

# Correlated quantum chemistry of extended systems

Towards generalized effective fragment potentials

Bartosz Błasiak

8th September 2016

## Abstract

I propose a generalization of the effective fragment potential (EFP) method originally introduced by Day et al. [1] and further developed by Gordon’s and Krylov’s groups. Original EFP method is applicable only to ground state systems at the Hartree–Fock theory or, with recent developments, at the Kohn–Sham density functional theory. Here proposed generalized effective fragment potential (GEFP) approach is derived from the charge density-based formulation of the interaction energy calculation of Mandado and Hermida-Ramon, [2] which makes it applicable to interactions between wave functions of an arbitrary electronic state described by an arbitrary (post-)Hartree–Fock methodology. Thus GEFP becomes essentially free from multipole expansion and perturbation theory. I achieved significant simplification of equations through introducing the concept of a one-electron effective potential operator that can be expanded in a known auxiliary basis set. Thus, the energy-dependent observable can be generally expressed as

$$P = \sum_{ij} \sum_p f_p[\langle i | \hat{\mathcal{O}}_p(1) | j \rangle] , \quad (1)$$

where  $\hat{\mathcal{O}}_p(1)$  is a one-electron operator,  $f$  is the functional representing environment and  $i, j$  refer to atomic or molecular orbitals. Pauli deformation effects are approximated up to second-order in overlap matrix whereas polarization-induced charge deformation is predicted from the density matrix polarization susceptibility tensors and the distribution of the external electric field. Pure exchange is modeled by invoking the spherical Gaussian overlap approximation for localized orthogonalized sets of molecular orbitals. The resulting model is completely free from electron repulsion integrals (ERI’s), while rigorously taking into account electrostatic, induction, exchange-repulsion, exchange-induction and exchange-correlation effects. Moreover, the computational cost of GEFP is predicted to be only slightly larger than EFP2 approach. Owing to those features, I anticipate that the GEFP approach could be a significant improvement over the EFP and could meet various applications in the fields of (i) dynamics of extended systems, (ii) simulating the resonance energy transfer, (iii) (photo)chemical reactions in condensed phases, (iv) modeling the non-linear optical properties of crystals as well as (v) the electronic, vibrational and magnetic solvatochromism.

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Fragment-Based Approach . . . . .	4
1.2	Applications of GEFP Beyond Interaction Energies . . . . .	5
1.3	The Concept of Effective One-Electron Potentials . . . . .	5
1.4	Effective One-Electron Potentials: Examples . . . . .	6
1.4.1	Repulsive Interactions . . . . .	6
1.4.2	EET Coupling Constants . . . . .	8
1.5	Limitations of Effective One-Electron Potentials . . . . .	11
<b>2</b>	<b>Energetic Properties of Molecules</b>	<b>11</b>
2.1	Full Intermolecular Interaction Energy Expression . . . . .	11
2.2	Hartree–Fock Intermolecular Interaction Energy Based on GEFP . . . . .	12
2.2.1	Electrostatics . . . . .	13
2.2.2	Repulsion . . . . .	14
2.2.3	Polarization . . . . .	18
2.2.4	Exchange . . . . .	25
2.2.5	Total Hartree–Fock Interaction Energy Expression . . . . .	26
2.3	Post-Hartree–Fock Intermolecular Interaction Energy Based on GEFP . . . . .	27
2.3.1	Pauli Deformation Density . . . . .	27
2.3.2	Polarization Effects . . . . .	28
2.3.3	Exchange Effects . . . . .	29
2.3.4	Dispersion Effects . . . . .	29
2.3.5	Summary . . . . .	29
<b>3</b>	<b>Electronic Solvatochromism</b>	<b>30</b>
<b>4</b>	<b>Polarization Density Model: Test Results</b>	<b>30</b>
<b>5</b>	<b>Planned Course of Research</b>	<b>38</b>
	<b>References</b>	<b>40</b>

# 1 Introduction

Modern times brought us to the turning point when accurate molecular-level description of extended systems of biological and technological importance becomes rather unavoidable. One might say that the milestone in the development of new intelligent drugs and materials is deeply rooted in understanding the interplay between not a few but hundreds or even thousands of molecules.

Recently, significant portion of quantum chemistry community is focused on the development of highly accurate schemes of solving the Schrödinger equation as accurately as possible. However, understanding the chemical reactions requires studying families of molecular (ground and electronically excited) states in a deep and wide sea of conformational manifolds that span millions of structural varieties. This is mostly because chemical reactions are not simple chains of end-to-end molecular events that finally occur to produce products. They are rather an enormous coupled collections of such events, each having its own probability of occurrence that depends on other events and surrounding molecules.

Exploring fully the dynamics of molecular systems is the intuitive way of studying the nature at genuine molecular level. Dealing with all the relevant molecular states and event scenarios requires prohibitively large amount of computational effort if one wants to use current state-of-the-art highly accurate approaches based on coupled cluster or similar methods, even including quickly growing family of cost-reducing methodologies such as resolution of identity (RI), density fitting (DF), chain of spheres (COS) algorithm and many others. Therefore, it seems reasonable to adopt a well known hybrid approach in modeling extended systems with chemical rearrangements, describing the further ‘environment’ in a simple but still rigorous manner.

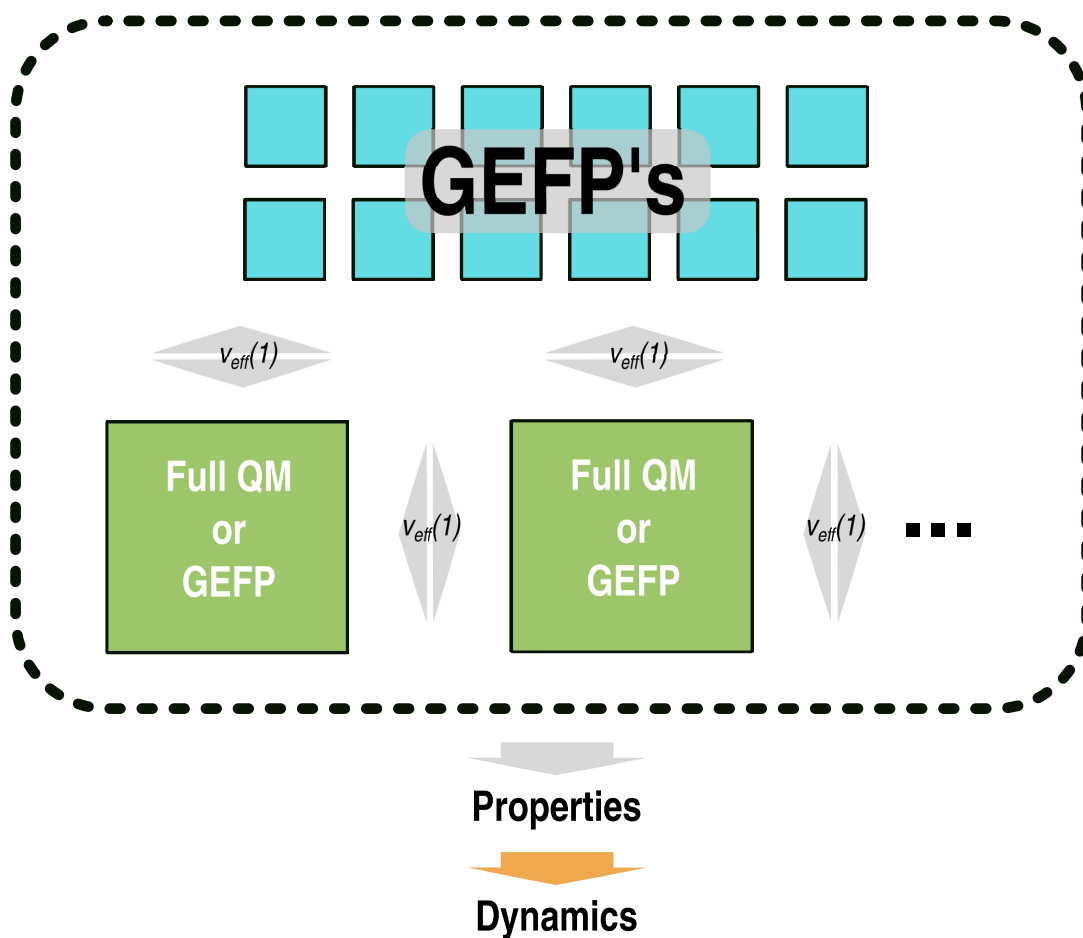
In this Document, we attempt to design the new methodology for describing the molecular environment or simple extended system as a whole by making use of the fragment-based strategy. The proposed model, *the generalized effective fragment potential method* or GEFP, inspired by the second generation of the effective fragment potential (EFP2) method, is to go beyond the EFP2 that is solely based on single-determinantal wave function such as Hartree-Fock (HF) or density functional theory (DFT). In principle, our GEFP method is targeted towards any kind of correlated determinantal wave function bringing the potential applications beyond the ground state chemistries. The key features of the proposed GEFP method are as follows:

- The necessary approximations are limited to the absolute minimum that enables partitioning of the system into independent fragments but without using the perturbation theory.
- We make use of the electronic charge density deformations recently elaborated by Mandado and Hermida-Ramon [2] in their density-based interaction energy decomposition scheme. Owing to this, we allow electrons to move (as for the present stage, only within a fragment but we hope it can be extended to include also intermolecular charge flows)
- We generalize the technique of defining the *one-electron potentials* in quantum chemistry of aggregates. We show that in many cases it is possible to define the *effective* one-electron potentials with making little or no approximations to the original equations. Then, the complicated and lengthy summations over the electron repulsion integrals (ERI’s) are drastically reduced to a few summations over the one-electron integrals (OEI’s).

Once the above features are adequately combined, it could be possible to construct an efficient and accurate model of an extended system composed of a set of relatively small, rigid molecules. The price we need to pay is the need of a separate theories or models for distinct molecular properties. However, once these theories are developed, tested and implemented in a computer software, it becomes possible to use them in a black-box manner much like EFP2, and study extended molecular systems that are either out-of-reach for state-of-the-art methods, or require colossal networks of parallelized computer infrastructures, not available for most of scientists.

## 1.1 Fragment-Based Approach

Our strategy can be depicted as done in Figure 1. It is clear that the strategy is no different from the conventional fragment-based calculation approach. The only difference lies in the mathematical formulation of the generalized one-electron potentials,  $v_{\text{eff}}(1)$ . If designed appropriately, such calculations could be fast, easy to paralelize and scale almost linearly with the size of the system. The bottleneck part of the calculations would be reduced to the effort associated with the full QM calculations (if any) of a fragment which should remain structurally and electronically flexible. Note that GEFP's are intrinsically structurally rigid entities by construct. The key



**Figure 1:** Scheme of calculations utilizing GEFP's as fragments. Full QM calculations can be performed on a small fragment whereas structurally rigid GEFP's interact with each other and Full QM fragments via the generalized one-electron potentials describing electrostatic and exchange effects.

question remains regarding the overall accuracy of the GEFP approach. It is believed that, in the limit of the strength of the conventional intermolecular interactions, including electrostatic

interactions and first-order exchange-repulsion, when combined with the molecular properties at highly correlated level, our proposed model could be accurate and reliable.

## 1.2 Applications of GEFP Beyond Interaction Energies

Definition of the effective one-electron potentials is one of central aspects of our approach and is not limited to strictly energetic quantities. Therefore, I believe that this can be used not only to obtain efficient working models for intermolecular exchange-repulsion forces but also other specific molecular properties. In this proposal I shall in addition

- provide an example of constructing a special model for a given property of a system such as coupling between electronic transitions, which short-range component based just on the construction of one-electron potentials. Coupling of electronic transitions is important for understanding the resonance energy transfer (RET) which is central to the functioning of light-harvesting complexes.
- consider the couplings between two potential energy surfaces (PES) in terms of the effective potentials. Such couplings are very important for non-adiabatic molecular dynamics that needs to be considered when analyzing the time-evolution of nearly degenerate electronic states.

I believe that GEFP scheme can be also applicable to other challenges of computational chemistry such as the magnetic solvatochromism, interaction-induced nonlinear optical properties and so on. In particular, I believe that GEFP would be further stage in the development of the vibrational solvatochromism model based on conventional EFP2 – SolEFP method [3, 4] that have been developed for the recent 4 years. [3, 5, 6]

## 1.3 The Concept of Effective One-Electron Potentials

In this Section, I describe the motivation for introducing the effective one-electron potentials into quantum chemistry of large, extended systems. In brief, the main principle is to rewrite the arbitrary function  $f$  as\*

$$f \left[ \left( \phi_i^A \phi_j^A | \phi_k^B \phi_l^B \right) \right] = \left( \phi_i^A | v_{kl}^B | \phi_j^A \right) \rightarrow \text{multipole expansion or density fitting} \quad (2a)$$

$$f \left[ \left( \phi_i^A \phi_j^B | \phi_k^B \phi_l^B \right) \right] = \left( \phi_i^A | v_{kl}^B | \phi_j^B \right) \rightarrow \text{density fitting} \quad (2b)$$

where  $A$  and  $B$  denote different molecules and  $\phi_i$  is the  $i$ th molecular orbital or basis function.  $v_{kl}^B$  is the effective one-electron potential derived from the partial effective densities  $\rho_{kl}^B(\mathbf{r}) = \phi_k^B(\mathbf{r})\phi_l^B(\mathbf{r})$ . The summations over  $k$  and  $l$  can be incorporated into the generalized one-electron potential  $v_{\text{eff}}^B$  to produce

$$\sum_{ij} \sum_{kl \in B} f \left[ \left( \phi_i \phi_j | \phi_k^B \phi_l^B \right) \right] = \sum_{ij} \left( \phi_i | v_{\text{eff}}^B | \phi_j \right) \quad (3)$$

Thus, the total computational effort is extremely reduced from the fourth-fold sum involving evaluation of ERIs to the two-fold sums of cheaper one-electron integrals. It is also possible to generalize the above expression even further by summing over all possible functions  $f_t$

$$\sum_t \sum_{ij} \sum_{kl \in B} f_t \left[ \left( \phi_i \phi_j | \phi_k^B \phi_l^B \right) \right] = \sum_{ij} \left( \phi_i | v_{\text{eff}}^B | \phi_j \right) \quad (4)$$

---

\* I use Coulomb notation for ERIs, i.e.,  $(\phi_i \phi_j | \phi_k \phi_l) \equiv \iint d\mathbf{r}_1 d\mathbf{r}_2 \phi_i^*(\mathbf{r}_1) \phi_j(\mathbf{r}_1) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \phi_k^*(\mathbf{r}_2) \phi_l(\mathbf{r}_2)$

The above design has the advantage that the effective potentials are fully first-principles and no extensive case-dependent fitting procedures are necessary. Only the effective potentials of *independent* fragment  $B$  need to be determined once and for all and stored in a file. Note also, that, in principle, there is no approximation made here at that moment.

## 1.4 Effective One-Electron Potentials: Examples

Here I provide some illustrative examples of using the above rules. I discuss i) simple model of the Pauli repulsion energy in first order with respect to the overlap density matrix, and ii) evaluation of excitation energy transfer coupling that arises through the electron-hole transfer mechanism.

### 1.4.1 Repulsive Interactions

I shall describe the potential use of the effective one-electron potentials in the evaluation of the Pauli repulsion interaction energy between two molecules based on their unperturbed Hartree-Fock (HF) wave functions.

Repulsion energy was shown by Murrell et al. to be a power series of overlap matrix elements between MO's. [7, 8] Let us analyze the contribution which is first-order in the overlap matrix  $S$ ,

$$U_{\text{rep}}(\mathcal{O}(S)) = -2 \sum_{i \in A} \sum_{j \in B} S_{ij} \left\{ V_{ij}^A + V_{ij}^B - (\phi_i^A \phi_j^B | \phi_j^B \phi_j^B) - (\phi_i^A \phi_i^A | \phi_i^A \phi_j^B) \right. \\ \left. + 2 \sum_{k \in A} (\phi_i^A \phi_j^B | \phi_k^A \phi_k^A) + 2 \sum_{l \in B} (\phi_i^A \phi_j^B | \phi_l^B \phi_l^B) \right\} \quad (5)$$

This term is difficult to compute due to many nested summations and expensive ERI evaluation. I will significantly simplify the above formula by using one-electron effective potentials expanded in auxiliary bases.

The one-electron parts are assumed to be easy to evaluate and I focus our attention on ERIs. First, I can expand in AO basis as

$$(\phi_i^A \phi_j^B | \phi_j^B \phi_j^B) = \sum_{\mu \in A} \sum_{\sigma \in B} \sum_{\lambda \delta \in B} (C_{\mu i}^A)^* C_{\sigma j}^B (C_{\lambda j}^B)^* C_{\delta j}^B (\mu \sigma | \lambda \delta) \quad (6)$$

$$= \sum_{\mu \in A} \sum_{\sigma \in B} (C_{\mu i}^A)^* C_{\sigma j}^B (\mu | v_{\text{eff},1}^{B[j]} | \sigma) = (\phi_i^A | v_{\text{eff},1}^{B[j]} | \phi_j^B) \quad (7)$$

where I defined the *effective potential*

$$v_{\text{eff},1}^{B[j]}(\mathbf{r}) \equiv \sum_{\lambda \delta \in B} (C_{\lambda j}^B)^* C_{\delta j}^B \int d\mathbf{r}' \frac{\varphi_{\lambda}^*(\mathbf{r}') \varphi_{\delta}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (8)$$

The above potential is a function of fragment  $B$  *only* (more specifically on its  $j$ th molecular orbital). Thus, suitable parametrization of such potential can be very efficient. Similarly, the second integral can be simplified to

$$(\phi_i^A \phi_i^A | \phi_i^A \phi_j^B) = (\phi_i^A | v_{\text{eff},1}^{A[i]} | \phi_j^B) \quad (9)$$

with

$$v_{\text{eff},1}^{A[i]}(\mathbf{r}) \equiv \sum_{\nu \kappa \in A} (C_{\nu i}^A)^* C_{\kappa i}^A \int d\mathbf{r}' \frac{\varphi_{\nu}^*(\mathbf{r}') \varphi_{\kappa}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (10)$$

Now, I shall show that the above technique can be used to reduce the summations over MOs. Let us analyze

$$\sum_{l \in B} (\phi_i^A \phi_j^B | \phi_l^B \phi_l^B) = \sum_{l \in B} \sum_{\mu \in A} \sum_{\sigma \in B} \sum_{\lambda \delta \in B} (C_{\mu i}^A)^* C_{\sigma j}^B (C_{\lambda l}^B)^* C_{\delta l}^B (\mu \sigma | \lambda \delta) \quad (11)$$

$$= \sum_{\mu \in A} \sum_{\sigma \in B} \sum_{\lambda \delta \in B} (C_{\mu i}^A)^* C_{\sigma j}^B \underbrace{\sum_{l \in B} (C_{\lambda l}^B)^* C_{\delta l}^B}_{\frac{1}{2} P_{\delta \lambda}^B} (\mu \sigma | \lambda \delta) \quad (12)$$

$$= \sum_{\mu \in A} \sum_{\sigma \in B} (C_{\mu i}^A)^* C_{\sigma j}^B (\mu | v_{\text{eff},2}^B | \sigma) = (\phi_i^A | v_{\text{eff},2}^B | \phi_j^B) \quad (13)$$

where the effective potential is now defined as

$$v_{\text{eff},2}^B(\mathbf{r}) \equiv \frac{1}{2} \sum_{\lambda \delta \in B} P_{\delta \lambda}^B \int d\mathbf{r}' \frac{\varphi_{\lambda}^*(\mathbf{r}') \varphi_{\delta}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (14)$$

The quantity  $P_{\delta \lambda}^B$  is the reduced one-particle density matrix element. Analogously,

$$\sum_{k \in A} (\phi_i^A \phi_j^B | \phi_k^A \phi_k^A) = (\phi_i^A | v_{\text{eff},2}^A | \phi_j^B) \quad (15)$$

Now, I can combine three different functions:  $V_{ij}^B$ ,  $(\phi_i^A \phi_j^B | \phi_j^B \phi_j^B)$  and  $2 \sum_{l \in B} (\phi_i^A \phi_j^B | \phi_l^B \phi_l^B)$  (see  $f_t$  in Eq. (4)) and define the generalized effective one-electron potential due to  $j$ th occupied MO of  $B$  as follows

$$v_{\text{eff}}^{B[j]}(\mathbf{r}) = - \sum_{y \in B} \frac{Z_y^B}{|\mathbf{r} - \mathbf{r}_y|} + \sum_{\lambda \delta \in B} \{P_{\delta \lambda}^B - (C_{\lambda j}^B)^* C_{\delta j}^B\} \int d\mathbf{r}' \frac{\varphi_{\lambda}^*(\mathbf{r}') \varphi_{\delta}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (16)$$

which results in the following reduction

$$V_{ij}^B - (\phi_i^A \phi_j^B | \phi_j^B \phi_j^B) + 2 \sum_{l \in B} (\phi_i^A \phi_j^B | \phi_l^B \phi_l^B) \equiv (\phi_i^A | v_{\text{eff}}^{B[j]} | \phi_j^B) \quad (17)$$

The same reduction can be defined for molecule  $A$ . Thus, the first-order expression for Pauli repulsion interaction energy in Eq. (5) is greatly simplified into the following form

$$U(S^{-1}) = -2 \sum_{i \in A} \sum_{j \in B} S_{ij} [(\phi_i^A | v_{\text{eff}}^{A[i]} | \phi_j^B) + (\phi_i^A | v_{\text{eff}}^{B[j]} | \phi_j^B)] \quad (18)$$

It is extremely interesting to plot the distribution of potentials from Eq. (16) and examine whether it could be fit to a set of effective point multipoles. In the most simple formulation it could be approximated by ESP charges

$$v_{\text{eff}}^{B[j]}(\mathbf{r}) \approx \sum_D \frac{q_{D,j}^B}{|\mathbf{r} - \mathbf{r}_D|} \quad (19)$$

and one-electron integrals in Eq. (18) would be evaluated with great efficiency. Another scenario is to apply the density fitting scheme for the evaluation of one-electron integrals. That is, I can expand

$$v_{\text{eff}}^{B[j]} | \phi_j^B \rangle = \sum_u G_u^{(j)} | \gamma_u^{(j)} \rangle \quad (20)$$

For that, an appropriate auxiliary basis set  $\{|\gamma_u^{(j)}\rangle\}$  needs to be worked out and parameterized in terms of coefficients  $G_u^{(j)}$ . Here, I believe that, due to the nature of repulsive interactions, auxiliary functions could be just  $s$ -type functions and, in case of anisotropy or repulsion potential,  $p$ -type Gaussian lobes would be added as well.

### 1.4.2 EET Coupling Constants

GEFP technique of simplifying equations containing summations over ERIs can be also fruitfully applied to the evaluation of specific quantities such as excitation energy transfer (EET) couplings between electronic vertical transitions of coupled chromophores.

In the simplest case the EET coupling constant can be written as

$$V_{AB} = \iint d\mathbf{r}_1 d\mathbf{r}'_1 \frac{\rho_A^{eg*}(\mathbf{r}_1; \mathbf{r}_1) \rho_B^{eg}(\mathbf{r}'_1; \mathbf{r}'_1)}{|\mathbf{r}_1 - \mathbf{r}'_1|} - \frac{1}{2} \iint d\mathbf{r}_1 d\mathbf{r}'_1 \frac{\rho_A^{eg*}(\mathbf{r}_1; \mathbf{r}'_1) \rho_B^{eg}(\mathbf{r}'_1; \mathbf{r}_1)}{|\mathbf{r}_1 - \mathbf{r}'_1|} \quad (21)$$

$$\equiv J_{AB}(eg, eg) - K_{AB}(eg, eg) \quad (22)$$

It was shown before that the Coulombic coupling can be computed analytically by using *ab initio* distributed transition multipoles.[9] The exchange part is however not easy to evaluate, though it is relatively negligible in many cases. Eq. (21) can however be not sufficient when electron/hole/charge transfer mechanisms are dominant.

Let us analyze the contribution

$$V_{ET1} \cong t_{H \rightarrow L}^A \left[ (\phi_L^A | \hat{F} | \phi_L^B) + 2 (\phi_L^A \phi_H^A | \phi_H^A \phi_L^B) - (\phi_L^A \phi_L^B | \phi_H^A \phi_H^A) \right] \quad (23)$$

derived by Fujimoto. First of all,  $(\phi_L^A | \hat{F} | \phi_L^B)$  term is quite difficult to evaluate since it involves the intermolecular matrix elements of Fock operator,  $\hat{F}$ . However, I can notice that

$$(\phi_L^A | \hat{F} | \phi_L^B) = \sum_{\mu \in A} \sum_{\sigma \in B} (C_{\mu L}^A)^* C_{\sigma L}^B F_{\mu\sigma} \quad (24)$$

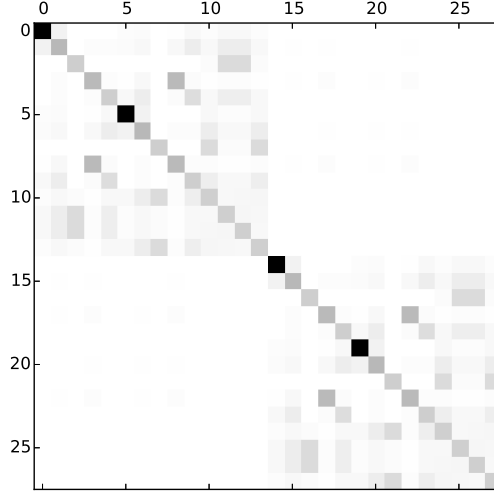
$$= \sum_{\mu \in A} \sum_{\sigma \in B} (C_{\mu L}^A)^* C_{\sigma L}^B \left[ h_{\mu\sigma}^{\text{core}} + \sum_{\alpha\beta}^{A,B} P_{\alpha\beta} \left\{ (\mu\sigma | \alpha\beta) - \frac{1}{2} (\mu\beta | \alpha\sigma) \right\} \right] \quad (25)$$

$$\cong \sum_{\mu \in A} \sum_{\sigma \in B} (C_{\mu L}^A)^* C_{\sigma L}^B \left[ h_{\mu\sigma}^{\text{core}} + \sum_{\nu\kappa \in A} P_{\nu\kappa}^A \left\{ (\mu\sigma | \nu\kappa) - \frac{1}{2} (\mu\kappa | \nu\sigma) \right\} \right. \\ \left. + \sum_{\delta\lambda \in B} P_{\delta\lambda}^B \left\{ (\mu\sigma | \delta\lambda) - \frac{1}{2} (\mu\lambda | \delta\sigma) \right\} \right] \quad (26)$$

where I assumed that the density matrix  $P$  is diagonal in AO representation and has non-zero blocks only for fragments  $A$  and  $B$ .

In Figure 2 I present the absolute values of  $P$  in AO basis computed for an ethylene dimer in stacked  $D_{2h}$  geometry with the distance between carbon atoms equal to 3 Å (HF/STO-3G level of theory). It can be noticed that  $P$  appears as a block matrix and its elements corresponding to bond orders between ethylene monomers are relatively vanishing. Thus, I believe that I can adopt the above approximation which was used to write Eq. (24).





**Figure 2:** Absolute magnitude of density matrix of ethylene dimer.

Now, I can also notice that the dummy indices can be rearranged in Eq. (24) so that I have

$$\sum_{\mu \in A} \sum_{\sigma \in B} (C_{\mu L}^A)^* C_{\sigma L}^B \sum_{\nu \kappa \in A} P_{\nu \kappa}^A \left\{ (\mu \sigma | \nu \kappa) - \frac{1}{2} (\mu \kappa | \nu \sigma) \right\} \quad (27)$$

$$= \sum_{\mu \in A} \sum_{\sigma \in B} \sum_{\nu \kappa \in A} (C_{\mu L}^A)^* C_{\sigma L}^B P_{\nu \kappa}^A (\mu \sigma | \nu \kappa) - \frac{1}{2} \sum_{\mu \in A} \sum_{\sigma \in B} \sum_{\nu \kappa \in A} (C_{\mu L}^A)^* C_{\sigma L}^B P_{\nu \kappa}^A (\mu \kappa | \nu \sigma) \quad (28)$$

$$= \text{Interchange dummy indices } \nu \text{ and } \mu \text{ in the second term} \quad (29)$$

$$= \text{Notice that } (\nu \kappa | \mu \sigma) = (\mu \sigma | \nu \kappa) \quad (30)$$

$$= \sum_{\mu \in A} \sum_{\sigma \in B} C_{\sigma L}^B \sum_{\nu \kappa \in A} \left\{ (C_{\mu L}^A)^* P_{\nu \kappa}^A - \frac{1}{2} (C_{\nu L}^A)^* P_{\mu \kappa}^A \right\} (\mu \sigma | \nu \kappa) \quad (31)$$

$$= \sum_{\mu \in A} \sum_{\sigma \in B} (\mu | v_{\text{el},1}^{A[\mu]} | \sigma) C_{\sigma L}^B \quad (32)$$

with the effective one-electron potential being defined as

$$v_{\text{el},1}^{A[\mu]}(\mathbf{r}) \equiv \sum_{\nu \kappa \in A} \left\{ (C_{\mu L}^A)^* P_{\nu \kappa}^A - \frac{1}{2} (C_{\nu L}^A)^* P_{\mu \kappa}^A \right\} \int d\mathbf{r}' \frac{\varphi_{\nu}^*(\mathbf{r}') \varphi_{\kappa}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (33)$$

Analogously, I define

$$\sum_{\mu \in A} \sum_{\sigma \in B} (C_{\mu L}^A)^* C_{\sigma L}^B \sum_{\delta \lambda \in B} P_{\delta \lambda}^B \left\{ (\mu \sigma | \delta \lambda) - \frac{1}{2} (\mu \lambda | \delta \sigma) \right\} \equiv \sum_{\mu \in A} \sum_{\sigma \in B} (C_{\mu L}^A)^* (\mu | v_{\text{el},1}^{B[\sigma]} | \sigma) \quad (34)$$

I can now see that I eliminated all ERIs from matrix elements of Fock operator from Eq. (24). I need to figure out still one-electron part. Let us rewrite

$$\sum_{\mu \in A} \sum_{\sigma \in B} (C_{\mu L}^A)^* C_{\sigma L}^B \left\{ V_{\mu \sigma}^A + V_{\mu \sigma}^B + T_{\mu \sigma} \right\} \quad (35)$$

$$= \sum_{\mu \in A} \sum_{\sigma \in B} \left\{ (C_{\mu L}^A)^* C_{\sigma L}^B T_{\mu \sigma} + (\mu | v_{\text{nuc-el}}^{A[\mu]} | \sigma) C_{\sigma L}^B + (C_{\mu L}^A)^* (\mu | v_{\text{nuc-el}}^{B[\sigma]} | \sigma) \right\} \quad (36)$$

and define the following potentials

$$v_{\text{nuc-el}}^{A[\mu]}(\mathbf{r}) = (C_{\mu L}^A)^* \sum_{x \in A} \frac{-Z_x^A}{|\mathbf{r} - \mathbf{r}_x|} \quad (37a)$$

$$v_{\text{nuc-el}}^{B[\sigma]}(\mathbf{r}) = C_{\sigma L}^B \sum_{y \in B} \frac{-Z_y^B}{|\mathbf{r} - \mathbf{r}_y|} \quad (37b)$$

Now I can notice that

$$(\phi_L^A | \hat{F} | \phi_L^B) = \left( \phi_L^A | -\frac{1}{2} \nabla^2 | \phi_L^B \right) + \sum_{\mu \in A} \sum_{\sigma \in B} \left[ (\mu | v_{\text{eff}}^{A[\mu]} | \sigma) C_{\sigma L}^B + (C_{\mu L}^A)^* (\mu | v_{\text{eff}}^{B[\sigma]} | \sigma) \right] \quad (38)$$

where the effective nuclear and electron potentials read

$$v_{\text{eff}}^{A[\mu]}(\mathbf{r}) \equiv v_{\text{nuc-el}}^{A[\mu]}(\mathbf{r}) + v_{\text{el},1}^{A[\mu]}(\mathbf{r}) \quad (39)$$

and similarly for  $B$ .

The last part of  $V_{\text{ET1}}$  is due to the ERIs evaluated in MO basis. It is already straightforward to notice that

$$(\phi_L^A \phi_H^A | \phi_H^A \phi_L^B) = \sum_{\mu \in A} \sum_{\sigma \in B} (\mu | v_{\text{el},2}^{A[\mu]} | \sigma) C_{\sigma L}^B \quad (40a)$$

$$(\phi_L^A \phi_L^B | \phi_H^A \phi_H^A) = \sum_{\mu \in A} \sum_{\sigma \in B} (C_{\mu L}^A)^* (\mu | v_{\text{el},3}^{A[\mu]} | \sigma) \quad (40b)$$

with the potentials defined as follows

$$v_{\text{el},2}^{A[\mu]}(\mathbf{r}) \equiv (C_{\mu H}^A)^* \sum_{\nu \kappa \in A} (C_{\nu L}^A)^* C_{\kappa H}^A \int d\mathbf{r}' \frac{\varphi_{\nu}^*(\mathbf{r}') \varphi_{\kappa}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (41a)$$

$$v_{\text{el},3}^{A[\mu]}(\mathbf{r}) \equiv (C_{\mu L}^A)^* \sum_{\nu \kappa \in A} (C_{\nu H}^A)^* C_{\kappa H}^A \int d\mathbf{r}' \frac{\varphi_{\nu}^*(\mathbf{r}') \varphi_{\kappa}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (41b)$$

Therefore, finally I recognize the generalized effective potentials due to  $A$  and  $B$  as

$$(\phi_L^A | \hat{F} | \phi_L^B) + 2 (\phi_L^A \phi_H^A | \phi_H^A \phi_L^B) - (\phi_L^A \phi_L^B | \phi_H^A \phi_H^A) \quad (42)$$

$$= \left( \phi_L^A | -\frac{1}{2} \nabla^2 | \phi_L^B \right) + \sum_{\mu \in A} \sum_{\sigma \in B} \left[ (\mu | v_{\text{ET1}}^{A[\mu]} | \sigma) C_{\sigma L}^B + (C_{\mu L}^A)^* (\mu | v_{\text{ET1}}^{B[\sigma]} | \sigma) \right] \quad (43)$$

with the following definitions

$$v_{\text{ET1}}^{A[\mu]}(\mathbf{r}) = v_{\text{el},1}^{A[\mu]}(\mathbf{r}) + v_{\text{nuc-el}}^{A[\mu]}(\mathbf{r}) + 2v_{\text{el},2}^{A[\mu]}(\mathbf{r}) - v_{\text{el},3}^{A[\mu]}(\mathbf{r}) \quad (44a)$$

$$v_{\text{ET1}}^{B[\sigma]}(\mathbf{r}) = v_{\text{el},1}^{B[\sigma]}(\mathbf{r}) + v_{\text{nuc-el}}^{B[\sigma]}(\mathbf{r}) \quad (44b)$$

To sum up these derivations, the ET1 contribution reads

$$V_{\text{ET1}} \cong t_{H \rightarrow L}^A \times \left\{ \left( \phi_L^A | -\frac{1}{2} \nabla^2 | \phi_L^B \right) + \sum_{\mu \in A} \sum_{\sigma \in B} \left[ (\mu | v_{\text{ET1}}^{A[\mu]} | \sigma) C_{\sigma L}^B + (C_{\mu L}^A)^* (\mu | v_{\text{ET1}}^{B[\sigma]} | \sigma) \right] \right\} \quad (45)$$

which is a significant simplification over the original expression given in Eq. (23), even if it is still formally equivalent under the assumption that the density matrix has a block form.

Here I remark that, contrary to repulsive potential discussed previously, I defined the one-electron effective potentials  $v_{\text{ET1}}^{A[\mu]}(\mathbf{r})$  and  $v_{\text{ET1}}^{B[\sigma]}(\mathbf{r})$  in AO basis. This is because the extent of simplification can be greater when explicit basis functions are used instead of molecular orbitals. Note also that, in practice, repulsion effective potentials need still to be evaluated in AO basis, but their total amount is much smaller.

Plausible further analysis is to expand

$$v_{\text{ET1}}^{B[\sigma]}(\mathbf{r}) \approx \sum_D \frac{q_{D,\sigma}^B}{|\mathbf{r} - \mathbf{r}_D|} \quad (46)$$

or follow the density fitting scheme

$$v_{\text{ET1}}^{B[\sigma]}|\sigma\rangle = \sum_u G_u^{(\sigma)}|\gamma_u^{(\sigma)}\rangle \quad (47)$$

Exactly the same manipulations can be continued for  $V_{\text{ET2}}$ , as well as  $V_{\text{HT1}}$ ,  $V_{\text{HT2}}$ ,  $V_{\text{CT1}}$  and  $V_{\text{CT2}}$ .

## 1.5 Limitations of Effective One-Electron Potentials

While defining smart one-electron effective potentials might drastically simplify equations and their evaluations for larger molecular systems, there are situations in which there is no way to simplify ERIs by using technique described in previous sections. One of the obstacles is tackling integrals of the form  $(\phi^A \phi^B | \phi^A \phi^B)$ . Example of use of such integrals can be in exchange energy, attractive part of exchange-repulsion potential. The solution we shall adopt later in this Proposal is to use the spherical Gaussian overlap (SGO) approximation by Jensen [10] which was shown by him to provide accurate estimations of the exchange energy at SCF level. However, this trick cannot be applied for fast evaluation of exchange contribution to EET coupling constant,  $K(ab)$ . It is however not so surprising - there are processes which have significantly two-electron character, thus preventing from using the 1-electron potentials. Nevertheless, there is some evidence that these 2-electron effects are rather rarely of significant strengths and virtually 1-electron effects, even if blurred by appearance of complicated 2-electron integrals, can be extracted and used extensively in basically all fields of quantum chemistry of large, extended molecular aggregates.

## 2 Energetic Properties of Molecules

Here I design the theory for a typical intermolecular interaction that is characterised by small degree of approximation. I express the intermolecular interaction potential by making use of only 1-electron effective potentials and approximating pure exchange integrals over MOs by using spherical Gaussian overlap approximation of Jensen. First I reinvestigate EFP model at Hartree-Fock level which is then generalized to arbitrary level of theory.

### 2.1 Full Intermolecular Interaction Energy Expression

The computation of intermolecular interaction energy was recently generalized to any wave function-based methodology by Mandado and Hermida-Ramón in their density-based parti-

tioning scheme, [2] which can be shortly written as

$$\Delta E = \underbrace{E_{\text{nuc-nuc}} + E_{\text{nuc-el}} + E_{\text{el-el}}}_{E_{\text{Elec}}} + \underbrace{E_{\text{nuc-Pauli}} + E_{\text{el-Pauli}} + E_{\text{Pauli-Pauli}} + E_{\text{Pauli}}^{\nabla^2}}_{E_{\text{Rep}}} + \underbrace{E_{\text{nuc-Pol}} + E_{\text{el-Pol}} + E_{\text{Pol-Pol}} + E_{\text{Pol}}^{\nabla^2} + E_{\text{Pauli-Pol}}}_{E_{\text{Pol}} \text{ without exchange-correlation}} + \underbrace{E_{\text{exch}} + E_{\text{Pol}}^{\text{XC}}}_{\text{exchange/correlation}} \quad (48)$$

Due to technical reasons, I consider the ‘polarization’ term  $E_{\text{Pol}}$  without the exchange-correlation term  $E_{\text{Pol}}^{\text{XC}}$  (which is a part of  $E_{\text{Pol}}$  in Ref.), and this contribution is analyzed separately along with the ‘exchange’ energy,  $E_{\text{exch}}$ . To obtain the interaction energy from Eq. (48) one has to perform extensive integrations over the one- and two-particle electronic densities

$$\rho(\mathbf{r}_1) = \rho_A(\mathbf{r}_1) + \rho_B(\mathbf{r}_1) + \Delta\rho_{\text{Pauli}}(\mathbf{r}_1) + \Delta\rho_{\text{Pol}}(\mathbf{r}_1) \quad (49a)$$

$$\rho(\mathbf{r}_1, \mathbf{r}_2) = \rho_{\text{XC},A}(\mathbf{r}_1, \mathbf{r}_2) + \rho_{\text{XC},B}(\mathbf{r}_1, \mathbf{r}_2) + \rho_{\text{XC},AB}(\mathbf{r}_1, \mathbf{r}_2) + \Delta\rho_{\text{XC}}(\mathbf{r}_1, \mathbf{r}_2) \quad (49b)$$

These operations involve computations of ERI’s and are prohibitively time consuming for condensed matter physics applications. Moreover, deformation densities and the intermolecular part of unpolarized exchange-correlation density need to be computed for each molecular cluster separately. In this Proposal, I shall work with Eq. (48) and apply techniques introduced in the previous sections to dramatically simplify the above expression, attempting to introduce very little approximations but still moving ERI’s *completely* from the working model to the generalized effective fragments of isolated monomers.

## 2.2 Hartree–Fock Intermolecular Interaction Energy Based on GEFP

Hartree–Fock theory is the basic prerequisite of understanding the general electronic structure problem, in which the many-body wave function is represented by Slater determinants. Therefore, in order to build the promising EFP-based model for an arbitrary wave function, one has to first elaborate such a model for ground state solutions to Hartree–Fock equations which could be then extended by introducing dynamic and non-dynamic electron correlation. Existing EFP2 model has many attractive advantages. However, its theoretical formulation makes it extremely difficult to develop an analogous model on top of a multiconfigurational wave function ansatz, inherently encoded within electronically excited states. This means that I need another, more general formulation. In this Section, I elaborate on the generalized EFP theory at HF level that, in my belief, meets the following criteriae:

1. In principle, the proposed model is not based on perturbation theory, but on supermolecular approach.
2. The model should provide interaction energies which are very close to the exact solution. The magnitude of error is believed to be of the order of the BSSE stabilization energy.
3. The necessary approximations are limited only to i) truncations of Taylor expansion of density matrices and ii) approximate treatment of exchange two-electron integrals.
4. The resulting working equations contain only one-electron effective operators or functions.
5. There is no inherent semi-empirical or empirical parameterization.
6. The model is believed to be extendable to electronically excited states.

The model has two obvious drawbacks and limitations:

1. The present formulation assumes monomer-centered basis sets (MCBS) which means that charge-transfer effects and BSSE are not accounted for.
2. The resulting effective fragment potentials are computed for a particular nuclear conformation. Therefore, all internal coordinates within GEFPs are frozen. This means that the method is applicable only to relatively small and structurally rigid molecules.

I first shortly discuss the pseudo-classical Coulombic part of the model. Then I focus on more challenging, remaining parts of interaction energy: exchange-repulsion and polarization energy.

### 2.2.1 Electrostatics

The easiest part of the model is electrostatics and it can be already very efficiently treated by using existing approaches. However, I shall use the one-electron effective potential technique to be consistent with other parts of the model. I can write

$$E_{\text{Elec}} = E_{\text{nuc-nuc}} + \iint \frac{\rho_A(\mathbf{r}_1)\rho_B(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \quad (50)$$

The double integral can be expressed in terms of one-particle density matrices of unperturbed molecules  $A$  and  $B$  as

$$\iint \frac{\rho_A(\mathbf{r}_1)\rho_B(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 = 4 \sum_{\mu\nu \in A} \sum_{\lambda\sigma \in B} D_{\mu\nu}^A D_{\lambda\sigma}^B (\mu\nu|\lambda\sigma) = 2 \sum_{\mu\nu \in A} (\mu|v^B|\nu) D_{\mu\nu}^A \quad (51)$$

where  $v^B$  is a typical electrostatic potential operator due to molecule  $B$ . As it can be seen, the above expression is asymmetric with respect to molecules  $A$  and  $B$ . However, I shall see in the due course of this Proposal that both the matrix elements  $v_{\mu\nu}^B$  and  $v_{\lambda\sigma}^A$  are needed. Therefore, I rewrite the electrostatic integral as

$$\iint \frac{\rho_A(\mathbf{r}_1)\rho_B(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 = \sum_{\mu\nu \in A} (\mu|v^B|\nu) D_{\mu\nu}^A + \sum_{\lambda\sigma \in B} (\lambda|v^A|\sigma) D_{\lambda\sigma}^B \quad (52)$$

Here I use monomer-centered basis sets (MCBS). This approximation is acceptable for large basis sets for which basis set superimposition errors (BSSE) are generally smaller than 1 kcal/mole. If I wanted to use dimer-centered basis sets (DCBS) here I couldn't derive GEFP model since monomer density matrices would be dependent on their relative orientations.

I also see that  $v^A$  is the simplest possible effective one-electron potential which can be parameterized here by point charges and stored in GEFP file along with unperturbed density matrix in AO basis. To summarize, electrostatic energy in GEFP with atomic site expansion of electrostatic potential is simply

$$E_{\text{Elec}} = \sum_{x \in A} \sum_{y \in B} \frac{Z_x Z_y}{|\mathbf{r}_x - \mathbf{r}_y|} + \sum_{y \in B} \sum_{\mu\nu \in A} v_{\mu\nu}^y D_{\mu\nu}^A + \sum_{x \in A} \sum_{\lambda\sigma \in B} v_{\lambda\sigma}^x D_{\lambda\sigma}^B \quad (53)$$

where  $v_{\lambda\sigma}^x$  is the standard potential one-electron integral. Note that, if the parameterization of electrostatic potential into atomic charges is accurate, the resulting Coulombic interaction energy is exact and takes into account charge-penetration effects systematically.

More accurate possibility is to use resolution of identity (RI) technique to parameterize  $v^A$  in a space of auxiliary basis functions, most probably composed of  $s$ -type orbitals.

### 2.2.2 Repulsion

Evaluation of repulsive interaction energy requires computation of the one-electron deformation density due to Pauli exclusion principle,

$$\Delta\rho_{\text{Pauli}}(\mathbf{r}) = 2 \sum_{\alpha\beta} \left( D_{\alpha\beta}^{oo} - D_{\alpha\beta}^{AB} \right) \phi_{\alpha}(\mathbf{r}) \phi_{\beta}(\mathbf{r}) = 2\phi^{\dagger} \Delta \mathbf{D}^{\text{Pauli}} \phi \quad (54)$$

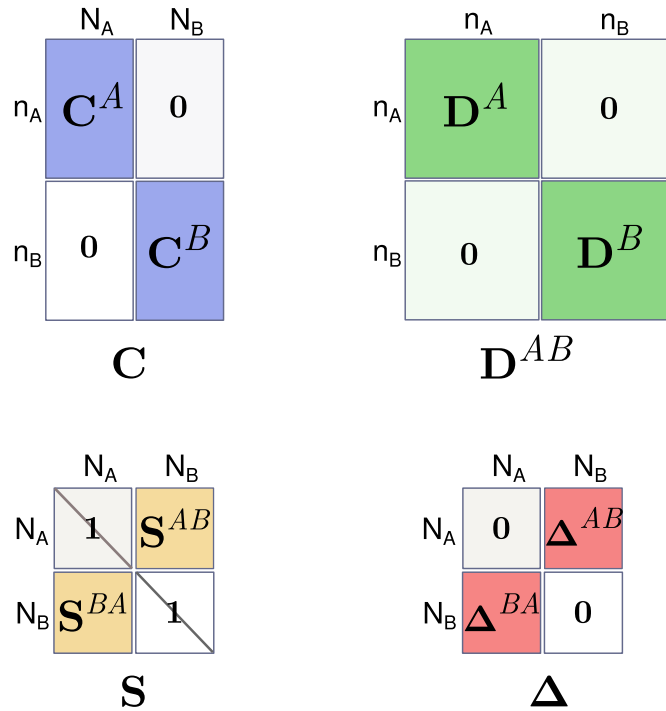
which is constructed from the difference between the anti-symmetrized densities and the sum of unperturbed densities. Here, labels  $\alpha$  and  $\beta$  run over all AOs,  $D_{\alpha\beta}^{oo}$  is the density matrix of  $AB$  complex formed from orthogonalized molecular orbitals (hence superscript  $oo$ ), whereas  $D_{\alpha\beta}^{AB}$  is assumed here to be the composite sum of the monomers' unperturbed density matrices,

$$\mathbf{D}^{AB} = \mathbf{D}^A \oplus \mathbf{D}^B \quad (55)$$

because I cannot use DCBS. The anti-symmetrized density matrix is

$$\mathbf{D}^{oo} = [\mathbf{C}\mathbf{X}] [\mathbf{C}\mathbf{X}]^{\dagger} \quad (56)$$

where  $\mathbf{C}$  is the LCAO-MO coefficient matrix shown schematically in Figure, and  $\mathbf{X}$  is the



**Figure 3:** The matrices that are needed to consider the expansion of repulsion interaction energy in terms of overlap integrals between unperturbed occupied molecular orbitals of the fragments  $A$  and  $B$ .

orthogonalizer obtained by means of the Löwdin orthogonalization procedure

$$\mathbf{X} = \mathbf{S}^{-\frac{1}{2}} \quad (57)$$

Here,  $S_{ij}$  is the overlap matrix in MO basis. It is clear that  $\mathbf{S}$  has  $AA$  and  $BB$  block equal to identity matrices. I can re-express  $\mathbf{D}^{oo}$  as

$$\mathbf{D}^{oo} = \mathbf{C}\mathbf{X}\mathbf{X}^{\dagger}\mathbf{C}^{\dagger} = \mathbf{C}\mathbf{S}^{-1}\mathbf{C}^{\dagger} \quad (58)$$

Therefore, the difference density matrix reads

$$\Delta \mathbf{D}^{\text{Pauli}} \equiv \mathbf{D}^{oo} - \mathbf{D}^{AB} = \mathbf{C} [\mathbf{S}^{-1} - \mathbf{1}] \mathbf{C}^\dagger \quad (59)$$

The computational cost of evaluation of  $\mathbf{D}^{oo} - \mathbf{D}^{AB}$  is not as great as supermolecular calculations involving ERIs and I could stop here. However, it will appear useful to analyse the expansion of  $\mathbf{D}^{oo}$  as a power series of the overlap matrix as done previously by Jensen,

$$\Delta \mathbf{D}^{\text{Pauli}} = \mathbf{C} [-\Delta + \Delta^2 - \dots] \mathbf{C}^\dagger \quad (60)$$

where  $\Delta_{ij} = S_{ij}$  for  $i \neq j$  and  $\Delta_{ii} = 0$ . Thus, I find

$$\begin{aligned} \frac{1}{2} \Delta \rho_{\text{Pauli}} = & \underbrace{-\phi_A^\dagger \mathbf{C}_A \Delta_{AB} \mathbf{C}_B^\dagger \phi_B - \phi_B^\dagger \mathbf{C}_B \Delta_{BA} \mathbf{C}_A^\dagger \phi_A}_{\text{first-order in } \Delta} \\ & + \underbrace{\phi_A^\dagger \mathbf{C}_A \Delta_{AB} \Delta_{BA} \mathbf{C}_A^\dagger \phi_A + \phi_B^\dagger \mathbf{C}_B \Delta_{BA} \Delta_{AB} \mathbf{C}_B^\dagger \phi_B}_{\text{second-order in } \Delta} + \dots \end{aligned} \quad (61)$$

which enables full factorization of  $A$  and  $B$  parts into effective potentials.

Now, let us derive  $E_{\text{nuc-Pauli}} + E_{\text{el-Pauli}}$  first. The contribution from  $A$ th charge-density distribution is given as

$$E_{\text{nuc-Pauli}}^A + E_{\text{el-Pauli}}^A = \sum_{x \in A} Z_x \int \frac{\Delta \rho_{\text{Pauli}}(\mathbf{r}_1)}{|\mathbf{r}_1 - \mathbf{r}_x|} d\mathbf{r}_1 + \iint \frac{\Delta \rho_{\text{Pauli}}(\mathbf{r}_1) \rho_A(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \quad (62)$$

This can be recast as follows

$$E_{\text{nuc-Pauli}}^A + E_{\text{el-Pauli}}^A = \sum_{\alpha\beta} (\alpha|v^A|\beta) \Delta D_{\alpha\beta}^{\text{Pauli}} \quad (63)$$

Note however, that now I cannot expand  $v^A$  in charges (or other multipoles) because the summation goes over all AOs. There are two routes here to test: (i) to employ an auxiliary basis set approach as proposed by Darden in his GEM model or (ii) to expand Pauli deformation matrix in power series of  $\Delta$  and truncate at low order. Here I consider explicitly the second possibility where I truncate the expansion of  $\Delta \rho_{\text{Pauli}}$  at second order:

$$\begin{aligned} \text{1st order due to } A = & \sum_{\mu \in A} \sum_{\lambda \in B} \sum_{i \in A} \sum_{j \in B} (\mu|v^A|\lambda) C_{\mu i}^A C_{\lambda j}^B S_{ij} \\ = & \sum_{\lambda, ij} S_{ij} C_{j\lambda}^B \left[ \sum_{\mu} C_{\mu i}^A (\mu|v^A|) \right] |\lambda\rangle \equiv \sum_{ij} S_{ij} \sum_{\zeta\lambda} C_{\lambda j}^B G_{\zeta i}^A(\zeta|\lambda) \end{aligned} \quad (64)$$

where I introduced the following generalized density fitting scheme

$$\sum_{\mu} C_{\mu i}^A (\mu|v^A|) \equiv \sum_{\zeta} G_{\zeta i}^A(\zeta|) \quad (65)$$

and the auxiliary basis set has to be worked out. I believe that due to the nature of repulsive interactions such basis set can be constructed from  $s$ -type functions only (perhaps including GLOs). Similar approach has been already developed by Darden's group for generation of fast semi-empirical forcefields. Here I attempt to develop fully first-principles forcefield because auxiliary basis functions  $(\zeta|$  and expansion coefficients  $G_{\zeta i}^A$  are functions of unperturbed molecule  $A$  only, making them one of generalized effective fragment potentials.

Similarly, at the second order I have

$$\begin{aligned} \text{2nd order due to } A = \sum_{ii' \in A} \sum_{j \in B} \sum_{\mu\nu \in A} (\mu|v^A|\nu) C_{\mu i}^A C_{\nu i'}^A S_{ij} S_{i'j} \\ + \sum_{jj' \in B} \sum_{i \in A} \sum_{\lambda\sigma \in B} (\lambda|v^A|\sigma) C_{\lambda j}^B C_{\sigma j'}^B S_{ij} S_{i'j'} \end{aligned} \quad (66)$$

Similar considerations apply for the contribution from molecule  $B$  charge density.

Next, I have to work out the contribution to the difference in kinetic energy due to Pauli deformation densities,

$$E_{\text{Pauli}}^{\nabla^2} = -\frac{1}{2} \int \nabla^2 \{ \Delta \rho_{\text{Pauli}}(\mathbf{r}_1, \mathbf{r}'_1) \}_{\mathbf{r}'_1 = \mathbf{r}_1} d\mathbf{r}_1 = 2 \sum_{\alpha\beta} \Delta D_{\alpha\beta}^{\text{Pauli}} \left( \alpha | -\frac{1}{2} \nabla^2 | \beta \right) \quad (67)$$

This could be already evaluated in the above form and compute kinetic one-electron integrals between all basis functions on the fly. However, to be consistent with the previous considerations, I expand  $\Delta D_{\alpha\beta}^{\text{Pauli}}$  and obtain

$$E_{\text{Pauli}}^{\nabla^2} \cong -4 \sum_{ij} S_{ij} \left( i | -\frac{1}{2} \nabla^2 | j \right) + \text{2nd-order terms} \quad (68)$$

which is easy to derive from that point and is not shown here in detail.

The most difficult term to evaluate is however the interaction of Pauli deformation densities of both molecules,

$$E_{\text{Pauli-Pauli}} = \frac{1}{2} \iint \frac{\Delta \rho_{\text{Pauli}}(\mathbf{r}_1) \Delta \rho_{\text{Pauli}}(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 = 2 \sum_{\alpha\beta\gamma\delta} \Delta D_{\alpha\beta}^{\text{Pauli}} \Delta D_{\gamma\delta}^{\text{Pauli}} (\alpha\beta|\gamma\delta) \quad (69)$$

Here I expand Pauli deformation density matrices up to first order only. This approximation is good if truncation up to quadratic terms with respect to overlap matrix is sufficient. Careful examination of this assumption leads to the result

$$E_{\text{Pauli-Pauli}} \cong -8 \sum_{ii' \in A} \sum_{jj' \in B} (ij|i'j') S_{ij} S_{i'j'} \quad (70)$$

which cannot be decomposed into separate contributions from unperturbed molecules  $A$  and  $B$  without further approximations. If I use localized orbitals then I can expect that  $(ij|i'j') \ll (ij|ij)$  for  $i' \neq i$  and  $j' \neq j$ . However, because the leading term here contains the product of these likely to be small repulsion integrals with the overlap matrix element squared, I assume this to be very small,

$$E_{\text{Pauli-Pauli}} \approx 0 \quad (71)$$

This expectation shall be tested in the future.

To summarize the above derivations, the total Pauli-repulsion energy is equal to

$$\begin{aligned} E^{\text{Rep}} \approx & \underbrace{-\text{Tr} [\mathbf{S}^{AB} \mathbf{S}^{BA}] - \text{Tr} [\mathbf{S}^{AB} \mathbf{S}^{BA}]}_{E_{\text{nuc-Pauli}} + E_{\text{el-Pauli}}; \text{1st-order in } \Delta} \\ & + \underbrace{\text{Tr} \left[ \left( \mathbf{v}^A + \mathbf{v}_{AA}^B \right) \mathbf{S}^{AB} \mathbf{S}^{BA} \right] + \text{Tr} \left[ \left( \mathbf{v}_{BB}^A + \mathbf{v}^B \right) \mathbf{S}^{BA} \mathbf{S}^{AB} \right]}_{E_{\text{nuc-Pauli}} + E_{\text{el-Pauli}}; \text{2nd-order in } \Delta} \\ & - \underbrace{4 \text{Tr} [\mathbf{T}^{AB} \mathbf{S}^{BA}] + 2 \text{Tr} [\mathbf{T}^A \mathbf{S}^{AB} \mathbf{S}^{BA}] + 2 \text{Tr} [\mathbf{T}^B \mathbf{S}^{BA} \mathbf{S}^{AB}]}_{E_{\text{Pauli}}^{\nabla^2}} \end{aligned} \quad (72)$$



or, by gathering terms with respect to orbital overlap power I get

$$E^{\text{Rep}} \approx \underbrace{\text{Tr} \left[ \left( 4\mathbf{T}^{AB} - \mathbf{s}^{AB} - [\mathbf{s}^{BA}]^T \right) \mathbf{S}^{BA} \right]}_{\text{first order}} + \underbrace{\text{Tr} \left[ \left( 2\mathbf{F}^A + \mathbf{v}_{AA}^B \right) \mathbf{S}^{AB} \mathbf{S}^{BA} \right] + \text{Tr} \left[ \left( 2\mathbf{F}^B + \mathbf{v}_{BB}^A \right) \mathbf{S}^{BA} \mathbf{S}^{AB} \right]}_{\text{second order}} \quad (73)$$

where I defined the following matrices

$$[\mathbf{S}^{AB}]_{ij} = S_{ij} = [\mathbf{S}^{BA}]_{ji} = S_{ji} \quad (74a)$$

$$[\mathbf{s}^{AB}]_{ij} = s_{ij} = \sum_{\zeta\lambda} G_{\zeta i}^A s_{\zeta\lambda} C_{\lambda j}^B \quad (74b)$$

$$[\mathbf{s}^{BA}]_{ji} = s_{ji} = \sum_{\eta\mu} G_{\eta j}^B s_{\eta\mu} C_{\mu i}^A \neq s_{ij} \quad (74c)$$

$$[\mathbf{v}_{AA}^B]_{ii'} = \sum_{\mu\nu} C_{\mu i}^A C_{\nu i'}^A (\mu|v_B|\nu) \quad (74d)$$

$$[\mathbf{v}_{BB}^A]_{jj'} = \sum_{\lambda\sigma} C_{\lambda j}^B C_{\sigma j'}^B (\lambda|v_A|\sigma) \quad (74e)$$

$$[\mathbf{v}^A]_{ii'} = \sum_{\mu\nu} C_{\mu i}^A C_{\nu i'}^A (\mu|v_A|\nu) \quad (74f)$$

$$[\mathbf{v}^B]_{jj'} = \sum_{\lambda\sigma} C_{\lambda j}^B C_{\sigma j'}^B (\lambda|v_B|\sigma) \quad (74g)$$

$$[\mathbf{T}^A]_{ii'} = \sum_{\mu\nu} C_{\mu i}^A C_{\nu i'}^A \left( \mu | -\frac{1}{2} \nabla^2 | \nu \right) \quad (74h)$$

$$[\mathbf{T}^B]_{jj'} = \sum_{\lambda\sigma} C_{\lambda j}^B C_{\sigma j'}^B \left( \lambda | -\frac{1}{2} \nabla^2 | \sigma \right) \quad (74i)$$

$$[\mathbf{T}^{AB}]_{ij} = \sum_{\mu\lambda} C_{\mu i}^A C_{\lambda j}^B \left( \mu | -\frac{1}{2} \nabla^2 | \lambda \right) \quad (74j)$$

$$\mathbf{F}^A = \mathbf{T}^A + \frac{1}{2} \mathbf{v}^A \quad (74k)$$

$$\mathbf{F}^B = \mathbf{T}^B + \frac{1}{2} \mathbf{v}^B \quad (74l)$$

Here, the effective fragment parameters are  $\mathbf{G}^X$  and  $\mathbf{F}^X$  (or linear combination of  $\mathbf{v}^X$  and  $\mathbf{T}^X$ ). To evaluate repulsion energy, I need to compute six matrices on the fly:  $\mathbf{S}^{AB}$ ,  $\mathbf{s}^{AB}$ ,  $\mathbf{s}^{BA}$ ,  $\mathbf{v}_{AA}^B$ ,  $\mathbf{v}_{BB}^A$  and  $\mathbf{T}^{AB}$ . Note that at HF level, the sums  $2\mathbf{T}^A + \mathbf{v}^A = 2\mathbf{F}^A$  and  $2\mathbf{T}^B + \mathbf{v}^B = 2\mathbf{F}^B$  are proportional to Fock matrices of isolated monomers and can be already stored as GEFP parameters just like in the case of conventional EFP2. Also,  $\mathbf{v}_{AA}^B$  and  $\mathbf{v}_{BB}^A$  are already needed for Coulombic interaction energy calculations. Thus, one has to compute 4 new matrices, perform two matrix multiplications,  $\mathbf{S}^{AB} \mathbf{S}^{BA}$  and  $\mathbf{S}^{BA} \mathbf{S}^{AB}$ , add the appropriate matrices and perform finally three traces of the three matrix multiplications results. I expect that the following procedure is only slightly more expensive than the conventional, more approximate in nature EFP2 repulsion energy:

$$E_{\text{EFP2}}^{\text{Rep}} = 4\text{Tr} [\mathbf{T}^{AB} \mathbf{S}^{BA}] - 2\text{Tr} [\mathbf{F}^B \mathbf{S}^{BA} \mathbf{S}^{AB}] - 2\text{Tr} [\mathbf{F}^A \mathbf{S}^{AB} \mathbf{S}^{BA}] - 2\text{Tr} [\mathbf{S}^{AB} \mathbf{R}^{BA}] \quad (75)$$

where

$$\Sigma_{ij} = S_{ij}^2 \quad (76)$$

and the matrix  $\mathbf{R}$ , which describes approximately the interactions between the charge-density distribution and Pauli deformation density is

$$R_{ij} = r_{ij}^{-1} - 2 \sum_{k \in A} r_{kj}^{-1} - 2 \sum_{l \in B} r_{il}^{-1} + \sum_{x \in A} r_{xj}^{-1} + \sum_{y \in B} r_{yi}^{-1} \quad (77)$$

Despite the repulsion interaction potential computed from the above EFP2 formulation was proven to be reasonably accurate, the more general formula from Eq. (73) has the advantage that it can be potentially extended to tackle electronic correlation. This is discussed in Section XXX.

### 2.2.3 Polarization

To compute polarization interaction energy from Eq. (48) in GEFP fashion, one needs to resolve the problem of computing the following class of integrals

$$\iint \frac{\rho^X(\mathbf{r}_1) \Delta \rho^{\text{Pol}}(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \quad (78)$$

where  $\rho^X(\mathbf{r}_1)$  is the arbitrary and known one-electron density,

$$\rho^X(\mathbf{r}_1) = 2 \sum_{\alpha\beta} D_{\alpha\beta}^X \phi_\alpha(\mathbf{r}_1) \phi_\beta(\mathbf{r}_1) \quad (79)$$

and

$$\Delta \rho^{\text{Pol}}(\mathbf{r}_1) = 2 \sum_{\alpha\beta} \delta D_{\alpha\beta} \phi_\alpha(\mathbf{r}_1) \phi_\beta(\mathbf{r}_1) \quad (80)$$

both expressed in terms of the associated one-particle density matrices,  $\mathbf{D}^X$  and  $\delta \mathbf{D}$ , respectively. Therefore, one needs to develop a methodology of predicting the density matrix change due to polarization process. This is a challenging task.

In this Proposal, I propose the model for prediction of the polarization-induced electron density change of structurally unrelaxed molecule. The associated density matrix is deconvoluted into a product form of the polarization susceptibility and induced dipole moments due to electric fields at the centroids of the occupied molecular orbitals,

$$\delta \mathbf{D} \propto \sum_i^{\text{occ}} [\text{polarization susceptibility of the } i\text{th MO}] \cdot \delta \boldsymbol{\mu}_i \quad (81)$$

While the latter contain critical information about the polarization of the electronic cloud, polarization susceptibility is a property of unperturbed molecule and can be precomputed once and for all. This dramatically simplifies the calculation of the exchange-induction interaction energies and can be adapted into an efficient effective fragment potential.

**Wavefunction Polarization: The Vector Model.** The total electronic dipole moment of a polarized closed-shell molecule can be expressed as

$$\boldsymbol{\mu}(\mathbf{r}_0) = -\text{Tr} [\mathbf{P}\mathbf{M}(\mathbf{r}_0)] = -2 \text{Tr} [\mathbf{D}\mathbf{M}(\mathbf{r}_0)] \quad (82)$$

where  $\mathbf{D}$  is a one-particle density matrix

$$\mathbf{D} = \mathbf{C}\mathbf{C}^\dagger \quad (83)$$

$\mathbf{C}$  is the LCAO-MO matrix constructed from orthogonal AOs and  $\mathbf{M}(\mathbf{r}_0)$  is certain representation of the dipole moment operator (defined with respect to origin at  $\mathbf{r}_0$ ),

$$[\mathbf{M}(\mathbf{r}_0)]_{\alpha\beta} = \langle \alpha | \mathbf{r} - \mathbf{r}_0 | \beta \rangle \quad (84)$$

in basis set denoted by Greek indices. If I work in unchanging basis (for example, AOs) I can write

$$\boldsymbol{\mu}(\mathbf{r}_0) = -2 \text{Tr} [\{\mathbf{D} + \delta\mathbf{D}\} \mathbf{M}(\mathbf{r}_0)] \quad (85)$$

or

$$\delta\boldsymbol{\mu}(\mathbf{r}_0) = -2 \text{Tr} [\delta\mathbf{D}\mathbf{M}(\mathbf{r}_0)] \quad (86)$$

which is the induced total dipole moment. The change in the molecular orbitals is parameterized as suggested by Haynes as

$$\delta\mathbf{C} = \boldsymbol{\Xi}\mathbf{C} \quad (87)$$

where  $[\mathbf{C}]_{\alpha i} = C_{\alpha i}$  is the LCAO-MO coefficient of unperturbed wave function and  $\boldsymbol{\Xi}_{\alpha\beta}$  is the square non-singular matrix of size  $m \times m$  with  $m$  denoting the number of AOs. This matrix is in general non-symmetric. If  $\boldsymbol{\Xi}$  is known, the change in the one-particle reduced density matrix can be found by considering idempotency restraint. The result can be written as

$$\delta\mathbf{D} = -\mathbf{D} + [\mathbf{D} + \mathbf{v}] [\mathbf{1} + \mathbf{v}^\dagger \mathbf{v}]^{-1} [\mathbf{D} + \mathbf{v}^\dagger] \quad (88)$$

where the auxiliary vector  $\mathbf{v}$  is defined as

$$\mathbf{v} \equiv [\mathbf{1} - \mathbf{D}] \boldsymbol{\Xi} \mathbf{D} \quad (89)$$

Eq. (88) can be approximated by Taylor expansion around  $\mathbf{v} = \mathbf{0}$  which gives

$$\delta\mathbf{D} = [\mathbf{v} + \mathbf{v}^\dagger] + [\mathbf{v}\mathbf{v}^\dagger - \mathbf{v}^\dagger\mathbf{v}] + \dots \quad (90)$$

Truncating the expansion from Eq. (90) at linear terms with respect to  $\mathbf{v}$  and inserting this into Eq. (86) I obtain

$$-\frac{1}{2}\delta\boldsymbol{\mu} \cong \text{Tr} [\mathbf{M}\boldsymbol{\Xi}\mathbf{D}] + \text{Tr} [\mathbf{M}\mathbf{D}\boldsymbol{\Xi}^\dagger] - \text{Tr} [\mathbf{M}\mathbf{D}\boldsymbol{\Xi}\mathbf{D}] - \text{Tr} [\mathbf{M}\mathbf{D}\boldsymbol{\Xi}^\dagger\mathbf{D}] \quad (91)$$

where I omitted the dipole origin for notational simplicity. I rewrite Eq. (91) by realising that

$$\text{Tr} [\mathbf{M}\boldsymbol{\Xi}\mathbf{D}] + \text{Tr} [\mathbf{M}\mathbf{D}\boldsymbol{\Xi}^\dagger] = 2 \text{Tr} [\widetilde{\mathbf{M}}\widetilde{\boldsymbol{\Xi}}] \quad (92a)$$

$$\text{Tr} [\mathbf{M}\mathbf{D}\boldsymbol{\Xi}\mathbf{D}] + \text{Tr} [\mathbf{M}\mathbf{D}\boldsymbol{\Xi}^\dagger\mathbf{D}] = 2 \text{Tr} [\widetilde{\mathbf{A}}\widetilde{\boldsymbol{\Xi}}] \quad (92b)$$

for

$$\bar{\Xi} \equiv \Xi \mathbf{C} \quad (93a)$$

$$\tilde{\mathbf{M}} \equiv \mathbf{C}^\dagger \mathbf{M} \quad (93b)$$

$$\tilde{\mathbf{A}} \equiv \mathbf{C}^\dagger \mathbf{M} \mathbf{D} \quad (93c)$$

This leads to the following equation

$$-\frac{1}{4} \delta \boldsymbol{\mu}(\mathbf{r}_0) = \text{Tr} \left[ \left\{ \underbrace{\tilde{\mathbf{M}} - \tilde{\mathbf{A}}}_{\tilde{\mathbf{L}}} \right\} \bar{\Xi} \right] \equiv \sum_i^{\text{occ}} \sum_\alpha \tilde{\mathbf{L}}_{i\alpha} \bar{\Xi}_{\alpha i} \quad (94)$$

In this way, I obtained the partitioning of the total induced dipole moment into separate contributions associated with each occupied MO, which is *exact* up to first order in  $\Xi$ . The quantity  $\tilde{\mathbf{L}}$  contains the projector onto unoccupied MO space,  $\mathbf{1} - \mathbf{D}$ , from which the new (perturbed) orbitals are formed. The distributed moment is

$$-\frac{1}{4} \delta \boldsymbol{\mu}_i = \sum_\alpha \tilde{\mathbf{L}}_{i\alpha} \bar{\Xi}_{\alpha i} \quad (95)$$

The above equation can be re-written in a matrix form as

$$-\frac{1}{4} \boldsymbol{\mu}_i^T(\mathbf{r}_0) = \bar{\Xi}_i^T \mathbf{L}_i \quad (96)$$

where I defined the vector

$$[\bar{\Xi}_i]_\alpha = \sum_\beta \Xi_{\alpha\beta} C_{\beta i} \quad (97)$$

of length  $m$ , matrix

$$[\mathbf{L}_i]_\alpha = \sum_\beta C_{\beta i} \mathbf{M}_{\beta\alpha} \quad (98)$$

of size  $m \times 3$ , and the vector of matrices

$$[\mathbf{L}]_{\alpha\zeta\beta} = L_{\beta\alpha}^{(\zeta)} \quad (99)$$

of size  $m \times 3 \times m$ . Note that, in this notation,  $\boldsymbol{\mu}_i^T$  and  $\bar{\Xi}_i^T$  are row vectors whereas  $\bar{\Xi}_i^p$  is a column vector. The solution of Eq. (96) is

$$\bar{\Xi}_i^T = -\frac{1}{4} \boldsymbol{\mu}_i^T [\tilde{\mathbf{L}}_i]_{\text{Left}}^{-1} \quad (100)$$

where  $[\tilde{\mathbf{L}}_i]_{\text{Left}}^{-1}$  is a *left inverse* of  $\tilde{\mathbf{L}}_i$  matrix and is given as

$$[\tilde{\mathbf{L}}_i]_{\text{Left}}^{-1} \equiv [\tilde{\mathbf{L}}_i^T \tilde{\mathbf{L}}_i]^{-1} \tilde{\mathbf{L}}_i^T \quad (101)$$

and has size  $3 \times m$ . Note that the *right inverse* of  $\tilde{\mathbf{L}}_i$  does not exist because  $\tilde{\mathbf{L}}_i \tilde{\mathbf{L}}_i^T$  is singular.

**Density Matrix Change.** Now I can obtain the final expression for the density matrix change due to external electrostatic perturbation. This can be written in terms of  $\Xi$  which is solved in Eq. (100). That is,

$$\delta \mathbf{D} \cong [\mathbf{1} - \mathbf{D}] \Xi \mathbf{D} + \mathbf{D} \Xi^T [\mathbf{1} - \mathbf{D}] = \Xi \mathbf{C}^T + \mathbf{C} \Xi^T - \mathbf{D} \Xi \mathbf{C}^T - \mathbf{C} \Xi^T \mathbf{D} \quad (102)$$

which is correct up to first order in  $\Xi$ . This result can be recast as

$$\delta D_{\alpha\beta} = -\frac{1}{4} \sum_i^{\text{occ}} \mathbf{F}_i \cdot \mathbf{B}_{\alpha\beta}^{(i)} \quad (103)$$

where the **density polarization susceptibility tensor** is

$$\mathbf{B}_{\alpha\beta}^{(i)} = \alpha_i \cdot \left\{ C_{\alpha i} \tilde{\mathbf{L}}_{i\beta} + C_{\beta i} \tilde{\mathbf{L}}_{i\alpha} - \sum_{\gamma} (D_{\alpha\gamma} C_{\beta i} + D_{\beta\gamma} C_{\alpha i}) \tilde{\mathbf{L}}_{i\gamma} \right\} \quad (104)$$

I mark that all elements  $\mathbf{B}_{\alpha\beta}^{(i)}$  are unique properties of electronically unperturbed molecule. This makes them perfect candidates for general effective fragment parameters.

**Origin-Dependence of Induced Dipole Moment.** I need to be aware that multipole integrals depend on the origin with respect to which they are evaluated. The induced dipole moments defined in Eq. (95) have to be origin-independent. I can find that

$$[\tilde{\mathbf{M}}]_{i\alpha}(\mathbf{r}_0) = [\tilde{\mathbf{M}}]_{i\alpha}(\mathbf{0}) - \mathbf{r}_0 [\mathbf{C}^T]_{i\alpha} \quad (105)$$

Also I have

$$[\tilde{\mathbf{A}}]_{i\alpha}(\mathbf{r}_0) = [\tilde{\mathbf{A}}]_{i\alpha}(\mathbf{0}) - \mathbf{r}_0 [\mathbf{C}^T \mathbf{D}]_{i\alpha} \quad (106)$$

But  $\mathbf{C}^T \mathbf{D} = \mathbf{C}^T \mathbf{C} \mathbf{C}^T = \mathbf{C}^T$  which means that

$$[\tilde{\mathbf{L}}]_{i\alpha}(\mathbf{r}_0) = [\tilde{\mathbf{M}}]_{i\alpha}(\mathbf{r}_0) - [\tilde{\mathbf{A}}]_{i\alpha}(\mathbf{r}_0) = [\tilde{\mathbf{M}}]_{i\alpha}(\mathbf{0}) - [\tilde{\mathbf{A}}]_{i\alpha}(\mathbf{0}) = [\tilde{\mathbf{L}}]_{i\alpha}(\mathbf{0}) \quad (107)$$

Therefore, it is clear that the polarization-induced distributed dipole moments defined in Eq. (95) are origin-independent. Thus, one can just compute dipole integrals with respect to any origin and resulting susceptibilities  $\mathbf{B}_{\alpha\beta}^{(i)}$  from Eq. (104) will be uniquely defined.

**Polarization Energy from Predicted Density Matrix Change.** The total polarization deformation density can be expressed as the sum of the polarization deformation densities of each monomer and the charge-transfer term. Mainly,

$$\Delta \rho^{\text{Pol}}(\mathbf{r}) = \Delta \rho^{\text{Pol},A}(\mathbf{r}) + \Delta \rho^{\text{Pol},B}(\mathbf{r}) + \Delta \rho^{\text{Pol},AB}(\mathbf{r}) \quad (108)$$

In this Proposal, I neglect charge transfer and I leave the analysis of the problem for the future. However, in most cases of typical intermolecular interactions, charge transfer energies are the least important contributions. Therefore, we assume that the following approximation

$$\Delta \rho^{\text{Pol}}(\mathbf{r}) \cong \Delta \rho^{\text{Pol},A}(\mathbf{r}) + \Delta \rho^{\text{Pol},B}(\mathbf{r}) \quad (109)$$

is acceptable at that moment.

Using the deformation susceptibility tensor derived above we can now drastically simplify Eq. (78) as follows

$$\begin{aligned} \iint \frac{\rho^X(\mathbf{r}_1)\Delta\rho^{\text{Pol}}(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 &= 4 \sum_{\alpha\beta} \sum_{\mu\nu \in A} D_{\alpha\beta}^X \delta D_{\mu\nu}^A (\alpha\beta|\mu\nu) \\ &= - \sum_{\alpha\beta} \sum_{\mu\nu \in A} \sum_{i \in A}^{\text{occ}} D_{\alpha\beta}^X \mathbf{B}_{\mu\nu}^{(i)} \cdot \mathbf{F}_i (\alpha\beta|\mu\nu) \equiv \sum_{\alpha\beta} D_{\alpha\beta}^X \sum_{i \in A}^{\text{occ}} (\alpha|\mathbf{v}_{\text{Pol}}^{(i)}|\beta) \cdot \mathbf{F}_i \end{aligned} \quad (110)$$

where the new effective one-electron potential is defined without loss of generality as

$$\mathbf{v}_{\text{Pol}}^{(i)}(\mathbf{r}) \equiv - \sum_{\mu\nu \in A} \left[ \int \frac{\phi_\mu(\mathbf{r}')\phi_\nu(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \right] \mathbf{B}_{\mu\nu}^{(i)} \quad (111)$$

Analogous potential can be defined for molecule  $B$ . Such potentials, which are associated with each occupied orbital, can be either expanded in an auxiliary basis set with expansion coefficients determined for isolated monomer, or even fit to the set of effective charges. These two kinds of generalized effective fragment parameters could be calculated only once and for all.

Now, I can derive the working formulae for the polarization interaction energy. First, let us consider the semi-classical induction energy that is given by the sum  $E_{\text{nuc-Pol}} + E_{\text{el-Pol}}$  in which

$$E_{\text{nuc-Pol}} = \iint \frac{[\rho^{\text{nuc},A}(\mathbf{r}_1) + \rho^{\text{nuc},B}(\mathbf{r}_1)] [\Delta\rho^{\text{Pol},A}(\mathbf{r}_2) + \Delta\rho^{\text{Pol},B}(\mathbf{r}_2)]}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \quad (112)$$

and

$$E_{\text{el-Pol}} = \iint \frac{[\rho^{\text{el},A}(\mathbf{r}_1) + \rho^{\text{el},B}(\mathbf{r}_1)] [\Delta\rho^{\text{Pol},A}(\mathbf{r}_2) + \Delta\rho^{\text{Pol},B}(\mathbf{r}_2)]}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \quad (113)$$

Each of the above two equations will give four different terms. It is beneficial to consider the  $AA$  and  $BB$ -type contributions (or ‘self-interactions’) in different way than the ‘mixed’ terms, and treat the charge density as a whole, not dividing it into nuclei and electrons. Thus, by noting that  $\rho^A(\mathbf{r}_1) = \rho^{\text{nuc},A}(\mathbf{r}_1) + \rho^{\text{el},A}(\mathbf{r}_1)$  I have

$$\begin{aligned} E_{\text{nuc-Pol}}^{AA} &= \iint \frac{\rho^A(\mathbf{r}_1)\Delta\rho^{\text{Pol},A}(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 = \int d\mathbf{r}_2 \hat{v}_A(\mathbf{r}_1) \Delta\rho^{\text{Pol},A}(\mathbf{r}_2) \\ &= -\frac{1}{4} \sum_{i \in A}^{\text{occ}} \sum_{\mu\nu \in A} \mathbf{B}_{\mu\nu}^{(i)} \cdot \mathbf{F}_i (\mu|\hat{v}_A|\nu) \equiv \sum_{i \in A}^{\text{occ}} \mathbf{b}^{(i)} \cdot \mathbf{F}_i \end{aligned} \quad (114)$$

where I define the new effective fragment to be

$$\mathbf{b}^{(i)} \equiv -\frac{1}{4} \sum_{\mu\nu \in A} (\mu|\hat{v}_A|\nu) \mathbf{B}_{\mu\nu}^{(i)} \quad (115)$$

Note that the vector  $\mathbf{b}^{(i)}$  has just length of  $N_{\text{occ}}$  and the associated cost of evaluation of  $E_{\text{nuc-Pol}}^{AA}$  is negligible once  $\mathbf{b}^{(i)}$  had been computed and stored.

Since the same can be done for molecule  $B$  self-energy, I am left for the remaining four terms:

$$\begin{aligned} \iint \frac{\rho^{\text{nuc},A}(\mathbf{r}_1)\Delta\rho^{\text{Pol},B}(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}d\mathbf{r}_1d\mathbf{r}_2 & \quad \iint \frac{\rho^{\text{el},A}(\mathbf{r}_1)\Delta\rho^{\text{Pol},B}(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}d\mathbf{r}_1d\mathbf{r}_2 \\ \iint \frac{\rho^{\text{nuc},B}(\mathbf{r}_1)\Delta\rho^{\text{Pol},A}(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}d\mathbf{r}_1d\mathbf{r}_2 & \quad \iint \frac{\rho^{\text{el},B}(\mathbf{r}_1)\Delta\rho^{\text{Pol},A}(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}d\mathbf{r}_1d\mathbf{r}_2 \end{aligned} \quad (116)$$

but only the integrals in the upper row have distinct functional forms. Hence, I consider explicitly only these integrals here. First I have

$$\begin{aligned} \iint \frac{\rho^{\text{nuc},A}(\mathbf{r}_1)\Delta\rho^{\text{Pol},B}(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}d\mathbf{r}_1d\mathbf{r}_2 \\ = 2 \int d\mathbf{r}_1 \int d\mathbf{r}_2 \sum_{x \in A} \sum_{\lambda \sigma \in B} Z_x \delta(\mathbf{r}_x - \mathbf{r}_1) \delta D_{\lambda\sigma}^B \frac{\phi_\lambda(\mathbf{r}_2)\phi_\sigma(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|} \end{aligned} \quad (117)$$

Further I have

$$\begin{aligned} \iint \frac{\rho^{\text{nuc},A}(\mathbf{r}_1)\Delta\rho^{\text{Pol},B}(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}d\mathbf{r}_1d\mathbf{r}_2 \\ = 2 \int d\mathbf{r}_2 \sum_{x \in A} \sum_{\lambda \sigma \in B} \delta D_{\lambda\sigma}^B \frac{\phi_\lambda(\mathbf{r}_2)\phi_\sigma(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|} Z_x = -\frac{1}{2} \sum_{\lambda \sigma \in B} \left[ \sum_{j \in B}^{\text{occ}} \mathbf{B}_{\lambda\sigma}^{(j)} \cdot \mathbf{F}_j \right] \left[ \sum_{x \in A} V_{\lambda\sigma}^{(x)} \right] \end{aligned} \quad (118)$$

where  $V_{\lambda\sigma}^{(x)}$  is the potential one-electron integral evaluated for the  $x$ th atomic centre of  $A$ . It is expected that the computational cost of the final result in Eq. (118) is negligible. The integral associated with the interaction of electronic density can be obtained readily by using the result from Eq. (110) and realizing that the summations over AOs belong only to either molecule,

$$\iint \frac{\rho^{\text{el},A}(\mathbf{r}_1)\Delta\rho^{\text{Pol},B}(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}d\mathbf{r}_1d\mathbf{r}_2 = \sum_{\mu\nu \in A} D_{\mu\nu}^A \sum_{j \in B}^{\text{occ}} (\mu|\mathbf{v}_{\text{Pol}}^{(j)}|\nu) \cdot \mathbf{F}_j \quad (119)$$

This is the most time consuming part of pseudo-classical induction energy evaluation within proposed GEFP. One can see that the costs of such calculations are small.

Now let us consider the contribution

$$E_{\text{Pol-Pol}} = \frac{1}{2} \iint \frac{\Delta\rho^{\text{Pol}}(\mathbf{r}_1)\Delta\rho^{\text{Pol}}(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}d\mathbf{r}_1d\mathbf{r}_2 \quad (120)$$

This also contains two self-interaction terms and two ‘mixing’ terms. It is straightforward to notice that in the former case I can define new effective potential

$$E_{\text{Pol-Pol}}^{AA} = \frac{1}{2} \iint \frac{\Delta\rho^{\text{Pol},A}(\mathbf{r}_1)\Delta\rho^{\text{Pol},A}(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}d\mathbf{r}_1d\mathbf{r}_2 \equiv \sum_{ii' \in A}^{\text{occ}} \mathbf{B}^{(ii')} : \mathbf{F}_i \otimes \mathbf{F}_{i'} \quad (121)$$

with

$$\boxed{\mathbf{B}^{(ii')} \equiv \frac{1}{8} \sum_{\mu\nu\mu'\nu' \in A} \mathbf{B}_{\mu\nu}^{(i)} \otimes \mathbf{B}_{\mu'\nu'}^{(i')} (\mu\nu|\mu'\nu')} \quad (122)$$

To derive the ‘mixed’ terms I use Eq. (110) and I obtain

$$\frac{1}{2} \iint \frac{\Delta\rho^{\text{Pol},A}(\mathbf{r}_1)\Delta\rho^{\text{Pol},B}(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}d\mathbf{r}_1d\mathbf{r}_2 \equiv \frac{1}{2} \sum_{\mu\nu \in A} \sum_{j \in B}^{\text{occ}} \delta D_{\mu\nu}^A (\mu|\mathbf{v}_{\text{Pol}}^{(j)}|\nu) \cdot \mathbf{F}_j \quad (123)$$

Another contribution to polarization energy is associated with the change in kinetic energy contribution to the overall polarization energy

$$E_{\text{Pol}}^{\nabla^2} = -\frac{1}{2} \int d\mathbf{r}_1 \nabla^2 [\Delta\rho_{\text{Pol}}(\mathbf{r}_1, \mathbf{r}'_1)]_{\mathbf{r}'_1=\mathbf{r}_1} \quad (124)$$

It is easy to see that

$$E_{\text{Pol}}^{\nabla^2} = 2 \sum_{\mu\nu \in A} \delta D_{\mu\nu}^A T_{\mu\nu} + 2 \sum_{\lambda\sigma \in B} \delta D_{\lambda\sigma}^B T_{\lambda\sigma} \quad (125)$$

evaluation of which is expected to be very efficient.

The most problematic part of polarization energy is due to the interaction with Pauli deformation density and due to exchange effects. Let us first elaborate on the first, that is

$$E_{\text{Pol-Pauli}} = \iint \frac{\Delta\rho^{\text{Pol}}(\mathbf{r}_1) \Delta\rho^{\text{Pauli}}(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \quad (126)$$

Here, I assume that truncating the series in Eq. (61) already at the first order term is acceptable since the extent of  $\Delta\rho^{\text{Pol}}(\mathbf{r}_1)$  is small as compared to  $\rho(\mathbf{r}_1)$ . Therefore, I can write for the contribution due to  $\Delta\rho^{\text{Pol},A}(\mathbf{r}_1)$  that

$$E_{\text{Pol-Pauli}}^A \cong -4 \sum_{i \in A} \sum_{j \in B} \sum_{\mu\nu \in A} \sum_{\kappa \in A} \sum_{\lambda \in B} \delta D_{\mu\nu}^A (\mu\nu|\kappa\lambda) C_{\kappa i}^A C_{\lambda j}^B \quad (127)$$

which is correct up to first order with respect to overlap matrix between MOs. By applying expansion of  $\delta D_{\mu\nu}^A$  in terms of the deformation susceptibilities I find the convenient to handle expression

$$E_{\text{Pol-Pauli}}^A \cong - \sum_{ij} S_{ij} \sum_{k \in A} (i|\mathbf{v}_{\text{Pol}}^{(k)}|j) \quad (128)$$

In summary, all the above contributions to the polarization interaction energy are gathered in the following expression

$$\begin{aligned} E_{\text{Pol}} = & \underbrace{\sum_{i \in A} \left[ \mathbf{b}_i + \sum_{k \in A} \mathbf{B}_{ik} \cdot \mathbf{F}_k \right] \cdot \mathbf{F}_i + \sum_{j \in B} \left[ \mathbf{b}_j + \sum_{l \in B} \mathbf{B}_{jl} \cdot \mathbf{F}_l \right] \cdot \mathbf{F}_j}_{\text{self-interaction contributions}} \\ & + \underbrace{\text{Tr} \left[ \delta \mathbf{D}^A \left\{ 2\mathbf{T}^A + 2\mathbf{V}_{AA}^{B,\text{nuc}} + \frac{1}{2} \bar{\mathbf{V}}_{AA}^{\text{Pol}} \right\} \right] + \text{Tr} \left[ \delta \mathbf{D}^B \left\{ 2\mathbf{T}^B + 2\mathbf{V}_{BB}^{A,\text{nuc}} + \frac{1}{2} \bar{\mathbf{V}}_{BB}^{\text{Pol}} \right\} \right]}_{\text{Pol-nuc, Pol-Pol and Pol-}\nabla^2 \text{ contributions}} \\ & + \underbrace{\text{Tr} \left[ \mathbf{D}^A \bar{\mathbf{V}}_{AA}^{\text{Pol}} \right] + \text{Tr} \left[ \mathbf{D}^B \bar{\mathbf{V}}_{BB}^{\text{Pol}} \right]}_{\text{Pol-el contributions}} - \underbrace{\text{Tr} \left[ \mathbf{S}^{AB} \bar{\mathbf{V}}_{BA}^{\text{Pol}} \right] - \text{Tr} \left[ \mathbf{S}^{BA} \bar{\mathbf{V}}_{AB}^{\text{Pol}} \right]}_{\text{Pol-Pauli contributions - 1st order}} \quad (129) \end{aligned}$$

where we introduced auxiliary definitions:

$$[\bar{\mathbf{V}}_{AA}^{\text{Pol}}]_{ii'} = \sum_j \mathbf{V}_{AA}^{(j),\text{Pol}} \cdot \mathbf{F}_j \quad (130a)$$

$$[\bar{\mathbf{V}}_{BB}^{\text{Pol}}]_{jj'} = \sum_i \mathbf{V}_{BB}^{(i),\text{Pol}} \cdot \mathbf{F}_i \quad (130b)$$

$$[\bar{\mathbf{V}}_{AB}^{\text{Pol}}]_{ij} = \sum_l \mathbf{V}_{AB}^{(l),\text{Pol}} \cdot \mathbf{F}_l \quad (130c)$$

$$[\bar{\mathbf{V}}_{BA}^{\text{Pol}}]_{ji} = \sum_k \mathbf{V}_{BA}^{(k),\text{Pol}} \cdot \mathbf{F}_k \quad (130d)$$



It is clear that the evaluation cost of the self-interaction contributions in Eq. (129) is negligible. Thus, the bottleneck of the computation of  $E_{\text{Pol}}$  is the construction of the polarization 1-electron matrices and evaluation of a few traces. I believe that the above expression is superior to the polarization energy calculation in the conventional EFP2 method, based on the distributed dipole-dipole expansion with semi-empirical damping parameters to account for charge-penetration effects. Expression if Eq. (129) is free from multipole expansion and charge-penetration effects are inherently present in the formulation. Moreover, the overlap effects as well as electronic kinetic energy contributions are included from first principles. The computational cost is however greater of the proposed GEFP formulation. This is because the induced dipole moments are still necessary to compute the electronic density change due to polarization (note that the energy itself is not computed from multipole expansion). Thus, the overall cost of the method is similar in order to the cost of computation of the Pauli repulsion energy  $E_{\text{Rep}}$  from Eq. (73).

### 2.2.4 Exchange

Exchange effects due to Pauli exclusion principle can be estimated at SCF level as described by Mandado and Hermida-Ramon earlier. [2] In the first-order, exchange energy is equal to

$$E_X = -2 \sum_{ij} (i^{oo} j^{oo} | j^{oo} i^{oo}) + 2 \sum_{ii'} (ii' | i'i) + 2 \sum_{jj'} (jj' | j'j) \quad (131)$$

It was however showed that the Jensen's formula based on the spherical Gaussian overlap (SGO) [10]

$$E_{\text{ex}} \cong 2 \sum_{ij} \sqrt{\frac{-2 \ln |S_{ij}|}{\pi}} \frac{S_{ij}^2}{r_{ij}} \quad (132)$$

is quite accurate for closed shell molecules with errors smaller than 0.1 kcal/mole. In this approximation the MO's are assumed to be localized.

More challenging is the evaluation of the exchange energy contribution to the electron polarization,

$$E_{X,\text{Pol}} = \frac{1}{2} \iint \frac{\Delta \rho_X^{\text{Pol}}(\mathbf{r}_1, \mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \quad (133)$$

because it requires the evaluation of the change in the exchange density due to the mutual polarization of the electronic clouds of  $A$  and  $B$ ,

$$\Delta \rho_X^{\text{Pol}}(\mathbf{r}_1, \mathbf{r}_2) = \rho_X^{\text{polar}}(\mathbf{r}_1, \mathbf{r}_2) - \rho_X^{\text{unpol}}(\mathbf{r}_1, \mathbf{r}_2) \quad (134)$$

Thus,

$$E_{X,\text{Pol}} = -2 \sum_{ij} \left[ \left( i_{\text{polar}}^{oo} j_{\text{polar}}^{oo} | j_{\text{polar}}^{oo} i_{\text{polar}}^{oo} \right) - (i^{oo} j^{oo} | j^{oo} i^{oo}) \right] \quad (135)$$

where the label 'polar' was used to mark that the polarization process is taken into account. By using the above introduced density matrix polarization model we can express the molecular 1-electron density matrices in terms of the electric field. Thus, in order to determine the orthogonalized set of polarized orbitals  $\{\chi_{i,\text{polar}}^{oo}\}$  the following procedure has to be performed: (i) the non-orthogonal polarized molecular orbitals  $\{\chi_{i,\text{polar}}\}$  have to be obtained by diagonalization of the one-particle density matrices of each molecule; (ii) the weighted orthogonalization

procedure needs to be undertaken to obtain orthogonal polarized set of orbitals  $\{\chi_{i,\text{polar}}^{oo}\}$ . Now, we can localize the resulting orbitals using Pipek-Mezey procedure [11] which is very cheap (involves only one-electron integrals and converges fast). Also, we can localize the set of unpolarized orthogonal orbitals  $\{\chi_i^{oo}\}$ . Once these orbitals become localized, we can approximate Eq. (135) as

$$E_{X,\text{Pol}} \approx 2 \sum_{ij}^{\text{polar}} \sqrt{\frac{-2 \ln |S_{ij}^{oo}|}{\pi}} \frac{[S_{ij}^{oo}]^2}{r_{ij}^{oo}} - 2 \sum_{ij}^{\text{unpolar}} \sqrt{\frac{-2 \ln |S_{ij}^{oo}|}{\pi}} \frac{[S_{ij}^{oo}]^2}{r_{ij}^{oo}} \quad (136)$$

This formula is computationally cheap as compared to the expression involving ERI's that need to be computed for both sets of molecular orbitals: polarized and unpolarized. The bottlenecks here are the orthogonalization of the unperturbed HF orbitals as well as the diagonalization of the density matrices to obtain the polarized non-orthogonal HF orbitals.

To avoid additional orthogonalization of the unpolarized MO's the net exchange energy originating from the first-order and polarization contributions can be considered as a whole. We apply SGO approximation to the first term in Eq. (131) and the result is

$$E_X \approx 2 \sum_{ij}^{\text{unpolar}} \sqrt{\frac{-2 \ln |S_{ij}^{oo}|}{\pi}} \frac{[S_{ij}^{oo}]^2}{r_{ij}^{oo}} + 2 \sum_{ii'} (ii'|i'i) + 2 \sum_{jj'} (jj'|j'j) \quad (137)$$

This leads to the following expression

$$E_X + E_{X,\text{Pol}} \approx 2 \sum_{ij}^{\text{polar}} \sqrt{\frac{-2 \ln |S_{ij}^{oo}|}{\pi}} \frac{[S_{ij}^{oo}]^2}{r_{ij}^{oo}} + 2 \sum_{ii'} (ii'|i'i) + 2 \sum_{jj'} (jj'|j'j) \quad (138)$$

Note that the second and third terms in Eq. (138) are the intra-molecular exchange energies and can be computed just once and stored as two numbers. Therefore, only the first term in Eq. (138) requires evaluation on the fly.

### 2.2.5 Total Hartree–Fock Interaction Energy Expression

To summarize the above derivations of the electrostatic, Pauli-repulsive, polarization and exchange interaction energies, we can cast the total HF interaction energy within the GEFP model as follows

$$\begin{aligned} \Delta E_{\text{HF}} \approx & \text{Tr} \left[ \left( 4\mathbf{T}^{AB} - \mathbf{s}^{AB} - [\mathbf{s}^{BA}]^T - \bar{\mathbf{V}}_{AB}^{\text{Pol}} - [\bar{\mathbf{V}}_{BA}^{\text{Pol}}]^T \right) \mathbf{S}^{BA} \right] \\ & + \text{Tr} \left[ \left( 2\mathbf{F}^A + \mathbf{v}_{AA}^B \right) \mathbf{S}^{AB} \mathbf{S}^{BA} \right] + \text{Tr} \left[ \left( 2\mathbf{F}^B + \mathbf{v}_{BB}^A \right) \mathbf{S}^{BA} \mathbf{S}^{AB} \right] \\ & + \sum_{i \in A} \left[ \mathbf{b}_i + \sum_{k \in A} \mathbf{B}_{ik} \cdot \mathbf{F}_k \right] \cdot \mathbf{F}_i + \sum_{j \in B} \left[ \mathbf{b}_j + \sum_{l \in B} \mathbf{B}_{jl} \cdot \mathbf{F}_l \right] \cdot \mathbf{F}_j \\ & + \text{Tr} \left[ \delta \mathbf{D}^A \left\{ 2\mathbf{T}^A + 2\mathbf{V}_{AA}^{B,\text{nuc}} + \frac{1}{2} \bar{\mathbf{V}}_{AA}^{\text{Pol}} \right\} \right] + \text{Tr} \left[ \delta \mathbf{D}^B \left\{ 2\mathbf{T}^B + 2\mathbf{V}_{BB}^{A,\text{nuc}} + \frac{1}{2} \bar{\mathbf{V}}_{BB}^{\text{Pol}} \right\} \right] \\ & + \text{Tr} \left[ \mathbf{D}^A \left\{ \mathbf{v}_{AA}^B + \bar{\mathbf{V}}_{AA}^{\text{Pol}} \right\} \right] + \text{Tr} \left[ \mathbf{D}^B \left\{ \mathbf{v}_{BB}^A + \bar{\mathbf{V}}_{BB}^{\text{Pol}} \right\} \right] \\ & + 2 \sum_{ij}^{\text{polar}} \sqrt{\frac{-2 \ln |S_{ij}^{oo}|}{\pi}} \frac{[S_{ij}^{oo}]^2}{r_{ij}^{oo}} + 2 \sum_{ii'} (ii'|i'i) + 2 \sum_{jj'} (jj'|j'j) + \sum_{xy} \frac{Z_x Z_y}{|\mathbf{r}_x - \mathbf{r}_y|} \quad (139) \end{aligned}$$

Note that the above expression is completely free from direct evaluation of ERIs. The approximations used in the above formulation are listed here for clarity: (i) density matrix of the

system is expanded in Taylor series with respect to  $\Delta$  and truncated at quadratic terms for first-order Pauli-repulsion and linear terms for Pauli repulsion contribution to the polarization; (ii) interaction between Pauli deformation densities of two monomers is neglected; (iii) charge transfer process is neglected; (iv) SGO approximation is invoked to estimate exchange interaction energy; (v) resulting total interaction energy suffers from BSSE. These simplifications are believed to introduce small errors to the total HF energy of the  $AB$  complex and, given the reasonably fast performance of the code, it is expected that the resulting model is superior to EFP2.

## 2.3 Post-Hartree–Fock Intermolecular Interaction Energy Based on GEFP

The above discussed GEFP model at HF level is now a starting point for a construction of an analogous model based on the interaction between correlated wave functions, in which each determinant is expanded in terms of the natural molecular orbitals (NMO's). To address the problem of electron correlation one has to consider not only the orbitals, but also their occupation numbers which also change upon orthogonalization and polarization of electronic density distribution. In addition, dispersion interactions, a purely electron correlation effect, need to be addressed. Therefore, the following new features of the model need to be worked out in detail:

1. Extension of the Taylor expansion of the density matrix by inclusion of the fractionally occupied *virtual* orbitals and *occupied* orbitals with occupancies close to 1;
2. Extension of the vector polarization model by including (i) the change in the occupation numbers and (ii) the density deformation due to dispersion effects.

Once the above problems are solved, it is believed that the GEFP theory for a correlated system could have similar appearance than in Eq. (139) with a notable difference that the trace-performing summations would additionally include the virtual orbitals. The problems 1. and 2.(i) can be partially solved in the case of ground state Density Functional Theory (DFT) calculations because the electron correlation can be introduced semi-empirically at the DFT level and the resulting wave function has convenient single-determinantal functional form with singly occupied MO's. However, the problem with the dispersion-induced deformations and the interactions between electronically excited states cannot be easily resolved by using DFT and the more refined wave function-based approach needs to be developed.

In the following section I design the model that could be used to extend the Taylor expansion density matrix.

### 2.3.1 Pauli Deformation Density

By using the weighted orthogonalization procedure, the orthogonal molecular orbitals can be written as

$$\mathbf{C}^{oo} = \mathbf{C}\mathbf{W}[\mathbf{W}\mathbf{S}\mathbf{W}]^{-\frac{1}{2}} \quad (140)$$

where the weights form the diagonal matrix with entries  $W_{ii} = \sqrt{n_i}$  and  $n_i$  is the unperturbed occupation number of the  $i$ th MO. The resulting density matrix is then given by

$$\mathbf{D}^{oo} = \mathbf{C}^{oo}\mathbf{W}\mathbf{W}\mathbf{C}^{oo\dagger} = \mathbf{C}\mathbf{W}[\mathbf{W}\mathbf{S}\mathbf{W}]^{-\frac{1}{2}}\mathbf{W}\mathbf{W}[\mathbf{W}\mathbf{S}\mathbf{W}]^{-\frac{1}{2}}\mathbf{W}\mathbf{C}^\dagger \quad (141)$$

I note that I use the orthogonal AO basis here.<sup>†</sup> The fragment-based GEFP approach requires the above series expansion to be product-separable between molecules  $A$  and  $B$ . To achieve this I can Taylor expand the square root matrix function around  $\Delta$  as

$$[\mathbf{WSW}]^{-\frac{1}{2}} = \mathbf{1} - \frac{1}{2}\mathbf{W}^{-\frac{1}{2}}\Delta\mathbf{W}^{-\frac{1}{2}} + \frac{3}{8}\mathbf{W}^{-\frac{1}{2}}\Delta^2\mathbf{W}^{-\frac{1}{2}} - \dots \quad (142)$$

By inserting the above series truncated at the quadratic terms into Eq. (143) and subtracting from it the unperturbed density matrix, I obtain the deformation density matrix

$$\Delta\mathbf{D}^{\text{Pauli}} \cong \Delta\mathbf{D}^{\text{Pauli}}(\Delta) + \Delta\mathbf{D}^{\text{Pauli}}(\Delta^2) \quad (143)$$

where

$$\Delta\mathbf{D}^{\text{Pauli}}(\Delta) = -\frac{1}{2}\mathbf{C} \left[ \mathbf{W}^{\frac{1}{2}}\Delta\mathbf{W}^{\frac{5}{2}} + \mathbf{W}^{\frac{5}{2}}\Delta\mathbf{W}^{\frac{1}{2}} \right] \mathbf{C}^\dagger \quad (144a)$$

$$\Delta\mathbf{D}^{\text{Pauli}}(\Delta^2) = \frac{1}{8}\mathbf{C} \left[ 3\mathbf{W}^{\frac{1}{2}}\Delta^2\mathbf{W}^{\frac{5}{2}} + 3\mathbf{W}^{\frac{5}{2}}\Delta^2\mathbf{W}^{\frac{1}{2}} + 2\mathbf{W}^{\frac{1}{2}}\Delta\mathbf{W}\Delta\mathbf{W}^{\frac{1}{2}} \right] \mathbf{C}^\dagger \quad (144b)$$

Note that the radius of convergence of the series expansion from Eq. (142) can be strongly decreased when some of  $W_i$  are close to zero (virtual orbitals). Therefore, this series expansion needs to be tested and the appropriate procedure of determining the minimal orbital occupancy developed.

Assuming that the second-order expansion of the Pauli deformation density is accurate, I can now rewrite expressions for the Pauli repulsion energy in GEFP fashion. One has to notice the following identities

$$\Delta\mathbf{W}^m = \Delta^{BA}(\mathbf{W}^A)^m \oplus \Delta^{AB}(\mathbf{W}^B)^m \quad (145)$$

for  $m$  being a real number. Analogously I have

$$\Delta^2\mathbf{W}^m = \Delta^{AB}\Delta^{BA}\mathbf{W}^A \oplus \Delta^{BA}\Delta^{AB}\mathbf{W}^B \quad (146)$$

It is clear that the final expressions will be identical in structure as in Eq. (73) but will involve additionally the roots of orbital occupancies that multiply the overlap matrices in MO basis. This is straightforward but tedious task and, hence, the final formula will not be shown explicitly here.

Above I have shown that the Pauli repulsion energy between *any* correlated (ground and excited state) wave functions can be computed relatively easily from the unperturbed natural orbitals and occupancies, provided that the expansion in Eq. (142) is a reasonable approximation.

### 2.3.2 Polarization Effects

When the orbital occupancies are not equal to one, the 1-particle density matrix is no longer idempotent. However, it is *approximately* idempotent which means that for a relatively small electrostatic perturbations I could assume the vector model developed at HF level of theory. In other words, as a first approximation, it is to neglect the changes in the orbital occupancies and take into account only changes in the orbitals. If this assumption is reasonably accurate the expressions for polarization energy remain virtually identical with an exception of the Pauli repulsion contribution energy. In the latter case, the overlap matrix needs to be multiplied by roots of  $\mathbf{W}$  as described above and the summation here for the trace operation needs to include also virtual orbitals. It is however necessary to compute the distributed polarizabilities that are centered at the localized natural molecular orbital (NMOS) centroids.

<sup>†</sup>Otherwise the additional diagonalization of the product need to be performed to get the

### 2.3.3 Exchange Effects

To compute exchange-repulsion energy analogous to the first-order Heitler-London exchange-repulsion energy I need to evaluate the pure exchange energy. Since I have not developed yet the rigorous model of the polarization deformation for the correlated wave function, the pure exchange contribution can be computed as for now as presented by Mandado and Hermida-Ramon. [2] Mainly, the weighted orthogonalization procedure and the subsequent integration using ERI's is to be undertaken. To simplify these computations, I target to approximate the orthogonalized orbitals with the SGO and the resulting exchange energy is simply equal to:

$$E_{\text{exch}} \approx \sum_{ij}^{\text{unpolar}} \sqrt{\frac{-2n_i^{oo}n_j^{oo} \ln |S_{ij}^{oo}|}{\pi}} \frac{[S_{ij}^{oo}]^2}{r_{ij}^{oo}} + 2 \sum_{ii'} \sqrt{n_i n_{i'}} (ii'|i'i) + 2 \sum_{jj'} \sqrt{n_j n_{j'}} (jj'|j'j) \quad (147)$$

where the summation in the first term is performed over the *localized* orthogonal orbitals whereas the summations in the second and third terms are performed over localized unperturbed natural orbitals. As before, only the first term above needs to be evaluated on the fly because the latter two terms are constant numbers independent on the other monomer. However, here I cannot use simply Pipek-Mezey method [11] to localize orbitals and method of Reed and Weinhold called the *localized natural molecular orbitals* (NLMO's) [12] or similar one needs to be used here.

Exchange-correlation energy evaluation requires computation of the polarization deformation density at correlated level of theory. If the above discussed approximation based on the near-idempotency of the 1-particle density matrix can be used with a reasonable accuracy, then the exchange-correlation energy can be computed along with the first-order exchange energy from the following approximate formula

$$E_{\text{exch}} + E_{\text{XC}}^{\text{Pol}} \approx \sum_{ij}^{\text{polar}} \sqrt{\frac{-2n_i^{oo}n_j^{oo} \ln |S_{ij}^{oo}|}{\pi}} \frac{[S_{ij}^{oo}]^2}{r_{ij}^{oo}} + 2 \sum_{ii'} \sqrt{n_i n_{i'}} (ii'|i'i) + 2 \sum_{jj'} \sqrt{n_j n_{j'}} (jj'|j'j) \quad (148)$$

The polarized orthogonalized orbitals are obtained from (i) the polarized non-orthogonal NO's, (ii) their weighted orthogonalization to produce the polarized orthogonal orbitals, and (iii) a subsequent localization procedure.

### 2.3.4 Dispersion Effects

It is desirable to develop a dispersion deformation model analogous to the vector model for induction. I think it is possible to construct such a model from the fluctuating induced dipole moments computed from the frequency-dependent polarizabilities. The net deformation would be the expectation value of the total density deformation operator. However, as a first approximation dispersion interaction can be modeled in much simpler way by using Drude model developed before.

### 2.3.5 Summary

Here I showed that it is possible to design the formulae for the GEFP model that approximates the interaction energy between two arbitrary correlated wave functions. The following criteria

must be met: (i) Taylor expansion technique from Eq. (142) must be accurate; (ii) NLMO-distributed polarizabilities must be computable; (iii) upon electrostatic perturbation, 1-particle density matrix remains idempotent to a good degree of approximation; (iv) SGO approximation for exchange energy estimation must be reasonable; (v) dispersion effects can be modeled by a Drude model with a reasonable accuracy. If the above requirements are met, it is then possible to develop GEFP model that includes electrostatics, Pauli-repulsion, exchange, induction and dispersion interaction potentials and is not limited neither to HF level of theory nor electronic ground states. Therefore, in the next Section of this Proposal, I discuss the potential application of such GEFP model for a quantitative modelling of the electronic solvatochromism in condensed phases such as water.

### 3 Electronic Solvatochromism

Accurate calculations of the electronic excited state energies in condensed phases is a tremendous challenge because of prohibitively increasing costs of calculations. To properly describe many electronically excited states one needs to employ multireference wave functions and include dynamic correlation in some way. It is likely that the specific interactions with solute molecule change the shapes of the potential energy surfaces of electronic states. It is reasonable to assume that such effects cannot be described by continuum models of solvation. Therefore, GEFP model could be a good candidate in taking into account molecular granularity in an accurate and efficient way. It would enable computations for unprecedented before number of solvent molecules. Moreover, once the GEFP gradients are developed, it could be possible even to run molecular dynamics simulations! Non-adiabatic dynamics could be modelled as well by applying the generalized effective 1-electron potential technique introduced in this Proposal to quickly and accurately compute the inter-PES coupling constants.

Let us consider an isolated solute molecule ( $A$ ) in its  $n$ th electronic state with energy  $E_{A_n}^o$  and a solvent molecule ( $B$ ) in its ground state of energy  $E_B^o$ . The molecules are placed by a distance  $R$  apart from their centers of mass and their relative orientation is measured by  $\Omega$ . Here, we assume that the molecular internal structures are frozen and do not change with  $R$  and  $\Omega$ . Then, the total energy of such  $AB$  complex is

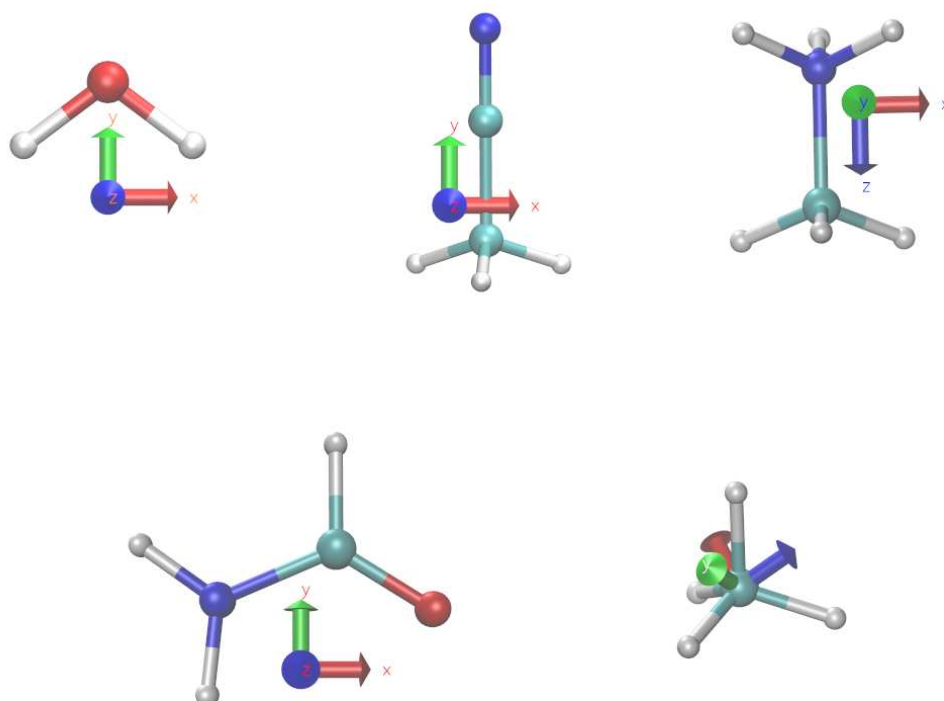
$$E_{A_nB}(R, \Omega) = E_{A_n}^o + E_B^o + \Delta E_{A_nB}(R, \Omega) \quad (149)$$

where  $\Delta E_{A_nB}$  defines the interaction energy between the solute's  $n$ th electronic state and the solvent's ground state. Therefore,  $E_{A_nB}(R, \Omega)$ , which could be estimated by GEFP, should be computed when one is interested in the electronic solvatochromism of solute. I believe that the GEFP method would allow to study bulk solutions and include hundreds of solvent molecule explicitly at a correlated level. Orthogonalization of NO's can be performed for more than two molecules if the non-additivity of exchange effects is of importance. Moreover, converging the induced dipoles can be done iteratively, which will take into account many body polarization effects on the polarization charge deformations. Therefore, I anticipate that the many body effects can be included in the treatment of exchange and polarization energies.

### 4 Polarization Density Model: Test Results

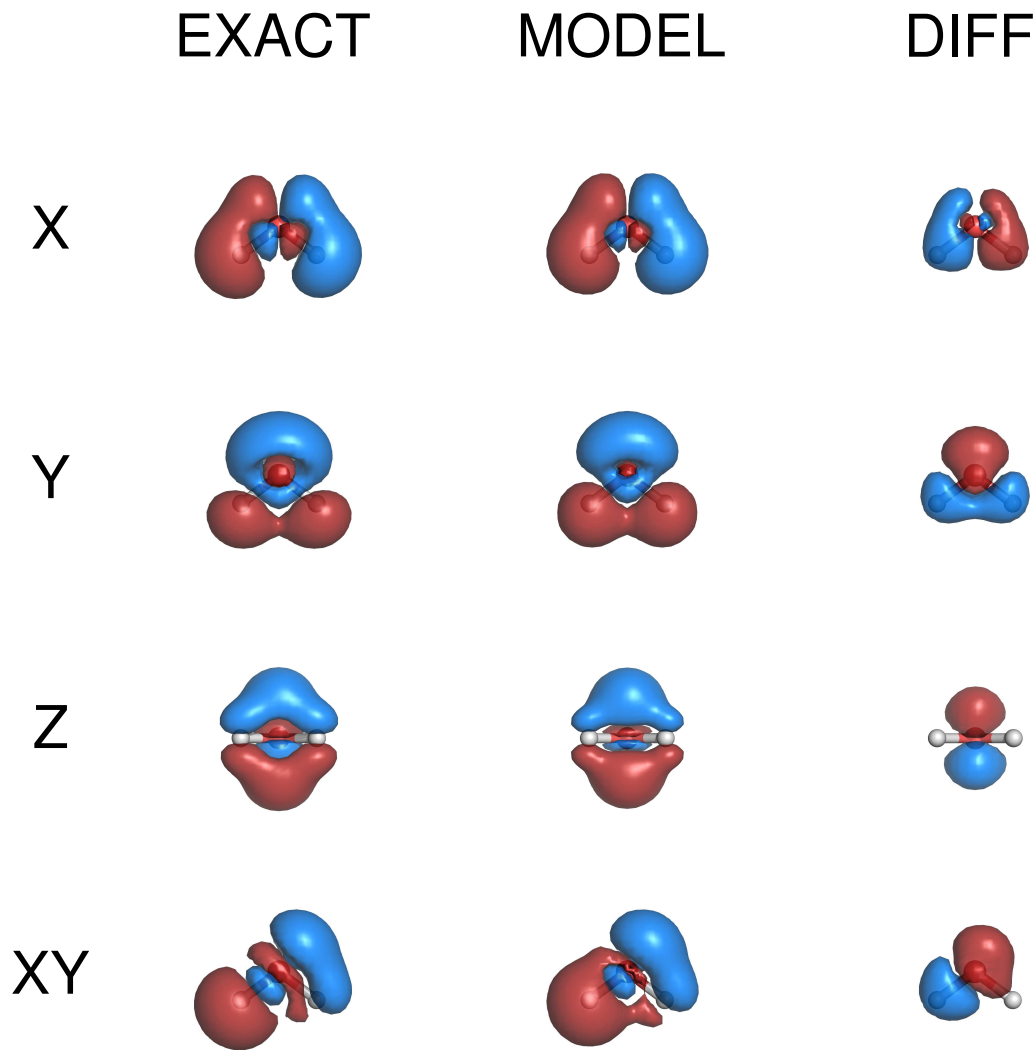
Here I test the vector model derived in Section 2.2.3. I chose five molecules:  $H_2O$ ,  $CH_3CN$ ,  $CH_3NH_3^+$ ,  $OHC-NH_2$  and  $CH_4$  that are placed in external uniform electric field applied in

various directions. The level of theory is assumed to be HF/6-311++G\*\* with Cartesian *d*-type basis functions. I assume the magnitude of electric field to be 0.015 au (77 MV/cm). Electric field of this order is exerted on molecules in aqueous environments. The molecular structures and their orientations with respect to the coordinate system are shown in Figure 4. See Figures for the deformation density plots.

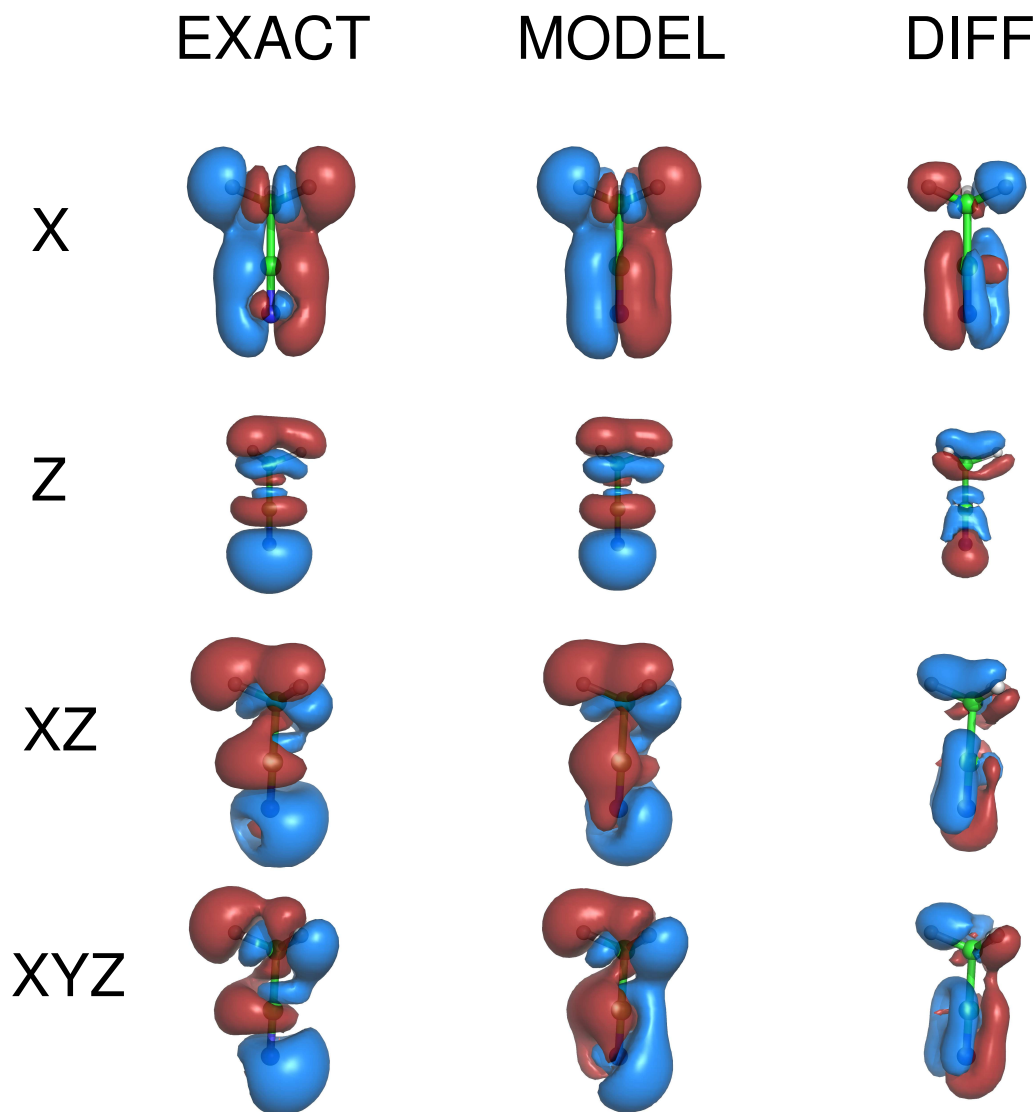


**Figure 4:** Molecules used for testing of density polarization model (Eq. (103)).

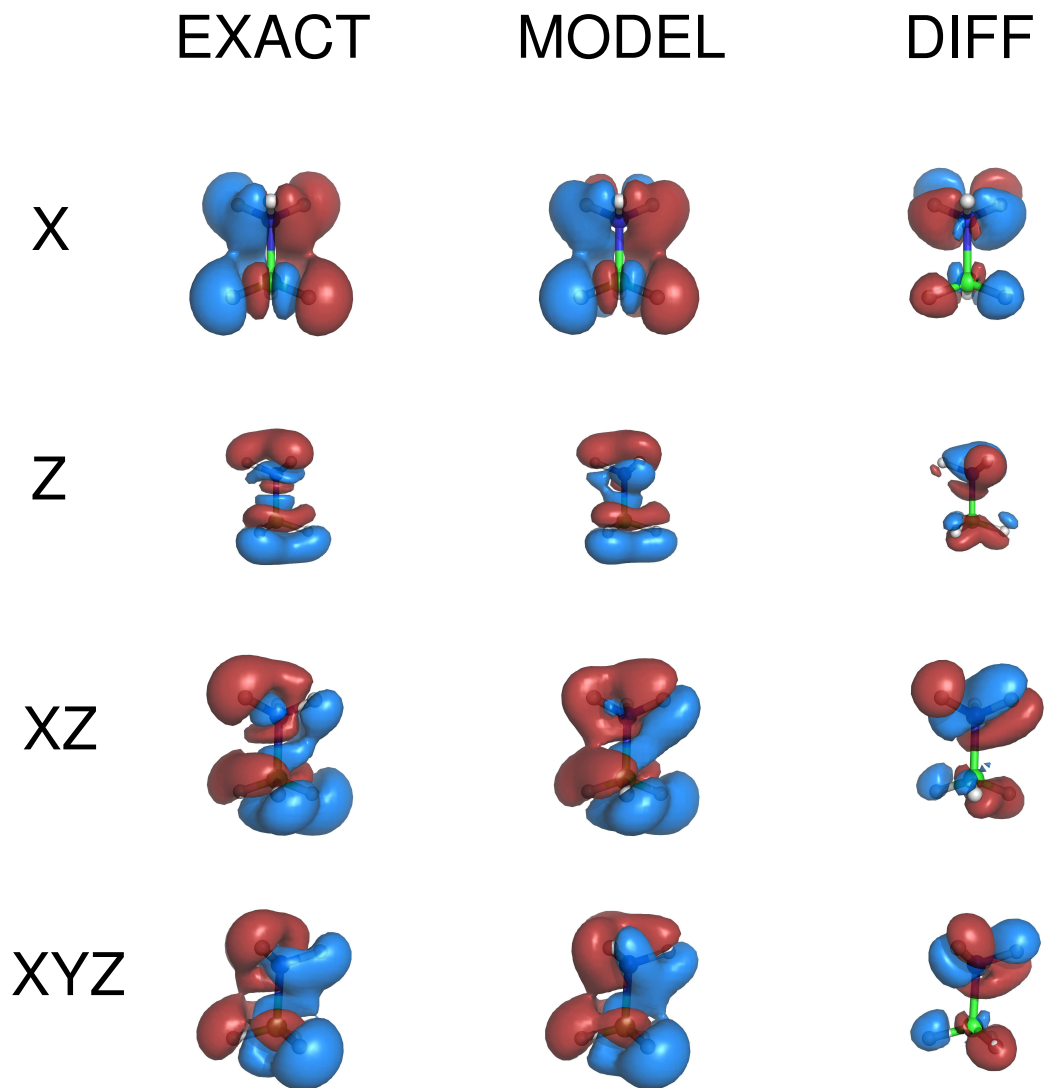




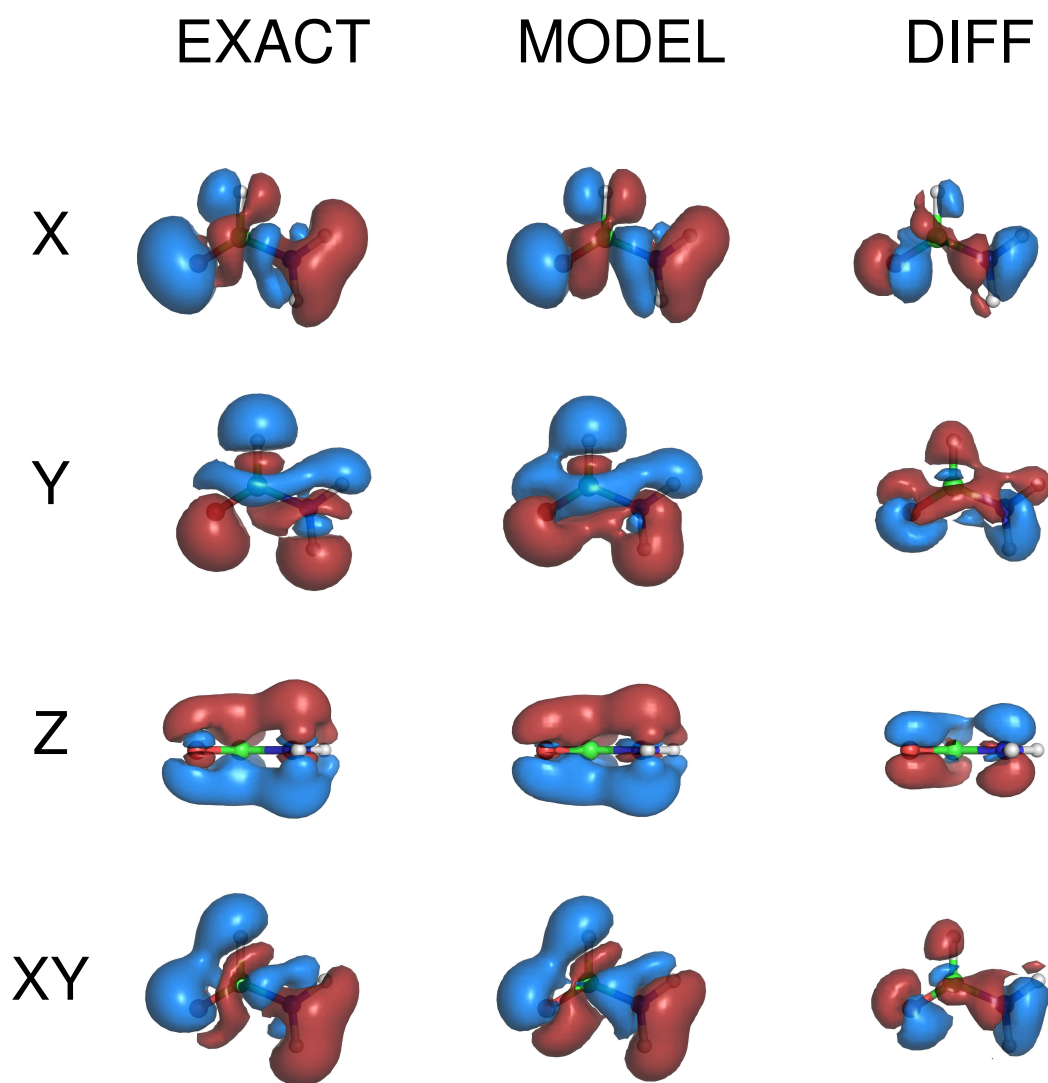
**Figure 5:** Polarization deformation densities of water ( $\text{H}_2\text{O}$ ) due to external uniform electric fields, that were obtained by using HF/6-311++G\*\* method ('EXACT') and proposed model ('MODEL'; Eq. (103)) derived at the same level. In this Figure, I also plot the difference  $\Delta\rho^{\text{DIFF}} \equiv \Delta\rho^{\text{EXACT}} - \Delta\rho^{\text{MODEL}}$ , which is denoted as 'DIFF'. The following electric field directions are tested ( $x, y, z$ , in au): 'X': (0.015, 0, 0); 'Y': (0, 0.015, 0); 'Z': (0, 0, 0.015); 'XY': (0.010607, 0.010607, 0). The isocontours were plotted for  $\pm 0.001$  au.



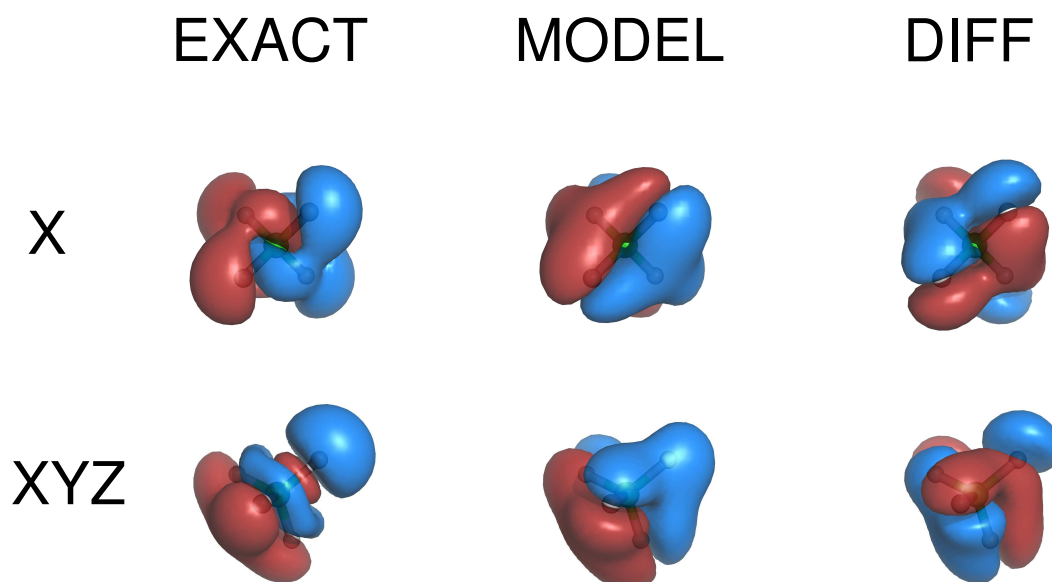
**Figure 6:** Polarization deformation densities of acetonitrile ( $\text{CH}_3\text{CN}$ ) due to external uniform electric fields, that were obtained by using HF/6-311++G\*\* method ('EXACT') and proposed model ('MODEL'; Eq. (103)) derived at the same level. In this Figure, I also plot the difference  $\Delta\rho^{\text{DIFF}} \equiv \Delta\rho^{\text{EXACT}} - \Delta\rho^{\text{MODEL}}$ , which is denoted as 'DIFF'. The following electric field directions are tested ( $x, y, z$ , in au): 'X': (0.015, 0, 0); 'Z': (0, 0, 0.015); 'XZ': (0.010607, 0, 0.010607); 'XYZ': (0.00866, 0.00866, 0.00866). The isocontours were plotted for  $\pm 0.001$  au.



**Figure 7:** Polarization deformation densities of methylamine cation ( $\text{CH}_3\text{NH}_3^+$ ) due to external uniform electric fields, that were obtained by using HF/6-311++G\*\* method ('EXACT') and proposed model ('MODEL'; Eq. (103)) derived at the same level. In this Figure, I also plot the difference  $\Delta\rho^{\text{DIFF}} \equiv \Delta\rho^{\text{EXACT}} - \Delta\rho^{\text{MODEL}}$ , which is denoted as 'DIFF'. The following electric field directions are tested ( $x, y, z$ , in au): 'X': (0.015, 0, 0); 'Z': (0, 0, 0.015); 'XZ': (0.010607, 0, 0.010607); 'XYZ': (0.00866, 0.00866, 0.00866). The isocontours were plotted for  $\pm 0.001$  au.



**Figure 8:** Polarization deformation densities of formamide ( $\text{HOC}-\text{NH}_2$ ) due to external uniform electric fields, that were obtained by using HF/6-311++G\*\* method ('EXACT') and proposed model ('MODEL'; Eq. (103)) derived at the same level. In this Figure, I also plot the difference  $\Delta\rho^{\text{DIFF}} \equiv \Delta\rho^{\text{EXACT}} - \Delta\rho^{\text{MODEL}}$ , which is denoted as 'DIFF'. The following electric field directions are tested ( $x, y, z$ , in au): 'X': (0.015, 0, 0); 'Y': (0, 0.015, 0); 'Z': (0, 0, 0.015); 'XY': (0.010607, 0.010607, 0). The isocontours were plotted for  $\pm 0.001$  au.



**Figure 9:** Polarization deformation densities of methane ( $\text{CH}_4$ ) due to external uniform electric fields, that were obtained by using HF/6-311++G\*\* method ('EXACT') and proposed model ('MODEL'; Eq. (103)) derived at the same level. In this Figure, I also plot the difference  $\Delta\rho^{\text{DIFF}} \equiv \Delta\rho^{\text{EXACT}} - \Delta\rho^{\text{MODEL}}$ , which is denoted as 'DIFF'. The following electric field directions are tested ( $x, y, z$ , in au): 'X': (0.015, 0, 0); 'XYZ': (0.00866, 0.00866, 0.00866). The isocontours were plotted for  $\pm 0.001$  au.

## 5 Planned Course of Research

Table 1: My caption

Period (months)	Task	Papers	Comments	Details
2	Vector model of density matrix polarization	1	SCF level only. Demonstrates construction of the effective potentials.	Implementation. Computation of polarization deformation matrices. Assessment of accuracy based on evaluation of the idempotency loss due to linear approximation. Evaluation of the induced multipole moments and comparison with benchmark results. Computation of polarization energy for model systems containing water, methanol, $\text{Na}^+$ , $\text{CH}_3\text{NH}_3^+$ etc. Comparison of induction interaction energy from vector model with the exact integral formulae. Direct comparisons with the distributed polarizability approach.
2	Calculation of EET coupling constant due to ET and HT processes	1	Using CIS benchmark. This work is to show the effective potentials. expanded in auxiliary basis set or approximated by ESP charges.	Derivation of all effective fragment potentials for ET1, ET2, HT1 and HT2 contributions. 3D visualisation of these potentials. Density fitting of these potentials to the special auxiliary <i>s</i> -type basis set. ESP fitting to generate effective charges. Practical implementations with density fitting or ESP charges. Computing EET coupling constants for model systems: ethylene, naphthalene, 5-aminocoumarin, formamide and chlorophyll dimers. Implementation of TDFI method and comparison.
3	Development of GEFP for Pauli repulsion.	1	SCF level only.	Working out the auxiliary basis set for $G_{\zeta i}$ or optionally developing the distributed effective charge model. Implementation of effective potentials. Computations of Pauli repulsion energies and comparison with EFP2 and exact calculations. Model systems: containing water, methanol, $\text{Na}^+$ , $\text{CH}_3\text{NH}_3^+$ etc.

6	Formulation of GEFP method at HF level.	1-2		Combining polarization vector model and GEFP Pauli repulsion. Implementing the localized polarized Löwdon-orthogonalized orbitals. Assessment of the validity of SGO approximation in calculation of the exchange and exchange-induction interaction energies. Computation of potential energy surfaces for model dimers containing water, methanol, $\text{Na}^+$ , $\text{CH}_3\text{NH}_3^+$ etc. Comparison with full HF results and EFP2 method. Assessment of the overall error of the GEFP/HF method. Analysis of the basis set extension effect on the total error with respect to full HF.
1	Working out the procedure of obtaining natural localized natural orbitals (NLMO's) and associated distributed polarizabilities.	0	Technical aspect.	Assessment of the vector polarization model at correlated levels. Assessment of the error due to non-idempotency of density matrix.
5	GEFP method at correlated level of theory	1	Ground states only.	Development of molecular occupancy model in the Pauli repulsion energy. Inclusion of polarization effects through the vector polarization model and/or Drude model. Implementing the localized polarized weight-orthogonalized orbitals. Computation of exchange and exchange-correlation energy. Computation of potential energy surfaces for model dimers containing water, methanol, $\text{Na}^+$ , $\text{CH}_3\text{NH}_3^+$ etc. Comparison with benchmark results and assessing the error of the method. Implementation of dispersion energy from Drude model.

5	GEFP method for electronically excited states: fragment-based embedding	0-1	Computation of distributed dipole moments for electronically excited state (Drude model or other technique). Application of the correlated GEFP method developed earlier to model the electronic energy levels as a function of the configuration of the solvent molecules. Analysis of model systems in which intermolecular charge transfer is small. Comparison with benchmark CIS, TDDFT or EOM-CC results.
---	---	-----	---

## References

- (1) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968–1986.
- (2) Mandado, M.; Hermida-Ramón, J. M. *J. Chem. Theory Comput.* **2011**, *7*, 633–641.
- (3) Błasiak, B.; Cho, M. *J. Chem. Phys.* **2015**, *143*, 164111.
- (4) Błasiak, B.; Ritchie, A. W.; Webb, L. J.; Cho, M. *Phys. Chem. Chem. Phys.* **2016**, *18*, 18094–18111.
- (5) Błasiak, B.; Lee, H.; Cho, M. *J. Chem. Phys.* **2013**, *139*, 044111.
- (6) Błasiak, B.; Cho, M. *J. Chem. Phys.* **2014**, *140*, 164107.
- (7) Murrell, J. N.; Randic, M.; Williams, D. R. *Proc. R. Soc. London, Ser. A* **1965**, *284*, 566–581.
- (8) Otto, P.; Ladik, J. *Chem. Phys.* **1975**, *8*, 192–200.
- (9) Błasiak, B.; Maj, M.; Cho, M.; Góra, R. W. *J. Chem. Theory Comput.* **2015**, *11*, 3259–3266.
- (10) Jensen, J. H. *J. Chem. Phys.* **1996**, *104*, 7795–7796.
- (11) Pipek, J.; Mezey, P. G. *J. Chem. Phys.* **1989**, *90*, 4916–4926.
- (12) Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 1736–1740.