**, *131*, 2818–2820.
- (10) Tawada, Y.; Tsuneda, T.; Yanagisawa, S.; Yanai, T.; Hirao, K. A long-range-corrected time-dependent density functional theory. *The Journal of chemical physics* **2004**, *120*, 8425–8433.
- (11) Shao, Y.; Head-Gordon, M.; Krylov, A. I. The spin–flip approach within time-dependent density functional theory: Theory and applications to diradicals. *The Journal of chemical physics* **2003**, *118*, 4807–4818.
- (12) Theophilou, A. K. The energy density functional formalism for excited states. *Journal of Physics C: Solid State Physics* **1979**, *12*, 5419.
- (13) Gross, E. K.; Oliveira, L. N.; Kohn, W. Rayleigh–Ritz variational principle for ensembles of fractionally occupied states. *Physical Review A* **1988**, *37*, 2805.

- (14) Levy, M.; Nagy, A. Variational density-functional theory for an individual excited state. *Phys. Rev. Lett.* **1999**, *83*, 4361.
- (15) Gross, E. K.; Oliveira, L. N.; Kohn, W. Density-functional theory for ensembles of fractionally occupied states. I. Basic formalism. *Physical Review A* **1988**, *37*, 2809.
- (16) Görling, A. Density-functional theory for excited states. *Physical Review A* **1996**, *54*, 3912.
- (17) Ayers, P. W.; Levy, M.; Nagy, A. Time-independent density-functional theory for excited states of Coulomb systems. *Physical Review A* **2012**, *85*, 042518.
- (18) Pastorzak, E.; Pernal, K. Ensemble density variational methods with self-and ghost-interaction-corrected functionals. *The Journal of chemical physics* **2014**, *140*.
- (19) Gao, J.; Grofe, A.; Ren, H.; Bao, P. Beyond Kohn–Sham approximation: hybrid multistate wave function and density functional theory. *J. Phys. Chem. Lett.* **2016**, *7*, 5143–5149.
- (20) Lu, Y.; Gao, J. Multistate Density Functional Theory of Excited States. *J. Phys. Chem. Lett.* **2022**, *13*, 7762–7769.
- (21) Lu, Y.; Gao, J. Fundamental Variable and Density Representation in Multistate DFT for Excited States. *Journal of Chemical Theory and Computation* **2022**, *18*, 7403–7411.
- (22) Leininger, T.; Stoll, H.; Werner, H.-J.; Savin, A. Combining long-range configuration interaction with short-range density functionals. *Chemical physics letters* **1997**, *275*, 151–160.
- (23) Grofe, A.; Chen, X.; Liu, W.; Gao, J. Spin-multiplet components and energy splittings by multistate density functional theory. *The journal of physical chemistry letters* **2017**, *8*, 4838–4845.

- (24) Ref. 23, Eqn. (5).
- (25) Ref. 19, Eqn. (6).
- (26) Ziegler, T.; Rauk, A.; Baerends, E. J. On the calculation of multiplet energies by the Hartree-Fock-Slater method. *Theoretica chimica acta* **1977**, *43*, 261–271.
- (27) Zhu, H.; Zhao, R.; Liu, M.; Zhang, J.; Gao, J. Leveling the Mountain Range of Excited-State Benchmarking through Multistate Density Functional Theory. **2023**,
- (28) Li Manni, G.; Carlson, R. K.; Luo, S.; Ma, D.; Olsen, J.; Truhlar, D. G.; Gagliardi, L. Multiconfiguration pair-density functional theory. *Journal of chemical theory and computation* **2014**, *10*, 3669–3680.
- (29) Scott, M.; Rodrigues, G. L.; Li, X.; Delcey, M. G. Variational Pair-Density Functional Theory: Dealing with Strong Correlation at the Protein Scale. *Journal of Chemical Theory and Computation* **2024**,
- (30) Sand, A. M.; Hoyer, C. E.; Truhlar, D. G.; Gagliardi, L. State-interaction pair-density functional theory. *J. Chem. Phys.* **2018**, *149*.
- (31) Bao, J. J.; Zhou, C.; Varga, Z.; Kanchanakungwankul, S.; Gagliardi, L.; Truhlar, D. G. Multi-state pair-density functional theory. *Faraday discussions* **2020**, *224*, 348–372.
- (32) Lu, Y.; Gao, J. Multi-State Density Functional Theory for Ground and Excited States. *ChemRxiv* **2021**, DOI: 10.26434/chemrxiv-2021-h2bmc.
- (33) Ref. 20, theorem 1.
- (34) Ref. 20, (3).
- (35) Rinehart, R. The equivalence of definitions of a matrix function. *The American Mathematical Monthly* **1955**, *62*, 395–414.

- (36) Katriel, J.; Davidson, E. R. Asymptotic behavior of atomic and molecular wave functions. *Proceedings of the National Academy of Sciences* **1980**, *77*, 4403–4406.
- (37) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press, 1989.
- (38) Ref. 37, 6.1.20.
- (39) Dirac, P. A. Note on exchange phenomena in the Thomas atom. *Mathematical proceedings of the Cambridge philosophical society*. 1930; pp 376–385.
- (40) Ref. 37, 8.2.16.
- (41) Chachiyo, T. Communication: Simple and accurate uniform electron gas correlation energy for the full range of densities. *J. Chem. Phys.* **2016**, *145*.
- (42) Ceperley, D. M.; Alder, B. J. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.* **1980**, *45*, 566.
- (43) Trace inequality, Trace inequality — Wikipedia, The Free Encyclopedia. 2023; https://en.wikipedia.org/wiki/Trace_inequality, [Online; accessed 26-December-2023].
- (44) Ruelle, D. *Statistical mechanics: Rigorous results*; World Scientific, 1969; proposition 2.5.2 "Convexity Inequality".
- (45) Perdew, J. P.; Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **1981**, *23*, 5048.
- (46) Tsuneda, T.; Hirao, K. Self-interaction corrections in density functional theory. *The Journal of chemical physics* **2014**, *140*.
- (47) Ref. 37, section 6.7.
- (48) Werner, H.-J.; Meyer, W. MCSCF study of the avoided curve crossing of the two lowest $1\Sigma^+$ states of LiF. *The Journal of Chemical Physics* **1981**, *74*, 5802–5807.

- (49) Wu, Q.; Cheng, C.-L.; Van Voorhis, T. Configuration interaction based on constrained density functional theory: A multireference method. *The Journal of chemical physics* **2007**, *127*.
- (50) Sun, Q.; Zhang, X.; Banerjee, S.; Bao, P.; Barbry, M.; Blunt, N. S.; Bogdanov, N. A.; Booth, G. H.; Chen, J.; Cui, Z.-H. Recent developments in the PySCF program package. *The Journal of chemical physics* **2020**, *153*, 024109.
- (51) Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of chemical physics* **1989**, *90*, 1007–1023.
- (52) Kendall, R. A.; Dunning Jr, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *The Journal of chemical physics* **1992**, *96*, 6796–6806.
- (53) Pritchard, B. P.; Altarawy, D.; Didier, B.; Gibson, T. D.; Windus, T. L. New basis set exchange: An open, up-to-date resource for the molecular sciences community. *Journal of chemical information and modeling* **2019**, *59*, 4814–4820.
- (54) Becke, A. D. A multicenter numerical integration scheme for polyatomic molecules. *The Journal of chemical physics* **1988**, *88*, 2547–2553.
- (55) Lebedev, V. I. Quadratures on a sphere. *USSR Computational Mathematics and Mathematical Physics* **1976**, *16*, 10–24.
- (56) Becke, A. D.; Dickson, R. M. Numerical solution of Poisson's equation in polyatomic molecules. *The Journal of chemical physics* **1988**, *89*, 2993–2997.
- (57) Teale, A. M.; Helgaker, T.; Savin, A.; Adamo, C.; Aradi, B.; Arbuznikov, A. V.; Ayers, P. W.; Baerends, E. J.; Barone, V.; Calaminici, P. et al. DFT exchange: sharing

perspectives on the workhorse of quantum chemistry and materials science. *Phys. Chem. Chem. Phys.* **2022**, *24*, 28700–28781.

(58) Ref. 20, theorem (3).