





# Ab initio effective one-electron potential operators: Applications for charge-transfer energy in effective fragment potentials

Bartosz Błasiak  | Joanna D. Bednarska | Marta Chołuj  | Robert W. Góra  |  
Wojciech Bartkowiak 

Department of Physical and Quantum Chemistry, Faculty of Chemistry, Wrocław University of Science and Technology, Wrocław, Poland

## Correspondence

Bartosz Błasiak, Department of Physical and Quantum Chemistry, Faculty of Chemistry, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, Wrocław 50-370, Poland.  
Email: blasiak.bartosz@gmail.com

## Funding information

H2020 Marie Skłodowska-Curie Actions, Grant/Award Number: 665778; Narodowe Centrum Nauki, Grant/Award Number: 2016/23/P/ST4/01720

## Abstract

The concept of effective one-electron potentials (EOPs) has proven to be extremely useful in efficient description of electronic structure of chemical systems, especially extended molecular aggregates such as interacting molecules in condensed phases. Here, a general method for EOP-based elimination of electron repulsion integrals is presented, that is tuned toward the fragment-based calculation methodologies such as the second generation of the effective fragment potentials (EFP2) method. Two general types of the EOP operator matrix elements are distinguished and treated either via the distributed multipole expansion or the extended density fitting (DF) schemes developed in this work. The EOP technique is then applied to reduce the high computational costs of the effective fragment charge-transfer (CT) terms being the bottleneck of EFP2 potentials. The alternative EOP-based CT energy model is proposed, derived within the framework of intermolecular perturbation theory with Hartree–Fock noninteracting reference wavefunctions, compatible with the original EFP2 formulation. It is found that the computational cost of the EFP2 total interaction energy calculation can be reduced by up to 38 times when using the EOP-based formulation of CT energy, as compared to the original EFP2 scheme, without compromising the accuracy for a wide range of weakly interacting neutral and ionic molecular fragments. The proposed model can thus be used routinely within the EFP2 framework.

## KEYWORDS

ab initio force field, charge-transfer, effective fragment, electron repulsion integral, one-electron potential

## 1 | INTRODUCTION

In the framework of the theory of intermolecular forces, charge transfer (CT) is a rather technical term that originates due to algebraization in finite basis sets and arbitrary fragmentation of a quantum mechanical (QM) system.<sup>1–5</sup> Nonetheless, in supermolecular calculations of a complex

the charge transferred among interacting subsystems can be quite significant and the corresponding CT effect on the interaction energy is known to be often non-negligible. This is particularly evident in donor-acceptor systems such as H-bonded species and charged complexes.<sup>6,7</sup>

Since CT is associated with “interfragment” matrix elements, it constitutes a part of a variety of induction-related effects that are rigorously

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Journal of Computational Chemistry* published by Wiley Periodicals LLC.

defined in the symmetry adapted perturbation theory (SAPT).<sup>3</sup> Therefore, evaluation of CT contribution to the intermolecular interaction potential<sup>4</sup> is far from trivial due to its notably complex QM origins and cannot be realized in terms of any classical nor semiclassical approach. Even at the Hartree-Fock<sup>8</sup> (HF) approximation any separation of CT from intramolecular polarization is arbitrary and depends strongly on the chosen basis set,<sup>9,10</sup> to a point of becoming meaningless in the limit of a complete basis.<sup>2</sup> However, Stone and Misquitta showed that CT energy can in principle be extracted from SAPT calculations by comparing the induction energies of fictional systems in which basis functions are centered either solely on the monomers, or the entire interacting complex.<sup>5</sup> CT energy was also formulated by Murrell et al.<sup>11</sup> in their perturbation theory in the region of small wavefunction overlap up to second order. Nevertheless, all these theories are computationally too demanding to be applied in efficient calculations of intermolecular forces during molecular dynamics simulations, as they require evaluation of the electron repulsion integrals (ERIs) and their four-index transformation to molecular orbital (MO) basis.

The apparent difficulty in theoretically characterizing the CT energy in terms of the interacting molecular fragments is indeed a challenge in the development of modern force fields or ab initio fragmentation methods,<sup>12</sup> designed for modeling structure and dynamics of condensed phase systems.<sup>13</sup> This needs to be contrasted with the Coulombic electrostatics, non-Coulombic repulsion due to Pauli exclusion principle, or even dispersion and induction interactions, which all to a certain extent can be adequately described by relatively simple and computationally feasible mathematical models, like the distributed multipole (DMTP) moments of charge densities, the model van der Waals repulsive and attractive potentials, or the various distributed polarizability models. For that reason, the CT effects are usually not explicitly included in most of molecular mechanics force fields developed up to date.<sup>14</sup> Only a few polarizable force fields exist which explicitly incorporate the CT effects in an ab initio manner and are readily applicable to the condensed-phase simulations.<sup>15</sup> The notable examples include the second generation of the effective fragment potential (EFP2) method,<sup>9,16-23</sup> the Sum of Interactions Between Fragments Ab initio Computed (SIBFA) method,<sup>24,25</sup> the Explicit Polarization (X-Pol) method,<sup>26</sup> and the S/G-1 approach.<sup>27</sup> In all of the above methods, CT energy formulation is based on the antisymmetrization of certain subsets of Hartree products of monomer wavefunctions and a few additional approximations. Naturally, CT can also be implicitly included by performing full QM electronic structure simulations or nonforce-field-based fragmentation techniques.<sup>28,29</sup>

EFP2, being the most commonly used ab initio force field, was derived from the first-principles at the HF level<sup>9,17,18,30-32</sup> and augmented with intermolecular dispersion effects by the response theory.<sup>33,34</sup> That is to say, the total intermolecular interaction potential, which is completely free of any semiempirical parameters, is approximated as

$$E^{\text{EFP2}} \approx E^{\text{Coul}} + E^{\text{Ex-Rep}} + E^{\text{Ind}} + E^{\text{Disp}} + E^{\text{CT}}, \quad (1)$$

where  $E^{\text{Coul}}$  is the Coulombic interaction energy of the unperturbed charge-density distributions of the monomers, represented by the

DMTP approximation with damping to account for the charge-penetration effects,<sup>35</sup>  $E^{\text{Ex-Rep}}$  is the exchange-repulsion energy originating from the Pauli exclusion principle,<sup>30,31</sup> whereas  $E^{\text{Ind}}$  and  $E^{\text{Disp}}$  are the induction and dispersion energies obtained from the distributed polarizability approximation,<sup>32-34</sup> and finally  $E^{\text{CT}}$  is the CT energy,<sup>17,18</sup> being in the limelight of this work.

Despite the considerable success of the EFP2 approach in accurately modeling the extended molecular systems like solutions<sup>21-23,36</sup> and recently even biomolecules<sup>37-40</sup> with the level of accuracy reaching in many cases<sup>19</sup> that of the second-order Møller-Plesset perturbation theory,<sup>41</sup> evaluation of the CT energy in EFP2 model is still relatively costly for typical applications in the molecular dynamics simulations. It has been reported that the implementation of the EFP2 CT energy and gradient in GAMESS (US) computer program<sup>42</sup> with canonical molecular orbitals (CMO) is on average 20-30 times more computationally demanding than the other components.<sup>17,19</sup> Recent advancement of Xu and Gordon<sup>18</sup> reduced this cost further by up to 50%-60% by minimizing the size of the virtual orbital space via the use of quasi-atomic minimal-basis orbitals.<sup>43</sup> (QUAMBOs) In this approach, the diagonalization of a Fock matrix in QUAMBO basis yields the original HF occupied orbitals and the virtual valence orbitals (VVO), which are then used in the CT energy evaluation instead of the original canonical virtual MOs (CVOs). Nevertheless, even with this improvement, being now a standard default in most of EFP2 applications, the EFP2 CT term still remains the most time-consuming to evaluate from among all the EFP2 contributions.

In effect, the CT energy component is sometimes ignored when applying EFP2 to chemical problems.<sup>36-38,44-49</sup> In fact, EFP2 CT term is available only in the GAMESS (US) quantum chemistry program,<sup>42</sup> whereas it is neither supported in the official release of the recent LIBEFP library for linking quantum chemistry packages with the EFP2 functionalities,<sup>50</sup> nor in the Q-CHEM quantum chemistry program,<sup>51</sup> contrary to the remaining electrostatic, exchange-repulsion, induction, and dispersion EFP2 terms.

One of the main goals of this work, apart from developing a more efficient model of the CT energy that is compatible with the EFP2 approach, is to extend the definition of effective one-electron potential (here referred to as the EOP) technique, that has been widely explored in the past,<sup>8,10,17,52-60</sup> to simplify the rigorous and costly fragment-based quantum chemical models of extended systems with a particular emphasis on solvation phenomena and molecular dynamics simulations.

The presented EOP technique of removing ERIs from the working equations follows the notion of the importance of one-electron densities in chemistry,<sup>53,54</sup> thus reducing the complicated summations involving ERIs to much shorter expressions involving only one-electron integrals (OEl). Therefore, the generalized EOP computational method is first outlined in Section 2.1. Next, in Section 2.2, this new extended technique is used to derive an alternative formulation of the CT energy compatible with the EFP2 method. Subsequently, after details of computations are discussed in Section 3, in Section 4 the validation of the EOP-based CT model is discussed and its performance in terms of accuracy and computational speed is compared

against the EFP2 model. Finally, a few concluding remarks and outlook of future work are given in Section 5.

## 2 | THEORY

### 2.1 | Incorporating electron repulsion integrals into effective potentials

#### 2.1.1 | Generalized definition of EOPs for fragment-based methods

EOP operator can be expressed as  $\hat{v}^{\text{eff}} = \lambda \hat{v}_{\text{nuc}} + \int dr |r\rangle v_{\text{el}}^{\text{eff}}(\mathbf{r}) \langle r|$ , where the electronic part, associated with a certain effective density  $\rho^{\text{eff}}$ , is given by Equation (S1), whereas the nuclear part is defined in Equation (S5)—see Supporting Information for details about the notation. Consider now an arbitrary linear functional  $\mathcal{F}$  that explicitly depends on the ERIs. In this work, EOP-based ERI elimination procedure is defined by the following expression

$$\sum_{kl \in A} \mathcal{F} \left[ \langle ij | k^A l^A \rangle \right] = \langle i | \hat{v}^{\text{eff}, A} | j \rangle, \quad (2)$$

where  $\hat{v}^{\text{eff}, A}$  is the EOP operator associated with molecule (fragment) A. Note that, its mathematical form depends on the linear functional  $\mathcal{F}$ . In Equation (2), the summations over  $k$  and  $l$  orbitals are incorporated into a single one-electron matrix element. Thus, the total computational effort is, in principle, reduced from the two-fold sum involving evaluation of ERIs to just one much easier to compute OEI. It is also possible to generalize the above expression even further by summing over distinct linear functionals  $\mathcal{F}_t$ , as well as overall one-electron operators  $\hat{o}^A$ ,

$$\sum_t \mathcal{F}_t \langle BX | AA \rangle + \sum_s \langle B | \hat{o}_s^A | X \rangle = \sum_{ij \in A} \langle B | \hat{v}_i^{\text{eff}, A} | j^A \rangle \left[ \mathbf{S}^{-1} \right]_{ij} \langle j^A | X \rangle. \quad (3)$$

where the capital italic letters denote subsets of orbitals associated with a particular fragment and  $X = A$  or  $B$  (see the discussion below). The above design has the advantage that it opens the possibility to define first-principles effective fragments as long as the functionals  $\mathcal{F}_t$  are well defined, computable and can be partitioned in between the interacting fragments. The derivation of Equation (3) is shown in Appendix A.

Three unique classes of ERIs can be recognized based on the basis function partitioning scheme within the system composed of two molecules (shall be A and B throughout the course of this work). They are as follows:

1. the Coulomb-like ERIs of the type  $\langle AA | BB \rangle \rightarrow \langle i^A j^A | k^B l^B \rangle$ ,
2. the overlap-like ERIs of the type  $\langle AA | AB \rangle \rightarrow \langle i^A j^A | k^A l^B \rangle$ , and
3. the exchange-like ERIs of the type  $\langle AB | AB \rangle \rightarrow \langle i^A j^B | k^A l^B \rangle$ .

In contrast to the first two classes of ERIs, exchange-like ERIs cannot be incorporated into EOPs. The Coulomb and overlap-like classes, which are listed in Table 1, are usually approximated by expanding the EOP operator in DMTP expansion series, and integrating over one remaining electron coordinate<sup>17</sup> (see Equation [S13]). In this work,

**TABLE 1** Types of matrix elements with EOP operators

Matrix element	Overlap-like	Coulomb-like
	$\langle ij   \hat{v}^{\text{eff}, A}   lj \rangle$	$\langle ij   \hat{v}^{\text{eff}, A}   ll \rangle$
Partitioning scheme	$i \in A, j \in B$	$j, l \in B$
ERI class	$\langle AA   AB \rangle$	$\langle AA   BB \rangle$
DF/RI form	$\sum_{\xi \in A} V_{\xi i}^A S_{\xi j}^{AB}$	$\sum_{\xi \in A} S_{\xi j}^{BA} V_{\xi \xi}^A S_{\xi l}^{AB}$
DMTP form	–	$\rho_{jj}^B \odot \rho^{\text{eff}, A}$

however, this method is considered already too expensive for application in the CT energy because evaluation of Equation (S13) requires calculation of electrostatic potential and electrostatic potential gradient(s) OEs. These kind of integrals are typically the most expensive when compared to other standard OEs such as overlap or kinetic energy integrals. Therefore, Coulomb EOP matrix elements will be treated via semiclassical DMTP expansion (see Section S4) which is more approximate but much less expensive approach. The overlap-like EOP matrix elements will be treated via the *extended DF* method, which is discussed next.

#### 2.1.2 | Extended DF of EOPs

Extended DF of EOPs, which will be referred to as the EDF scheme, is applicable in case of matrix elements of  $\langle i^A | \hat{v}^{\text{eff}, A} | j^B \rangle$  type. These matrix elements require ERIs of type  $\langle AA | AB \rangle$  only. To get the ab initio representation of such an overlap-like matrix element, one can use a procedure similar to the typical DF or resolution of identity (RI), which are nowadays routinely used to compute electron-repulsion integrals (ERIs) more efficiently, and reduce computational cost of post-HF methods.<sup>61</sup> DF was also applied to design ab initio force fields.<sup>57,58</sup>

An arbitrary one-electron potential acting on state vectors  $|A\rangle$  of molecule A can be expanded in an auxiliary and generally nonorthogonal AO space  $|a\rangle$  centered on A as

$$\hat{v}^{\text{eff}, A} |A\rangle = |a\rangle \mathbf{V}_a^A, \quad (4)$$

where

$$\mathbf{V}_a^A = \mathbf{S}_{aa}^{-1} \langle a | \hat{v}^{\text{eff}, A} | A \rangle \quad (5)$$

under the necessary assumption that the auxiliary basis set is nearly complete, that is,  $|a\rangle \mathbf{S}_{aa}^{-1} \langle a| \cong \mathbf{1}_{aa}$ . In the above equations,  $\mathbf{S}_{ab} = \langle a | b \rangle$  denotes the overlap AO integrals matrix. In practice, basis sets approximately fulfilling such a RI are relatively much larger than the primary AO basis sets. Therefore, it should be computationally more efficient to utilize smaller auxiliary AO basis set  $|m\rangle$ , which satisfies

$$\hat{v}^{\text{eff}, A} |A\rangle \approx |m\rangle \mathbf{V}_m^A. \quad (6)$$

In Equation (6), the EOP matrix  $\mathbf{V}_m^A$  is for the time being unknown.

To find an expression for  $\mathbf{V}_m^A$ , consider a certain orthonormal MO basis

$$|\mathbf{X}\rangle = |\mathbf{a}\rangle \mathbf{T}_{\mathbf{aX}} \quad (7)$$

for  $\mathbf{T}_{\mathbf{aX}} = \mathbf{S}_{\mathbf{aa}}^{-1} \langle \mathbf{a} | \mathbf{X} \rangle$ , which is as small as possible but accurately represents the EOP operator, that is,

$$\hat{v}^{\text{eff},A} |\mathbf{A}\rangle = |\mathbf{X}\rangle \langle \mathbf{X} | \hat{v}^{\text{eff},A} | \mathbf{A} \rangle, \quad (8)$$

with  $\mathbf{T}_{\mathbf{aX}}^\dagger \mathbf{S}_{\mathbf{aa}} \mathbf{T}_{\mathbf{aX}} = \mathbf{1}_{\mathbf{XX}}$  due to orthonormality. The similarity transformation matrix  $\mathbf{T}_{\mathbf{aX}}$  can be found from essential eigenvectors of the covariance matrix of  $\mathbf{V}_{\mathbf{a}}^A$  expressed in the orthogonal RI basis after Löwdin symmetric orthogonalization, that is,

$$\mathbf{T}_{\mathbf{aX}} = \mathbf{S}_{\mathbf{aa}}^{-\frac{1}{2}} \mathbf{Q} \mathbf{U}_{\mathbf{aX}}, \quad (9)$$

where

$$\mathbf{V}_{\mathbf{a}} \mathbf{V}_{\mathbf{a}}^\dagger = \mathbf{U}_{\mathbf{aX}} \mathbf{g}_{\mathbf{XX}} \mathbf{U}_{\mathbf{aX}}^\dagger \quad (10)$$

and

$$\mathbf{V}_{\mathbf{a}} = \mathbf{S}_{\mathbf{aa}}^{\frac{1}{2}} \mathbf{V}_{\mathbf{a}}. \quad (11)$$

The operator  $\mathbf{Q}$  in Equation (9) selects only eigenvectors  $\mathbf{U}_{\mathbf{aX}}$  associated with the nonvanishing eigenvalues stored in the diagonal matrix  $\mathbf{g}_{\mathbf{XX}}$ . Note that, the size of basis  $\mathbf{X}$  is bounded from above by the number of state vectors  $\mathbf{A}$ , which is also an upper bound for the size of the auxiliary AO basis  $\mathbf{m}$ . The latter can be found by maximizing the overlap between the  $\mathbf{X}$  MOs and their approximate expansion in basis  $\mathbf{m}$ , that is,

$$Z[\mathbf{m}] = \max\{\text{Tr } \mathbf{S}_{\mathbf{XX}}\}, \quad (12)$$

where the overlap matrix is

$$\mathbf{S}_{\mathbf{XX}} = \left( \mathbf{T}_{\mathbf{aX}}^\dagger \mathbf{S}_{\mathbf{am}} \mathbf{S}_{\mathbf{mm}}^{-1} \mathbf{S}_{\mathbf{ma}} \mathbf{T}_{\mathbf{aX}} \right)^{\frac{1}{2}}. \quad (13)$$

The approximate transformation matrix in  $\mathbf{m}$  basis can be found by using the basis set projection method of Polly et al.,<sup>62</sup>

$$\mathbf{T}_{\mathbf{mX}} = \mathbf{S}_{\mathbf{mm}}^{-1} \mathbf{S}_{\mathbf{ma}} \mathbf{T}_{\mathbf{aX}} \mathbf{S}_{\mathbf{XX}}^{-1}. \quad (14)$$

From the above analysis, the approximate identity operator is

$$\mathbf{1}_{\mathbf{mm}} \approx |\mathbf{m}\rangle \mathbf{T}_{\mathbf{mX}} \mathbf{S}_{\mathbf{XX}}^{-1} \mathbf{T}_{\mathbf{mX}}^\dagger \langle \mathbf{m} |, \quad (15)$$

which results in the following expression for the EOP matrix,

$$\mathbf{V}_{\mathbf{m}}^A = \mathbf{T}_{\mathbf{mX}} \mathbf{S}_{\mathbf{XX}}^{-1} \mathbf{T}_{\mathbf{mX}}^\dagger \mathbf{S}_{\mathbf{ma}} \mathbf{V}_{\mathbf{a}}^A. \quad (16)$$

Equations (9)–(16) define the EDF method and once basis  $\mathbf{m}$  is found Equation (6) can be used to effectively eliminate ERIs and

replace them with products of EOP matrix  $\mathbf{V}_{\mathbf{m}}^A$  and overlap matrix involving  $\mathbf{m}$  AOs. Note that, in the limiting case of  $|\mathbf{m}\rangle = |\mathbf{a}\rangle$  the EDF method reduces to the usual DF in the nearly-complete AO basis according to Equation (5).

It is emphasized here that other possibilities of formulating the  $\mathbf{V}_{\mathbf{m}}^A$  matrix exist (see Section S5). However, it was found that Equation (16) seems to be the most computationally efficient because only overlap integrals are required.

## 2.2 | Charge transfer interaction energy for fragment potentials

As the theory that is necessary to eliminate ERIs from fragment-based models has been given in the previous section, we shift our attention to the CT energy formulation for the EFP2 model. In the CT treatments of bi-molecular complexes, the CT energy can be expressed as a sum of the stabilization energy due to excitations from molecule  $A$  to  $B$  and vice versa, that is,

$$E^{\text{CT}} = E^{A \rightarrow B} + E^{B \rightarrow A}. \quad (17)$$

### 2.2.1 | EFP2 model

To provide a complete account on the new CT energy expression proposed here, the original EFP2 formulation is briefly reviewed first. Li, Gordon, and Jensen used the expansion of the overlap density in Taylor series and found four different approximate theories for the CT energy.<sup>17</sup> The optimal theory, which was shown to well reproduce the CT energies obtained by using the reduced variational space (RVS) analysis of Stevens and Fink,<sup>63</sup> reads as

$$E^{A \rightarrow B} \approx 2 \sum_{i \in A}^{\text{Occ}} \sum_{n \in B}^{\text{Vir}} \frac{|U_{in}^{A \rightarrow B}|^2}{\epsilon_i - T_{nn}}, \quad (18)$$

where

$$|U_{in}^{A \rightarrow B}|^2 \approx \frac{u_{in}}{1 - \sum_{m \in A}^{\text{All}} S_{mn}^2} \left\{ u_{in} + \sum_{j \in B}^{\text{Occ}} S_{ij} \left( T_{nj} - \sum_{m \in A}^{\text{All}} S_{nm} T_{mj} \right) \right\}, \quad (19)$$

and

$$u_{in} \equiv U_{in}^{\text{eff},B} - \sum_{m \in A}^{\text{All}} U_{im}^{\text{eff},B} S_{mn}, \quad (20)$$

in which the summations extend over occupied (denoted by “Occ”), virtual (denoted by “Vir”) or both (denoted by “All”) sets of MOs. Note that, in case of using QUAMBOs as MO basis for Fock matrix, VVOs and resulting orbital energies (instead of the original canonical HF

orbitals and energies) are to be used. The effective potential energy matrix elements are defined by

$$U_{in}^{\text{eff},B} \equiv -\langle i | \hat{v}_{\text{tot}}^B | n \rangle, \quad (21)$$

and are evaluated by expanding the  $\hat{v}_{\text{tot}}^B$  operator in DMTP series according to Equation (S13). The apparent success of the EFP2 scheme is rooted in the dramatic simplifications of the ab initio expressions for interaction energy, in which the relatively costly ERIs have been effectively removed from the working models while maintaining the required accuracy. That is, in the case of the EFP2 CT energy component, the MO energies  $\epsilon_i$  are constant parameters, whereas the overlap  $S_{nm}$ , kinetic energy  $T_{mn}$ , and effective one-electron electrostatic potential  $U_{in}^{\text{eff}}$  matrix elements are all certain types of OEIs, orders of magnitude cheaper to evaluate than ERIs. Unfortunately, due to extensive summations over virtual orbitals, evaluating Equation (18) is still considerable in cost because typically large basis sets need to be used for generating the EFP2 parameters. In effect, calculation of  $U_{in}^{\text{eff}}$  is more expensive as compared to other types of OEIs and is the bottleneck of EFP2 CT energy calculation, even when using VVOs. In the following subsections, the alternative model of the CT energy is proposed by introducing EOPs. Although application of the EOP method to the CT formulation in EFP2 method is probably possible, it would be relatively difficult to discuss the resulting EOP-based EFP2 models because there are four distinct versions of this theory with a set of different approximations, selected based on performance assessment rather than a rigorous derivation manner.<sup>17</sup> Instead, perturbation theory of Murrell et al.<sup>11</sup> with the explicit formulation for closed shell systems by Otto and Ladik,<sup>10</sup> which is somewhat more rigorous than the EFP2 CT model, is chosen as a starting point in this work. It is believed that this choice will enable a clear demonstration of the EOP technique in fragment-based modeling.

## 2.2.2 | Otto-Ladik's model: Starting point

The CT energy at HF level of theory can be expressed by<sup>10,11</sup>

$$E^{A \rightarrow B} = 2 \sum_{i \in A} \sum_{n \in B}^{\text{Occ}} \frac{|U_{in}^{A \rightarrow B}|^2}{\epsilon_i - \epsilon_n}, \quad (22)$$

where the coupling constant is given by Otto and Ladik,<sup>10</sup> here referred to as the OL method, as

$$U_{in}^{A \rightarrow B} = -\langle i | \hat{v}_{\text{tot}}^B | n \rangle + \sum_{j \in B}^{\text{Occ}} \langle i | \hat{v}_{jn}^B | j \rangle + \sum_{k \in A}^{\text{Occ}} S_{nk} \langle k | \hat{v}_{\text{tot}}^B | i \rangle + \sum_{j \in B}^{\text{Occ}} S_{ij} \langle j | \hat{v}_i^A | n \rangle - \sum_{k \in A}^{\text{Occ}} \sum_{j \in B}^{\text{Occ}} S_{kj} (1 + \delta_{ik}) \langle i | \hat{v}_{jn}^B | k \rangle, \quad (23)$$

where  $\langle i | \hat{v}_{kl}^B | j \rangle \equiv -\langle ij | kl \rangle$  (see the notation convention explained in Supporting Information). In the above expression, the following EOP operators,

$$\hat{v}_{\text{tot}}^B = \hat{v}_{\text{nuc}}^B + 2 \sum_{j \in B}^{\text{Occ}} \hat{v}_{jj}^B \quad \text{and} \quad (24a)$$

$$\hat{v}_i^A = \hat{v}_{\text{tot}}^A - 2\hat{v}_{ii}^A, \quad (24b)$$

were introduced without making any approximation to the original equation from Ref. [10] Note that, ERIs in MO basis are necessary to evaluate all terms in Equation (23).

## 2.2.3 | Otto-Ladik's model: Application of EOP technique

One can immediately notice that the five summation terms from Equation (23) can be classified based on Table 1 into three groups regarding the type of ERIs that are required: (i) overlap-like  $\langle AB | BB \rangle$ —the first two terms; (ii) Coulomb-like  $\langle AA | BB \rangle$ —the third term and (iii) Coulomb-like  $\langle BB | AA \rangle$ —the two last terms. Note also that, there are no exchange-like terms needed in this case. Therefore, all the contributions can be recast in terms of the EOPs.

*Group (i).* Group (i) can be rewritten by using Equation (3) to eliminate the interfragment ERIs of the overlap-like type:

$$\langle i | \left[ -\hat{v}_{\text{tot}}^B | n \rangle + \sum_{j \in B}^{\text{Occ}} \hat{v}_{nj}^B | j \right] \rangle \cong \left\langle i | \sum_{\eta \in B} V_{\eta n}^B | \eta \right\rangle \quad (25)$$

where the EOP matrix is given according to Equation (5) by

$$V_{\eta n}^B = \sum_{\zeta} [\mathbf{S}^{-1}]_{\eta\zeta} V_{\zeta n}^B \quad (26)$$

(with  $[\mathbf{S}_{aa}]_{\eta\zeta} \equiv S_{\eta\zeta}$ ) and

$$V_{\zeta n}^B = -\sum_{y \in B} W_{\zeta n}^{(y)} + \sum_{j \in B}^{\text{Occ}} \{2\langle \zeta n | jj \rangle - \langle \zeta j | nj \rangle\}. \quad (27)$$

Note that, all the calculations that are required to obtain  $V_{\zeta n}^B$  are performed solely on the densities and basis sets associated with the unperturbed molecule B. Therefore,  $V_{\zeta n}^B$  can be considered as effective fragment parameters used to compute the first two terms of Equation (23) and the final expression reads

$$-\langle i | \hat{v}_{\text{tot}}^B | n \rangle + \sum_{j \in B}^{\text{Occ}} \langle i | \hat{v}_{nj}^B | j \rangle \cong \sum_{\eta \in B} S_{i\eta} V_{\eta n}^B, \quad (28)$$

which is a great simplification over the original form of group (i) because only the overlap integrals between the  $i$ th MO on molecule A and  $\eta$ th auxiliary orbital on molecule B need to be evaluated. Note that, the only approximation made so far was the application of DF and RI. If the RI basis set is sufficiently large, the errors due to this approximation can be minimal and negligible in principle. Alternatively, the optimized auxiliary AO basis set instead of the RI AO basis set can be used according to the EDF method developed in Section 2.1.2. In this case, the matrix elements of the EOP matrices are given by Equation (16) instead of Equation (5), and the labels  $\eta$  and  $\zeta$  in Equation (26) refer to the minimal auxiliary basis set of fragment B. This variant is easiest to realize when the VVOs are used because their amount is always less than the minimal basis set size, making the lower bound of the size of the auxiliary AO basis set very small.

*Group (ii).* The term belonging to this group can be considered as a sum of interaction energies between the total charge density distribution of molecule B and the partial density  $\rho_{ik}(\mathbf{r})$  of molecule A, weighted by the overlap integrals  $S_{nk}$ . Using the DMTP expansion based at the charge centroids of the localized molecular orbitals (LMOs),  $\mathbf{r}_i = \langle i|\hat{\mathbf{r}}|i\rangle$  with  $\chi_i(\mathbf{r})$  being localized, this group can be approximated by

$$\sum_{k \in A}^{\text{Occ}} S_{nk} \langle k|\hat{v}_{\text{tot}}^B|i\rangle \approx S_{ni} \rho_{ii}^A \odot \rho_{\text{tot}}^B. \quad (29)$$

Here, it was assumed that  $|\rho_{ik}(\mathbf{r})| \ll \rho_{ij}(\mathbf{r})$  for  $i \neq k$  in most locations in the case of LMOs, which allows one to conjecture that

$$\rho_{ki}^A \odot \rho_{\text{tot}}^B \approx \delta_{ik} \rho_{ii}^A \odot \rho_{\text{tot}}^B. \quad (30)$$

Now, the DMTP expansion of the interaction energy in the right hand side of Equation (30) can be expressed as

$$\rho_{ii}^A \odot \rho_{\text{tot}}^B \approx q_i \left[ \sum_{y \in B}^{\text{At}} \frac{Z_y}{|\mathbf{r}_y - \mathbf{r}_i|} + 2 \sum_{j \in B}^{\text{Occ}} \frac{q_j}{|\mathbf{r}_j - \mathbf{r}_i|} \right], \quad (31)$$

because the distributed charges  $q_i = -1$  whereas the distributed dipole moments centered at their respective LMO charge centroids vanish.<sup>64</sup> This means that Equation (29) can be finally given as follows:

$$\sum_{k \in A}^{\text{Occ}} S_{nk} \langle k|\hat{v}_{\text{tot}}^B|i\rangle \approx -S_{ni} \left[ \sum_{y \in B}^{\text{At}} \frac{Z_y}{r_{yi}} - \sum_{j \in B}^{\text{Occ}} \frac{2}{r_{ji}} \right]. \quad (32)$$

Therefore, only overlap integrals and relative distances between atomic and LMO centroid positions are needed, which leads to a great reduction of the calculation cost, as compared either to the original expression or to the multipole expansion (left- and right-hand sides of Equation (29), respectively). We shall refer to this approximation as to the local overlap approximation (LOA) resulting in similar expressions to the ones obtained by Jensen and Gordon in their exchange-repulsion interaction energy EFP2 model (see eq. [39] in ref. [30]).

Note that, to make this approximation valid, occupied molecular orbitals need to be spatially localized.

*Group (iii).* The terms with the overlap integrals involving the occupied MO on A can be combined into a single summation term, that is,

$$\begin{aligned} & \sum_{j \in B}^{\text{Occ}} S_{ij} \langle j|\hat{v}_i^A|n\rangle - \sum_{k \in A}^{\text{Occ}} \sum_{j \in B}^{\text{Occ}} S_{kj} (1 - \delta_{ik}) \langle j|\hat{v}_{ik}^A|n\rangle = \sum_{k \in A}^{\text{Occ}} \sum_{j \in B}^{\text{Occ}} S_{kj} \langle j|\hat{v}_{ik}^{\text{eff}}|n\rangle \\ & \cong \sum_{k \in A}^{\text{Occ}} \sum_{j \in B}^{\text{Occ}} S_{kj} \rho_{jn}^B \odot \rho_{ik}^{\text{eff}}, \end{aligned} \quad (33)$$

where the effective potential  $\hat{v}_{ik}^{\text{eff}}$  (with the associated effective density  $\rho_{ik}^{\text{eff}}$ ) is defined by

$$\hat{v}_{ik}^{\text{eff}} \equiv \delta_{ik} \left[ \hat{v}_{\text{tot}}^A - 2\hat{v}_{kk}^A + \hat{v}_{ik}^A \right] - \hat{v}_{ik}^A. \quad (34)$$

In order to include the  $\rho_{jn}^B$  density, it is approximately represented here by a set of effective cumulative atomic charges  $\{q_v^{B,(jn)}\}$  associated with the effective one-particle density matrix

$$\mathbf{P}_{\beta\delta}^{B,(jn)} = \mathbf{C}_{\beta j} \mathbf{C}_{\delta n}. \quad (35)$$

In this work, the effective charges were defined via the Mulliken method as discussed in Supporting Information with  $\lambda = 0$ . By applying the LOA for the  $\rho_{ik}^A$  density the effective potential from Equation (34) simplifies to

$$\hat{v}_{ik}^{\text{eff}} \approx \delta_{ik} \left\{ \hat{v}_{\text{tot}}^A - 2\hat{v}_{kk}^A \right\}, \quad (36)$$

which leads to

$$\sum_{j \in B}^{\text{Occ}} S_{ij} \langle j|\hat{v}_{\text{tot}}^A|n\rangle - \sum_{k \in A}^{\text{Occ}} \sum_{j \in B}^{\text{Occ}} S_{kj} \langle j|\hat{v}_{ik}^A|n\rangle \approx \sum_{j \in B}^{\text{Occ}} S_{ij} \sum_{y \in B}^{\text{At}} q_y^{B,(jn)} \left[ \sum_{x \in A}^{\text{At}} \frac{Z_x}{r_{xy}} + \frac{2}{r_{iy}} - \sum_{k \in A}^{\text{Occ}} \frac{2}{r_{ky}} \right]. \quad (37)$$

*Final EOP-based forms of the coupling constant.* Gathering the results from previous paragraphs, the coupling constant can be given as follows

$$\mathbf{V}_{in}^{A \rightarrow B} \approx \mathbf{G}_{1,in}^{A \rightarrow B} + \sum_{i' \in A}^{\text{LMO}} \mathbf{L}_{i'i}^A \left\{ \mathbf{G}_{2,i'n}^{A \rightarrow B} + \mathbf{G}_{3,i'n}^{A \rightarrow B} \right\}, \quad (38)$$

where the symbols  $\mathbf{G}_n$  ( $n = 1, 2, 3$ ) denote a particular group of terms from Equations (28), (32), and (37), and the primes denote the localized MOs with the  $\mathbf{L}^A$  matrix being the CMO-LMO transformation matrix of molecule A. For completeness, we list all the EOP-approximated contributions to the coupling constants below:

$$G_{1,in}^{A \rightarrow B} \equiv \sum_{\eta \in B}^{Rl} V_{n\eta}^B S_{\eta i}, \quad (39a)$$

$$G_{2,fn}^{A \rightarrow B} \equiv -S_{n\eta} U_{\eta}^{BA}, \quad (39b)$$

$$G_{3,fn}^{A \rightarrow B} \equiv \sum_{j \in B}^{Occ} S_{ij} \sum_{y \in B}^{At} q_y^{B,(jn)} w_{y\eta}^{BA}, \quad (39c)$$

where the auxiliary variables are defined as

$$U_{\eta}^{BA} \equiv \sum_{y \in B}^{At} \frac{Z_y}{r_{y\eta}} - \sum_{j \in B}^{Occ} \frac{2}{r_{ji}}, \quad (40a)$$

$$w_{y\eta}^{BA} \equiv \sum_{x \in A}^{At} \frac{Z_x}{r_{xy}} + \frac{2}{r_{iy}} - \sum_{k \in A}^{Occ} \frac{2}{r_{ky}}. \quad (40b)$$

Note that, the LOA-based contributions  $G_2$  and  $G_3$  need to be transformed back to the canonical MO basis. Thus, the final working formula for the interaction energy due to CT with excitations from A to B reads

$$E^{A \rightarrow B} \approx 2 \sum_{i \in A}^{Occ} \sum_{n \in B}^{Vir} \frac{1}{\epsilon_i - \epsilon_n} \left( G_{1,in}^{A \rightarrow B} + \sum_{\eta \in A}^{LMO} L_{i\eta}^A \left\{ G_{2,fn}^{A \rightarrow B} + G_{3,fn}^{A \rightarrow B} \right\} \right)^2. \quad (41)$$

The total CT energy is given by the sum of the above contribution and the twin contribution due to CT from molecule B to A according to Equation (17).

### 3 | METHODS

#### 3.1 | Implementation

The benchmark CT energy was assumed to be the CT energy defined in the RVS method.<sup>63</sup> All the models that were used to test the theory presented in this work, that is, the EFP2, OL, EOP, and RVS methods, as well as the EDF method, were implemented in our in-house plugin to Psi4 quantum chemistry program.<sup>65</sup> For the CT EFP2 component, potential energy integrals from Equation (S13) were calculated with the CAMM up to quadrupoles (distributed centers are atoms), instead of the DMA (distributed centers are atoms and mid-bond points) as implemented in most of quantum chemistry programs. The choice of CAMM versus DMA was due to convenience of implementation and, because the quantitative accuracy of our EFP2 CT energy code is comparable to the EFP2 code of GAMESS (US). Thus using only atomic distribution centers does not affect the interpretation of results in Section 4. It was also found that accuracy of the LOA from Equations (32) and (37) is usually slightly better when the Boys localization method<sup>66</sup> is used, as compared to the Pipek–Mezey method.<sup>67</sup> Henceforth, the former method was used for molecular orbital

localization throughout all the production calculations. For the sake of convenience, Cartesian  $d$  and  $f$  functions were utilized in all calculations. QUAMBOs were implemented based on the restricted open-shell Hartree–Fock<sup>68</sup> solutions of free atoms.  $1s^2 2s^2 2p^3$  electronic configuration was assumed for free carbon atom.

#### 3.2 | Validation: bi-molecular complexes

To perform statistical analysis of the accuracy of the EOP CT energy evaluation, structural databases of bi-molecular complexes in the noncovalent interactions database NCB31 developed by the Truhlar's group,<sup>69,70,70,71</sup> as implemented in the Psi4 program,<sup>65</sup> as well as the database for ionic systems of Řežáč and Hobza,<sup>72</sup> were utilized.

To analyze the asymptotic dependence of the CT energy, four complexes: (i)  $(H_2O)_2$ , (ii)  $H_2O-CH_3OH$ , (iii)  $H_2O-NH_4^+$ , and (iv)  $NO_3^- - NH_4^+$ , were chosen as model systems. The reference (zero-displacement) geometries were obtained by performing energy-optimizations at the HF/6-31+G(d,p) level,<sup>73–76</sup> as implemented in the GAUSSIAN 16 quantum chemistry program package.<sup>77</sup> Subsequently, 30 displaced geometries for each model complex were obtained by translating one of the monomers along the vector co-linear with the H-bond or N–N distance in the case of ammonium nitrate. The reference structures as well as the translation vectors are indicated in the insets of Figure 2. Throughout all the calculations for bi-molecular systems, 6-311++G(d,p) primary<sup>73,78–82</sup> and aug-cc-pVDZ-JKFit auxiliary<sup>83</sup> basis sets were used to obtain the EOP matrices (see Equation (5)) CT energies, unless the EDF scheme (Equation (16)) was utilized in the case of which the auxiliary basis sets were optimized for each species separately (see description below). For helium, 6-311+G(d,p) primary<sup>80</sup> and aug-cc-pVDZ-RI auxiliary<sup>84</sup> basis sets were used.

#### 3.3 | Validation: multi-fragment complexes

To test the total accuracy of the EFP2 method with the EOP-based CT term, three different multi-fragment model systems were selected. The structures were randomly sampled from classical molecular dynamics simulations of bulk water, DMSO, and  $(C_2mim)(NTf_2)$  ionic liquid in standard conditions, and subsequently optimized at EFP2/6-31++G(d,p) level by using the “GLOBOPT” routine of GAMESS (US) program.<sup>42</sup> Due to considerable size of the complexes chosen, 6-31++G(d,p) basis set<sup>73–76,78,82,85</sup> was used to obtain the interaction energies. Full RHF interaction energies were calculated as the reference values in the aggregate-centered basis set<sup>86</sup> (ACBS), to eliminate the basis set superposition error (BSSE). To superimpose EFP2 parameters (including the EOP matrices), the Kabsch method was used.<sup>87,88</sup> Generation of the EFP2 parameters for reference EFP2 interaction energy calculations was undertaken by using the “MAKEFP” routine of GAMESS (US) program.<sup>42</sup> The DMA method<sup>89,90</sup> was used to generate the DMTP expansion for electrostatics. Due to technical aspects of the implementation in our in-house code, the core MOs were also

included. Boys method was used to localize the MOs.<sup>66</sup> Charge-penetration effects for Coulomb and induction interactions were taken into account by using the overlap-correction method<sup>91</sup> and the Tang–Toennies method,<sup>92</sup> respectively. The Tang–Toennies damping constant was set to a default value of 0.6 in GAMESS (US). Infinite cut-off thresholds for the evaluation of Coulomb, induction, exchange-repulsion, and CT interactions were assumed. Since the reference level of theory is HF, dispersion interactions were not considered in this work and, therefore, they were omitted in all the EFP2 interaction energy calculations.

### 3.4 | Auxiliary AO basis set optimizations

The optimization routine of the EDF method was fully automatized in our in-house computer program. Basin hopping global optimization algorithm<sup>93–96</sup> was used, as implemented in the SciPy Python library.<sup>97</sup> Since the local minima of the objective function are usually separated by c.a.,  $0.002 \times Z_{\max}$  where  $Z_{\max}$  is equal to the number of VVOs, the consecutive basin hopping step was assumed to be accepted with the Metropolis probability of  $\exp\left(-\frac{Z_{\text{new}} - Z_{\text{old}}}{0.002 \times Z_{\max}}\right)$ . In all of our optimizations the optimal value of objective function exceeded  $0.99 \times Z_{\max}$  and was reached in only 10 basin hopping steps. To find the local minimum in each step, the Sequential Least Squares Programming method<sup>98</sup> was used with tolerance  $1.0 \times 10^{-9}$ , whereas bounds for orbital exponents and contraction coefficients were set as [0.0002,5000.0] and [0.001,1.000], respectively. Additional constraint for the normalization of contraction coefficients was also implemented. All optimized auxiliary basis sets were assumed to be triply-contracted minimal basis sets with a single set of polarization  $p$ -type functions on the 1-row atoms and  $d$ -type on larger atoms. The starting guess parameters were found to be only weakly dependent on the chemical composition and the primary basis set. The optimized basis sets for selected molecules are given in Supporting Information, Section S8.

### 3.5 | CPU time profiling

Time profiling of the code for the EFP2 and the EOP methods was performed for all the computational operations required for a single point energy calculation in a hypothetical sequential run on multiple geometries such as during a typical molecular dynamics calculation. Therefore, all the multi-fragment system dependent calculables were taken into account in the profiling, that is, calculations of OEIs in AO basis, their two-index transformations to MO basis and assembling of the interaction energy terms. On the other hand, calculations of the effective fragment parameters and initial computation of spherical harmonics transforms necessary for computation of the OEIs were not included, since they need to be evaluated only once. Time profiling was performed on 1.2 GHz AMD EPYC™ 7301 16-Core Processor a single core.

## 4 | RESULTS AND DISCUSSION

In this work, two variants of EOP-based CT energy model were considered that differ only in the treatment of the  $G_1$  terms: (i) the conventional DF from Equation (5) with the use of standard RI basis sets, referred here simply as to the “EOP” variant; (ii) the EDF method from Equation (16) with the use of optimal polarization auxiliary AO basis sets, referred here as to the “EOPm” variant.

### 4.1 | Accuracy

#### 4.1.1 | EOP variant

The root mean square errors (RMSE) of CT energy evaluations by using the OL, EFP2, and EOP formulations across various sets of bi-molecular systems are shown in Table 2, and the scatter plots against the reference (RVS) is shown in Figure 1 for the VVO basis, and Figure (S1) for the CVO basis. The overall accuracies of all the models are good and comparable within approximately 1 kcal/mol tolerance, with the OL/CVO method being the most accurate (RMSE of 1.13 kcal/mol), and EFP2/CVO and OL/VVO method being the least accurate (RMSE of about 2.6 kcal/mol). The EOP method performs on average slightly better than the EFP2 method by about 0.6 kcal/mol regardless of the type of virtual MOs (CVOs or VVOs) being used. Interestingly, although ideally the accuracies of the EOP method should be similar as the OL method, our results show that the accuracy of the EOP method is better when using the VVO basis, and worse when using the CVO basis, as compared to the OL method. Close inspection of this discrepancy shows that the ERI elimination technique works quantitatively well only for the overlap-like ERIs

**TABLE 2** Accuracy of approximate CT energy methods across wide range of bi-molecular complexes

Database	Root mean square error (kcal/mol)					
	CVO			VVO		
	OL	EOP	EFP2	OL	EOP	EFP2
HB6/04 <sup>a</sup>	0.85	0.86	1.94	2.06	1.63	0.49
DI6/04 <sup>a</sup>	0.71	0.51	0.71	1.42	1.14	0.90
CT7/04 <sup>a</sup>	0.94	1.37	5.44	4.06	1.52	5.19
WI7/04 <sup>b</sup>	0.03	0.05	0.04	0.04	0.04	0.04
PPS5/05 <sup>b</sup>	0.09	0.71	0.52	0.24	0.18	0.13
Ionic <sup>c</sup>	1.74	3.42	2.20	3.12	2.31	1.39
Total	1.13	2.07	2.58	2.55	1.60	2.20

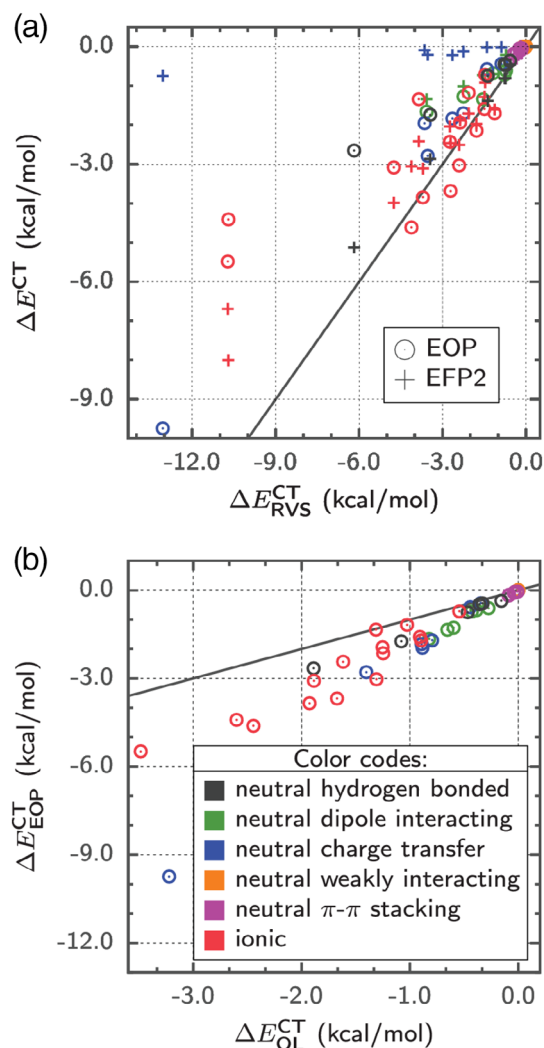
Note: RVS/6–311++G(d,p) CT energy is the reference. Abbreviations: “OL” - Otto and Ladik’s expression from Equation (4) in ref. [10]; “EOP” – this work, Equation (41); “EFP2” – ref. [17], here Equation (18). See Section 3 for more details about the AO basis set settings.

<sup>a</sup>Refs. [69–71].

<sup>b</sup>Refs. [69,70,99].

<sup>c</sup>Ref. [72].





**FIGURE 1** Performance of the EOP and EFP2 methods for the charge transfer (CT) interaction energy with VVOs in bi-molecular complexes. Structural datasets: The HB6/04 hydrogen bonding database (black),<sup>69–71</sup> the DI6/04 dipole interaction database (green),<sup>69–71</sup> the CT7/04 charge-transfer complex database (blue),<sup>69–71</sup> the WI7/05 weak interaction database (orange),<sup>69,70,99</sup> the PPS5/05 the  $\pi$ - $\pi$  stacking database (purple),<sup>69,70,99</sup> and the ionic systems database (red)<sup>72</sup>

( $G_1$  terms), whereas the contributions of the Coulomb-like ERIs ( $G_2$  and  $G_3$  terms) are typically overestimated relative to the corresponding OL results, especially when the VVOs are used. Therefore, the errors due to the LOA in  $G_2$  and  $G_3$  terms contribute to an accidental error cancellation in the VVO basis and loss of accuracy in the CVO basis. It is also worth noting that despite quite similar overall accuracy of EFP2 and EOP models in the VVO basis, only EOP model correctly describes the neutral charge-transfer complexes (compare blue dots with blue crosses in Figure 1(A)). However, in the case of H-bonded neutral systems as well as ionic systems (black and red markers in Figure 1(A), respectively) VVO EFP2 performs better than VVO EOP by roughly 1 kcal/mol.

The asymptotic dependence of the CT/VVO energy shown in Figure 2 indicates that in the region near the equilibrium and farther the EOP method (see solid green dots) is similarly accurate as the EFP2 method (see solid blue dots) except for the  $\text{H}_2\text{O} - \text{NH}_4^+$  system, in the case of which the EOP model overestimates the CT energies roughly by a factor of 2 whereas the EFP2 model well reproduces the RVS estimates. In the short-distance regions, the EFP2 method works quantitatively well for neutral systems, and slightly underestimates the CT energies for ionic systems, while the EOP model consistently underestimates the CT energies in all cases except for the  $\text{H}_2\text{O} - \text{NH}_4^+$  system. In contrast to the VVO basis, using the CVO basis with combination of the EOP formulation leads to unreasonably large overestimation of the CT energy in the ionic systems (see Figure [S2]). Therefore, the CVO basis is generally not suitable for the EOP model and, from that point onward, we focus our attention only on the VVO-based models.

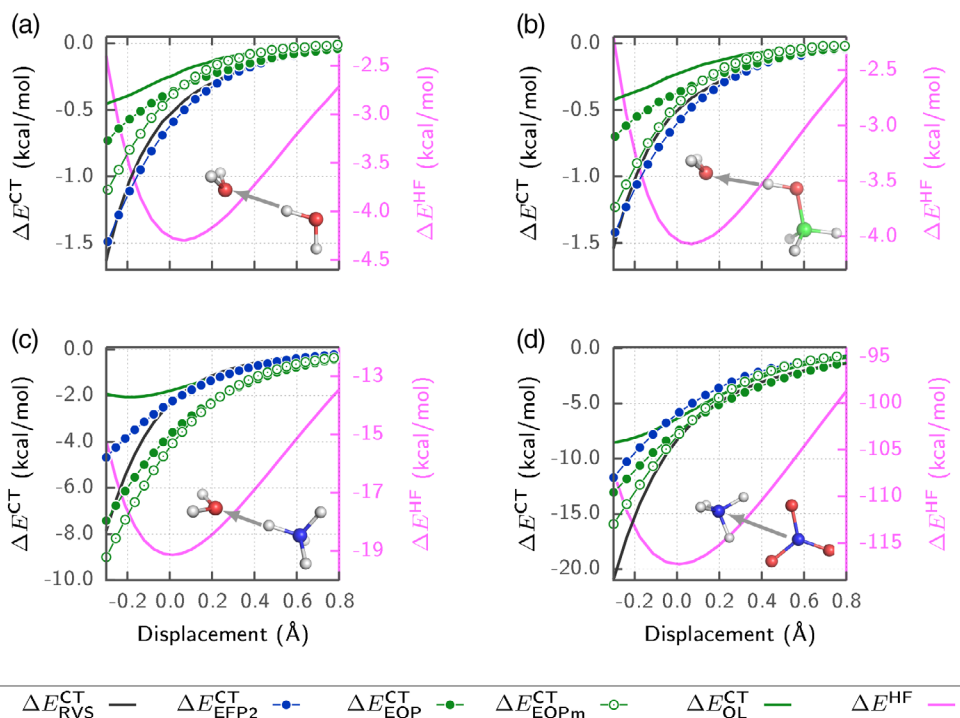
#### 4.1.2 | EOPm variant

Interestingly, using optimal polarization auxiliary basis sets as alternative (“EOPm” variant, green open circles in Figure 2) drastically improves the behavior of the EOP model in the short-distance region. This is likely due to the fruitful cancellation of errors upon truncation of the RI space in the  $G_1$  terms. Near the equilibrium and in the long-distance limit “EOP” and “EOPm” variants generate similar results.

#### 4.1.3 | Multi-fragment systems

From the above results one can conclude that the EFP2 model and the EOP CT model perform on overall comparably well when the VVO basis is used, except for the “EOP” variant in the short-range regions in the case of which CT energies are often underestimated. Nevertheless, accuracies near equilibrium geometries are acceptable for both “EOP” and “EOPm” variants. Therefore, we have applied the following three variants of the full EFP2 model for several multi-fragment model systems (Table 3): (i) original EFP2 formulation; (ii) original EFP2 but with the CT term replaced with the “EOP” variant—denoted as the EFP2<sub>EOP</sub> method; (iii) EFP2 with EOP-based term in “EOPm” variant—denoted as the EFP2<sub>EOPm</sub> method. As can be seen from Table 3, the accuracy of the EFP2<sub>EOPm</sub> method is good and comparable to the original EFP2 method in most of the cases studied. The total errors fall in the range of 2–4 kcal/mol as compared to the full HF results, except for the  $(\text{C}_2\text{mim}^+)_4(\text{NTf}_2^-)_4$  system where the EOP term introduced error of 7.1 kcal/mol. It is also found that the EFP2<sub>EOPm</sub> method performs rather similarly in  $(\text{H}_2\text{O})_{15}$  and  $(\text{DMSO})_9$  systems, and even slightly better in the  $(\text{C}_2\text{mim}^+)_4(\text{NTf}_2^-)_4$  system. For the latter, the error was reduced by 3 kcal/mol as compared to the EFP2<sub>EOP</sub> method, resulting in a total error of 4.1 kcal/mol. However, EFP2 method is more accurate in ionic liquid system, which is also consistent with Table 2 where the EFP2 model outperforms EOP

**FIGURE 2** Asymptotic dependence of the charge transfer energy in the CT formulation with VVOs for selected bi-molecular complexes. (A) Water dimer, (B) water-methanol complex, (C) water-ammonium complex, and (D) ammonium-nitrate complex, where one molecule has been translated along the vector specified in the insets relative to initial geometry, optimized at HF/6-31+G(d,p) level. The total interaction energy is also shown for comparison in purple color in this figure. All interaction energies were obtained at the HF/6-311++G(d,p) level of theory



**TABLE 3** Interaction energies and CPU timings for multi-fragment model systems

		(H <sub>2</sub> O) <sub>15</sub>		(DMSO) <sub>9</sub>		(C <sub>2</sub> mim <sup>+</sup> ) <sub>4</sub> (NTf <sub>2</sub> <sup>-</sup> ) <sub>4</sub>	
		ΔE	Time	ΔE	Time	ΔE	Time
Coul.	EFP2	-186.1		-66.3		-326.5	
Exch.-Rep.	EFP2	+126.3		+48.5		+56.6	
Ind.	EFP2	-45.2		-18.7		-21.6	
CT	EFP2	-12.9		-0.8		-1.3	
	EOP	-9.2		-3.4		-6.6	
	EOPm	-13.3		-2.1		-3.6	
Total	EFP2	-117.9	0.89	-37.3	14.2	-292.8	90.9
	EFP2 <sub>EOP</sub>	-114.2	0.11	-39.9	0.60	-298.1	2.7
	EFP2 <sub>EOPm</sub>	-118.3	0.07	-38.6	0.37	-295.1 <sup>a</sup>	2.4 <sup>a</sup>
Full HF		-115.0	-	-40.8	-	-291.0	-

Note: Interaction energies and CPU timings are given in kcal/mol and seconds, respectively. All the calculations were undertaken by using the 6-31++G(d,p) primary basis set and VVOs for the CT energy calculations. CPU timings for the reference RHF calculation results are assumed to be infinite due to a few orders of magnitude longer timings as compared to the EFP2 models studied here. See Section 3 for more details regarding the methods.

<sup>a</sup>Minimal polarization auxiliary AO basis applied only to C<sub>2</sub>mim<sup>+</sup> fragments.

in the ionic systems database set. Nevertheless, the general accuracy of the EOP and EOPm models is acceptable.

## 4.2 | Reduction of computational costs

The utmost goal of this work is to reduce the computational cost of the CT energy evaluation in the calculations involving effective fragment potentials. In Table 4, estimation of the computational cost of the EFP2 and EOP models is shown. It is apparent that

EFP2 requires much more quantities to be computed as compared to the EOP method (“Calculables” in the table). Clearly, evaluation of the EFP2 CT expression from Equation (18) involves quite a number of different types of OEIs. According to our estimations that assume sequential (two-step) two-index AO-MO transformations of OEI matrices and large AO basis sets, the computational cost is of an order of  $2p^3(s+t) + 3vop^2$ , where the  $o$  and  $p$  denote the number of occupied orbitals and the number of atomic basis functions, respectively. Here,  $s$ ,  $t$ , and  $v$  are the relative costs of evaluation of the overlap, kinetic energy, and multipole potential

**TABLE 4** Estimated computational cost of the EFP2 and EOP methods for calculation of CT energy

Method	EFP2	EOP
Constant Parameters <sup>a</sup>	$\epsilon_i$	$\epsilon_i, \epsilon_m, L_{ii}^A$
Superimposable Parameters <sup>a</sup>	$C_{ai}^A, C_{\beta n}^B, \{\alpha\}, \{\beta\}$	$C_{ai}^A, C_{\beta n}^B, \{\alpha\}, \{\beta\}, \{\eta\}, V_{n\eta}^{\text{eff},B}$
Calculables <sup>a</sup>	$S_{ij}, S_{nk}, S_{nw}, T_{nn}, T_{kj}, T_{vj}, T_{nj}, U_{in}^{\text{eff},B}, U_{ik}^{\text{eff},B}, U_{iw}^{\text{eff},B}$	$S_{ij}, S_{\eta i}, S_{ni}, U_i^{BA}, w_{yi}^{BA}$
Cost <sup>b</sup>	$sp(2p^2 + 2op + o^2) + tp$ $(2p^2 + 2op + o^2) + vop$ $(3p + o) + o^2p$	$sop(2p + o + a) +$ $op(a + oN + 2o)$

Note: Based on coupling constant expressions from Equation (19) and Equation (39) for EFP2 and EOP method, respectively.

<sup>a</sup>The subscript meaning is as follows: primary basis set functions of A:  $\alpha$ ; primary basis set functions of B:  $\beta$ ; auxiliary basis set functions of B:  $\eta$ , occupied MOs of A:  $i, i', k$ ; occupied MOs of B:  $j$ ; virtual MOs of A:  $w$ ; virtual MOs of B:  $n$ ; atoms of B:  $y$ . Analysis is based on  $E^{A-B}$  term.

<sup>b</sup>Numbers of: primary basis set functions— $p$ ; auxiliary basis set functions— $a$ ; occupied MOs— $o$ ; atoms— $N$ . Relative costs:  $v$ —multipole potential,  $t$ —kinetic energy; and  $s$ —overlap OELs. It was assumed that the number of virtual orbitals is equal to  $n$ .

OELs, respectively, with the latter being most expensive but necessary to compute  $U^{\text{eff}}$  matrices from Equation (21). On the contrary, EOP-based expression from Equation (41) requires only overlap OELs that are the least expensive, and has the cost of approximate magnitude of  $2so^2p^2$  for relatively small auxiliary basis sets. Note also that, among the calculables that are needed in each EOP-based CT energy evaluation, are the auxiliary vectors and matrices from Equations (40a) and (40b), the cost of which is negligible. The amount of effective fragment parameters, that needs to be superimposed during the calculations by applying rotation of orbitals and basis functions (“Superimposable parameters”) is rather the same in EFP2 and EOP models. This includes the LCAO-MO coefficients in canonical basis, and the primary basis set, with an addition of the auxiliary basis set for the EOP model. Therefore, the cost of parameter superimposition should not be significantly larger as in the EFP2 formulation, provided sufficiently small auxiliary basis set is used. For example, assuming a water dimer system and 6-311++G(d,p) primary and minimal auxiliary basis set with  $s = t = v \approx 1$ , the EOP CT method is predicted to be roughly 12–16 times faster than EFP2 CT method. In practice, the parameters  $t$  and  $v$  will have larger values, especially the latter.

Comparison of the time profiling of full EFP2 and EFP2<sub>EOP</sub> methods, shown in Table 3, reveals considerable CPU time savings when the EOP CT term is used. Based on our implementation, 8-, 23-, and 33-fold speed-ups in the case of  $(\text{H}_2\text{O})_{15}$ ,  $(\text{DMSO})_9$ , and  $(\text{C}_2\text{mim}^+)_4(\text{NTf}_2^-)_4$  systems were observed, respectively. The CPU time saving generally increases with the size of the fragment.

It was also found that using the “EOPm” variant leads to a further 50% reduction of the CPU time for the CT term evaluation as compared to the “EOP” variant. In effect, the CPU time for the total energy evaluation reduces by a factor of 1.6 when

using the EFP2<sub>EOPm</sub> method, since the EOP-based CT term is of comparable cost as the EFP2 exchange-repulsion term. As a result, the total CPU time required to evaluate total interaction energy with the EFP2<sub>EOP(m)</sub> methods gets dramatically reduced. The order of magnitude of these speed-ups is in good agreement with our theoretically predicted values based on Table 4 discussed above.

The time needed to compute the CT parameters in the preparatory EFP2 calculation is comparable when using the original EFP2 method and the EFP2<sub>EOPm</sub> method. However, minimization of the auxiliary basis set size through the EDF scheme, which is required for the “EOPm” variant, is more costly and, based on our current implementation, requires additional few CPU hours. Nevertheless, the overall CPU time required to compute the EOP CT parameters in the “EOPm” variant is still negligible as compared to the CPU time needed for a typical molecular dynamics simulation employing the EFP2 force-field.

### 4.3 | Memory requirements

The memory requirements for generating the EFP2 and EOP CT parameters in the “EOP” variant are virtually the same. Slightly more memory is required for generating parameters in the “EOPm” variant due to the additional storage of intermediate basin hopping solutions and steps during the gradient search when optimizing the auxiliary basis set. However, memory requirements are still much lesser than those to solve the coupled-perturbed Hartree-Fock (CPHF) equations,<sup>100,101</sup> which is necessary for the preparatory step of EFP2 parameters generation. Therefore, memory requirements are not expected to be larger as compared to the EFP2 in a typical implementation of the EOP and the EDF method.

## 5 | CONCLUSIONS

In this work, the EOP operator technique of eliminating ERIs in the fragment-based theories of intermolecular interactions was proposed. It was shown that in general case two types of EOPs can be defined and worked out either through the DF or the DMTP expansion. For the first group of EOPs, the DF was extended via the optimization of the auxiliary AO basis set to further reduce computational costs.

The EOP technique was then applied to calculate the CT energy in a variety of bi-molecular complexes as well as a few multi-fragment systems within the framework of the state-of-the-art EFP2 model. The presented validation of the EOP technique against the Otto-Ladik CT model as parent theory showed that in most cases elimination of ERIs is qualitatively correct and the proposed EOP-based Otto-Ladik model is of comparable accuracy to the CT formulation of the EFP2 model when the VVOs are used. More importantly, the EOP-based model significantly outperformed the original EFP2 model in terms of computational efficiency reaching overall speedups ranging from 8 to 38 times. On the other hand, the magnitude of the EOP-

based CT energy is found to be overestimated in ionic systems when using CVOs, which is not recommended.

Concluding, two new variants of the EFP2/VVO model with EOP-based CT term: (i) EFP2<sub>EOP</sub> and (ii) EFP2<sub>EOPm</sub> were introduced here, that are more efficient than the formulations proposed so far. The more efficient EFP2<sub>EOPm</sub> variant is recommended for small and medium fragments such as water and small organic molecules. For larger fragments, in the case of which optimization of auxiliary basis sets would be too costly, slightly slower EFP2<sub>EOP</sub> variant might be a better choice.

The proposed EOP-based model can be readily incorporated within the EFP2 method, strongly facilitating the complete (CT-including) EFP2 energy calculations and molecular dynamics simulations in large systems. The significant CPU time savings observed in our implementation demonstrate the high potential of the EOP-based ERI elimination technique and we expect comparable speed-ups for the subsequent implementations of our EOP CT formulation in other mainstream codes such as the GAMESS (US) program and the LIBEFP library.

Furthermore, the proposed EOP-based ERI elimination technique could in principle be used in virtually any other ab initio fragment-based approaches for condensed-phase simulations, where ERIs pose the computational challenge when confronted with the size of the system. Currently, we are applying the EOP technique developed in this work for efficient calculations of Frenkel excitonic Hamiltonians of electronically coupled chromophores, and the results will soon be published elsewhere.

## ACKNOWLEDGMENTS

This project is carried out under POLONEZ programme which has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 665778. This project is funded by National Science Centre, Poland (grant no. 2016/23/P/ST4/01720) within the POLONEZ 3 fellowship. Wrocław Centre for Networking and Supercomputing (WCSS) is acknowledged for granting computational resources.

## CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

## DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article and on request from the corresponding author. The computer code which contains implementation of the developed theory is available at <https://globulion.github.io/eopdev>.

## ORCID

Bartosz Błasiak  <https://orcid.org/0000-0003-1968-3465>

Marta Chołuj  <https://orcid.org/0000-0003-2461-4851>

Robert W. Góra  <https://orcid.org/0000-0003-0253-4295>

Wojciech Bartkowiak  <https://orcid.org/0000-0002-3442-9302>

## REFERENCES

[1] P. Otto, J. Ladik, *Int. J. Quant. Chem.* **1980**, *18*, 1143.

- [2] R. F. Frey, E. R. Davidson, *J. Chem. Phys.* **1989**, *90*, 5555.  
 [3] B. Jeziorski, R. Moszyński, K. Szalewicz, *Chem. Rev.* **1994**, *94*, 1887.  
 [4] A. J. Stone, *The Theory of Intermolecular Forces*, Clarendon Press, Oxford **1997**.  
 [5] A. J. Stone, A. J. Misquitta, *Chem. Phys. Lett.* **2009**, *473*, 201.  
 [6] A. Devarajan, T. L. Windus, M. S. Gordon, *J. Phys. Chem. A* **2011**, *115*, 13987.  
 [7] A. Devarajan, A. Gaenko, M. S. Gordon, T. L. Windus, *Nucleation Using the Effective Fragment Potential and Two-Level Parallelism*, John Wiley Sons, Ltd, Hoboken **2017**, chap. 7, p. 209.  
 [8] C. C. J. Roothaan, *Rev. Mod. Phys.* **1951**, *23*, 69.  
 [9] J. H. Jensen, *J. Chem. Phys.* **2001**, *114*, 8775.  
 [10] P. Otto, J. Ladik, *Chem. Phys.* **1975**, *8*, 192.  
 [11] J. N. Murrell, M. Randić, D. R. Williams, H. C. Longuet-Higgins, *Proc. R. Soc. Lond. A* **1965**, *284*, 566.  
 [12] M. S. Gordon, D. G. Fedorov, S. R. Pruitt, L. V. Slipchenko, *Chem. Rev.* **2012**, *112*, 632.  
 [13] O. Demerdash, Y. Mao, T. Liu, M. Head-Gordon, T. Head-Gordon, *J. Chem. Phys.* **2017**, *147*, 161721.  
 [14] O. Demerdash, E.-H. Yap, T. Head-Gordon, *Annu. Rev. Phys. Chem.* **2014**, *65*, 149.  
 [15] P. Xu, E. B. Guidez, C. Bertoni, M. S. Gordon, *J. Chem. Phys.* **2018**, *148*, 090901.  
 [16] P. N. Day, J. H. Jensen, M. S. Gordon, S. P. Webb, W. J. Stevens, M. Krauss, D. Garmer, H. Basch, D. Cohen, *J. Chem. Phys.* **1996**, *105*, 1968.  
 [17] H. Li, M. S. Gordon, J. H. Jensen, *J. Chem. Phys.* **2006**, *124*, 214108.  
 [18] P. Xu, M. S. Gordon, *J. Chem. Phys.* **2013**, *139*, 194104.  
 [19] M. S. Gordon, Q. A. Smith, P. Xu, L. V. Slipchenko, *Annu. Rev. Phys. Chem.* **2013**, *64*, 553.  
 [20] K. A. Nguyen, R. Pachter, P. N. Day, *J. Chem. Phys.* **2014**, *140*, 244101.  
 [21] M. K. Ghosh, S. G. Cho, C. H. Choi, *J. Phys. Chem. B* **2014**, *118*, 4876.  
 [22] N. Kuroki, H. Mori, *Chem. Lett.* **2016**, *45*, 1009.  
 [23] T. Sattasathuchana, P. Xu, M. S. Gordon, *J. Phys. Chem. A* **2019**, *123*, 8460.  
 [24] N. Gresh, P. Claverie, A. Pullman, *Theoret. Chim. Acta* **1984**, *66*, 1.  
 [25] J.-P. Piquemal, H. Chevreau, N. Gresh, *J. Chem. Theory Comput.* **2007**, *3*, 824.  
 [26] W. Xie, J. Gao, *J. Chem. Theory Comput.* **2007**, *3*, 1890.  
 [27] R. Chaudret, N. Gresh, C. Narth, L. Lagardère, T. A. Darden, G. A. Cisneros, J.-P. Piquemal, *J. Phys. Chem. A* **2014**, *118*, 7598.  
 [28] E. E. Dahlke, D. G. Truhlar, *J. Chem. Theory Comput.* **2008**, *4*, 1.  
 [29] H. R. Leverentz, K. A. Maerzke, S. J. Keasler, J. I. Siepmann, D. G. Truhlar, *Phys. Chem. Chem. Phys.* **2012**, *14*, 7669.  
 [30] J. H. Jensen, M. S. Gordon, *Mol. Phys.* **1996**, *89*, 1313.  
 [31] J. H. Jensen, M. S. Gordon, *J. Chem. Phys.* **1998**, *108*, 4772.  
 [32] H. Li, H. M. Netzloff, M. S. Gordon, *J. Chem. Phys.* **2006**, *125*, 194103.  
 [33] I. Adamovic, M. S. Gordon, *Mol. Phys.* **2005**, *103*, 379.  
 [34] P. Xu, F. Zahariev, M. S. Gordon, *J. Chem. Theory Comput.* **2014**, *10*, 1576.  
 [35] L. V. Slipchenko, M. S. Gordon, *J. Comput. Chem.* **2007**, *28*, 276.  
 [36] C. H. L. Błasiak, L. J. Webb, M. Cho, *Acc. Chem. Res.* **2017**, *50*, 968.  
 [37] D. Ghosh, D. Kosenkov, V. Vanovschi, C. F. Williams, J. M. Herbert, M. S. Gordon, M. W. Schmidt, L. V. Slipchenko, A. I. Krylov, *J. Phys. Chem. A* **2010**, *114*, 12739.  
 [38] P. K. Gurunathan, A. Acharya, D. Ghosh, D. Kosenkov, I. Kaliman, Y. Shao, A. I. Krylov, L. V. Slipchenko, *J. Phys. Chem. B* **2016**, *120*, 6562.  
 [39] B. Błasiak, A. W. Ritchie, L. J. Webb, M. Cho, *Phys. Chem. Chem. Phys.* **2016**, *18*, 18094.  
 [40] R. J. Xu, B. Błasiak, M. Cho, J. P. Layfield, C. H. Londergan, *J. Phys. Chem. Lett.* **2018**, *9*, 2560.

- [41] C. Møller, M. S. Plesset, *Phys. Rev.* **1934**, *46*, 618.
- [42] M. W. Schmidt, K. K. Baldrige, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, *J. Comput. Chem.* **1993**, *14*, 1347.
- [43] W. C. Lu, C. Z. Wang, M. W. Schmidt, L. Bytautas, K. M. Ho, K. Ruedenberg, *J. Chem. Phys.* **2004**, *120*, 2629.
- [44] T. Smith, L. V. Slipchenko, M. S. Gordon, *J. Phys. Chem. A* **2008**, *112*, 5286.
- [45] Q. A. Smith, M. S. Gordon, L. V. Slipchenko, *J. Phys. Chem. A* **2011**, *115*, 11269.
- [46] Q. A. Smith, M. S. Gordon, L. V. Slipchenko, *J. Phys. Chem. A* **2011**, *115*, 4598.
- [47] Š. Budzák, A. D. Laurent, C. Laurence, M. Medved, D. Jacquemin, *J. Chem. Theory Comput.* **2016**, *12*, 1919.
- [48] T. Giovannini, P. Lafiosca, C. Cappelli, *J. Chem. Theory Comput.* **2017**, *13*, 4854.
- [49] N. Kuroki, H. Mori, *J. Phys. Chem. B* **2019**, *123*, 194.
- [50] I. A. Kaliman, L. V. Slipchenko, *J. Comput. Chem.* **2015**, *36*, 129.
- [51] D. Ghosh, D. Kosenkov, V. Vanovschi, J. Flick, I. Kaliman, Y. Shao, A. T. Gilbert, A. I. Krylov, L. V. Slipchenko, *J. Comput. Chem.* **2013**, *34*, 1060.
- [52] P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, *136*, B864.
- [53] W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133.
- [54] A. Holas, N. H. March, *Phys. Rev. A* **1991**, *44*, 5521.
- [55] W. Weber, W. Thiel, *Theor. Chem. Acc.* **2000**, *103*, 495.
- [56] F. Neese, *J. Chem. Phys.* **2005**, *122*, 034107.
- [57] G. A. Cisneros, J.-P. Piquemal, T. A. Darden, *J. Chem. Phys.* **2005**, *123*, 044109.
- [58] J.-P. Piquemal, G. A. Cisneros, P. Reinhardt, N. Gresh, T. A. Darden, *J. Chem. Phys.* **2006**, *124*, 104101.
- [59] B. Błasiak, H. Lee, M. Cho, *J. Chem. Phys.* **2013**, *044111*, 139.
- [60] B. Błasiak, M. Maj, M. Cho, R. W. Góra, *J. Chem. Theory Comput.* **2015**, *11*, 3259.
- [61] A. Heßelmann, G. Jansen, M. Schütz, *J. Chem. Phys.* **2005**, *122*, 014103.
- [62] R. Polly, H.-J. Werner, F. R. Manby, P. J. Knowles, *Mol. Phys.* **2004**, *102*, 2311.
- [63] W. J. Stevens, W. H. Fink, *Chem. Phys. Lett.* **1987**, *139*, 15.
- [64] C. Etchebest, R. Lavery, A. Pullman, *Theoret. Chim. Acta* **1982**, *62*, 17.
- [65] R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford, C. D. Sherrill, *Journal of Chemical Theory and Computation* **2017**, *13*, 3185.
- [66] S. F. Boys, *Rev. Mod. Phys.* **1960**, *32*, 296.
- [67] J. Pipek, P. G. Mezey, *J. Chem. Phys.* **1989**, *90*, 4916.
- [68] C. C. J. Roothaan, *Rev. Mod. Phys.* **1960**, *32*, 179.
- [69] Y. Zhao, N. E. Schultz, D. G. Truhlar, *J. Chem. Phys.* **2005**, *123*, 161103.
- [70] Y. Zhao, N. E. Schultz, D. G. Truhlar, *J. Chem. Theory Comput.* **2006**, *2*, 364.
- [71] Y. Zhao, D. G. Truhlar, *J. Chem. Theory Comput.* **2005**, *1*, 415.
- [72] J. Řezáč, P. Hobza, *J. Chem. Theory Comput.* **2012**, *8*, 141.
- [73] T. Clark, J. Chandrasekhar, G. W. Spitznagel, P. V. R. Schleyer, *J. Comput. Chem.* **1983**, *4*, 294.
- [74] R. Ditchfield, W. J. Hehre, J. A. Pople, *J. Chem. Phys.* **1971**, *54*, 724.
- [75] P. C. Hariharan, J. A. Pople, *Theoret. Chim. Acta* **1973**, *28*, 213.
- [76] W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257.
- [77] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A., Jr. Montgomery, J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, *Gaussian 16 Revision C.01*, Gaussian Inc, Wallingford CT **2016**.
- [78] M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, J. A. Pople, *J. Chem. Phys.* **1982**, *77*, 3654.
- [79] P. M. Gill, B. G. Johnson, J. A. Pople, M. J. Frisch, *Chem. Phys. Lett.* **1992**, *197*, 499.
- [80] R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *J. Chem. Phys.* **1980**, *72*, 650.
- [81] A. D. McLean, G. S. Chandler, *J. Chem. Phys.* **1980**, *72*, 5639.
- [82] G. W. Spitznagel, T. Clark, P. von Ragué Schleyer, W. J. Hehre, *J. Comput. Chem.* **1987**, *8*, 1109.
- [83] F. Weigend, *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285.
- [84] F. Weigend, A. Köhn, C. Hättig, *J. Chem. Phys.* **2002**, *116*, 3175.
- [85] M. S. Gordon, J. S. Binkley, J. A. Pople, W. J. Pietro, W. J. Hehre, *J. Am. Chem. Soc.* **1982**, *104*, 2797.
- [86] G. Chalaśiński, M. Gutowski, *Mol. Phys.* **1985**, *54*, 1173.
- [87] W. Kabsch, *Acta Cryst. A* **1976**, *32*, 922.
- [88] W. Kabsch, *Acta Cryst. A* **1978**, *34*, 827.
- [89] A. Stone, M. Alderton, *Mol. Phys.* **1985**, *56*, 1047.
- [90] A. J. Stone, *J. Chem. Theory Comput.* **2005**, *1*, 1128.
- [91] V. Kairys, J. H. Jensen, *Chem. Phys. Lett.* **1999**, *315*, 140.
- [92] K. T. Tang, J. P. Toennies, *J. Chem. Phys.* **1984**, *80*, 3726.
- [93] Z. Li, H. A. Scheraga, *Proc. Nat. Acad. Sci.* **1987**, *84*, 6611.
- [94] D. J. Wales, J. P. K. Doye, *J. Phys. Chem. A* **1997**, *101*, 5111.
- [95] D. J. Wales, H. A. Scheraga, *Science* **1999**, *285*, 1368.
- [96] D. J. Wales, *Energy Landscapes*, Cambridge University Press, Cambridge, UK **2003**.
- [97] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimman, I. Henriksen, E. A. Qunitero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, *Nature Methods* **2020**, *17*(3), 261.
- [98] D. Kraft, Technical Report DFVLR-FB 88-28, **1988**.
- [99] Y. Zhao, D. G. Truhlar, *J. Phys. Chem. A* **2005**, *109*, 5656.
- [100] R. McWeeny, *Rev. Mod. Phys.* **1960**, *32*, 335.
- [101] L. Dodds, R. McWeeny, A. J. Sadlej, *Mol. Phys.* **1977**, *34*, 1779.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Błasiak B, Bednarska JD, Chołuj M, Góra RW, Bartkowiak W. Ab initio effective one-electron potential operators: Applications for charge-transfer energy in effective fragment potentials. *J Comput Chem.* 2021;42: 398–411. <https://doi.org/10.1002/jcc.26462>

**APPENDIX: A GENERAL FORM OF INTERFRAGMENT-SEPARABLE ERI FUNCTIONALS**

Here, we derive Equation (3) which defines the interfragment ERI elimination technique. To generate EOP matrix elements out of interfragment ERIs, the ERI functional must be linear with respect to ERIs and separable in between fragments so that the effective potential operators involving orbitals of same fragment can be defined. Here we consider a general example of such ERI functional, that involves a generalized Coulomb and exchange operators acting only within the same fragment orbital subspace. For the case of the  $\langle BA|AA \rangle$  ERI class,

$$\{\mathcal{F}_{t_j}[\langle BA|AA \rangle]\}_{ij} = \alpha_j^{A,(t_j)} \sum_{kl \in A} Q_{kl}^{A,(t_j)} f_{t_j}^A \left[ \langle i^B j^A | k^A l^A \rangle \right] \quad (\text{A1a})$$

$$\{\mathcal{F}_{t_k}[\langle BA|AA \rangle]\}_{ij} = \alpha_k^{A,(t_k)} \sum_{kl \in A} Q_{kl}^{A,(t_k)} f_{t_k}^A \left[ \langle i^B k^A | j^A l^A \rangle \right], \quad (\text{A1b})$$

where the linear functional of the electron–electron repulsion operator  $|r_1 - r_2|^{-1}$  is

$$f_t^A \left[ \langle i^B j^A | k^A l^A \rangle \right] \equiv \iint d\mathbf{r}_1 d\mathbf{r}_2 \phi_i^B(\mathbf{r}_1) \phi_j^{A*}(\mathbf{r}_1) f_t^A \left[ \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \right] \phi_k^A(\mathbf{r}_2) \phi_l^{A*}(\mathbf{r}_2). \quad (\text{A2})$$

Clearly, the above functionals are linear with respect to ERIs, and separable in between fragments A and B so that EOP operator of sole fragment A can be defined. The vectors  $\alpha_j^{A,(t_j)}$  and  $\alpha_k^{A,(t_k)}$ , matrices  $Q_{kl}^{A,(t_j)}$ , and  $Q_{kl}^{A,(t_k)}$ , as well as functionals  $f_{t_j}^A$  and  $f_{t_k}^A$  are arbitrary as long as they depend solely on fragment A. For instance, it immediately follows that

$$\{\mathcal{F}_{t_j}[\langle BA|AA \rangle]\}_{ij} = \langle i^B | \hat{v}_j^{\text{eff},A} | j^A \rangle \quad (\text{A3})$$

with

$$v_j^{\text{eff},A}(\mathbf{r}_1) \equiv \alpha_j^{A,(t_j)} \sum_{kl \in A} Q_{kl}^{A,(t_j)} \int d\mathbf{r}_2 f_{t_j}^A \left[ \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \right] \phi_k^A(\mathbf{r}_2) \phi_l^{A*}(\mathbf{r}_2). \quad (\text{A4})$$

Now consider the sum over Coulomb and exchange functionals,  $\mathcal{F}_{t_j} + \mathcal{F}_{t_k}$ . When ERIs are represented in the same basis a unified EOP

operator cannot be defined. However, one can make use of another (arbitrary) basis via the unitary transformation  $\mathbf{U}$  such that

$$\phi_{i\alpha}(\mathbf{r}) = \sum_{\mu \in A} U_{j\mu}^A \phi_{\mu}^A(\mathbf{r}), \quad (\text{A5})$$

and expand  $j$  and  $k$  orbitals in the expressions for the Coulomb and exchange functionals, respectively. Such basis sets might be for example molecular orbitals and atomic basis functions. Now, the summations over ERIs can be combined and after a suitable interchange of dummy summation variables the EOP is given by

$$\{\mathcal{F}_{t_j} + \mathcal{F}_{t_k}\}_{ij} = \sum_{\mu \in A} \langle i^B | \hat{v}_j^{\text{eff},A,(t)} | \mu^A \rangle, \quad (\text{A6})$$

where the EOP operator reads

$$v_j^{\text{eff},A,(t)}(\mathbf{r}_1) \equiv \sum_{kl \in A} \int d\mathbf{r}_2 \left\{ \alpha_j^{A,(t_j)} Q_{kl}^{A,(t_j)} f_{t_j}^A \left[ \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \right] \phi_k^A(\mathbf{r}_2) \phi_l^{A*}(\mathbf{r}_2) U_{j\mu}^A + \alpha_j^{A,(t_k)} Q_{kl}^{A,(t_k)} f_{t_k}^A \left[ \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \right] \phi_j^A(\mathbf{r}_2) \phi_l^{A*}(\mathbf{r}_2) U_{k\mu}^A \right\}. \quad (\text{A7})$$

Defining EOPs for the  $\langle BB|AA \rangle$  ERI class allows the following functionals

$$\{\mathcal{F}_{t_j}[\langle BB|AA \rangle]\}_{ij} = \sum_{kl \in A} Y_{kl}^{A,(t)} f_t^A \left[ \langle i^B j^B | k^A l^A \rangle \right], \quad (\text{A8})$$

( $Y_{kl}^{A,(t)}$  matrices being arbitrary) which immediately can be recast via an EOP

$$\{\mathcal{F}_{t_j}[\langle BB|AA \rangle]\}_{ij} = \langle i^B | \hat{v}_j^{\text{eff},A,(t)} | j^B \rangle \quad (\text{A9})$$

with

$$v_j^{\text{eff},A}(\mathbf{r}_1) \equiv \sum_{kl \in A} Y_{kl}^{A,(t)} \int d\mathbf{r}_2 f_t^A \left[ \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \right] \phi_k^A(\mathbf{r}_2) \phi_l^{A*}(\mathbf{r}_2). \quad (\text{A10})$$

By summing over  $t$  (all distinct functionals), one arrives to two-electron part of Equation (3). The derivation of the one-electron part of this equation is straightforward.