The Journal of Chemical Physics

Scalable *ab initio* fragmentation methods based on a truncated expansion of the non-orthogonal molecular orbital model

Cite as: J. Chem. Phys. **155**, 154101 (2021); https://doi.org/10.1063/5.0064864
Submitted: 27 July 2021 • Accepted: 20 September 2021 • Published Online: 15 October 2021

Bryce M. Westheimer and Mark S. Gordon

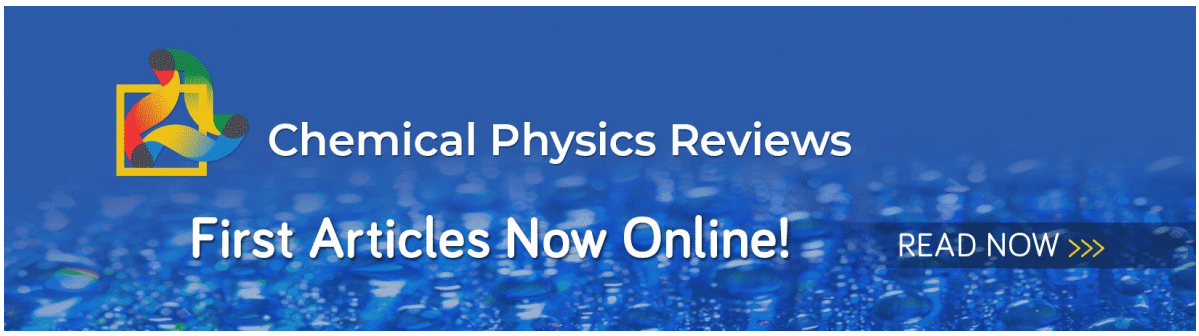
View Online Export Citation CrossMark

ARTICLES YOU MAY BE INTERESTED IN

Solving the Schrödinger equation using program synthesis
The Journal of Chemical Physics **155**, 154102 (2021); https://doi.org/10.1063/5.0062497

What the foundations of quantum computer science teach us about chemistry
The Journal of Chemical Physics **155**, 150901 (2021); https://doi.org/10.1063/5.0060367

Toward Laplace MP2 method using range separated Coulomb potential and orbital selective virtuals
The Journal of Chemical Physics **155**, 154104 (2021); https://doi.org/10.1063/5.0060099



AIP Publishing

© 2021 Author(s).

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Scalable *ab initio* fragmentation methods based on a truncated expansion of the non-orthogonal molecular orbital model

Cite as: J. Chem. Phys. 155, 154101 (2021); doi: 10.1063/5.0064864
Submitted: 27 July 2021 • Accepted: 20 September 2021 •
Published Online: 15 October 2021

View Online Export Citation CrossMark

Bryce M. Westheimer[1,2,a)] and Mark S. Gordon[1,2,b)]

AFFILIATIONS

[1] Department of Chemistry, Iowa State University, Ames, Iowa 50011, USA
[2] Ames Laboratory, Ames, Iowa 50011, USA

[a)] **Email:** bryce@si.msg.chem.iastate.edu
[b)] **Author to whom correspondence should be addressed:** mark@si.msg.chem.iastate.edu

ABSTRACT

An alternative formulation of the non-orthogonal molecular orbital model of electronic structure theory is developed based on the expansion of the inverse molecular orbital overlap matrix. From this model, a hierarchy of *ab initio* fragment-based quantum chemistry methods, referred to as the *n*th-order expanded non-orthogonal molecular orbital methods, are developed using a minimal number of approximations, each of which is frequently employed in intermolecular interaction theory. These novel methods are compared to existing fragment-based quantum chemistry methods, and the implications of those significant differences, where they exist, between the methods developed herein and those already existing methods are examined in detail. Computational complexities and theoretical scaling are also analyzed and discussed. Future extensions for the hierarchy of methods, to account for additional intrafragment and interfragment interactions, are outlined.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0064864

I. INTRODUCTION

The last century of development in theoretical models of electronic structure theory as well as computational science has led to the ability to quantitatively study a large number of systems using computational chemistry. These systems, however, span a small subspace of the full chemical space due to the limitations in system size that existing quantitatively accurate methods impose. Therefore, one major focus of research in the field of theoretical quantum chemistry in recent decades has been the development of various approximations, which may be employed to develop models and corresponding methods that may be applied to realistic chemical systems. In general, these approximations all seek to reduce the scaling of a given method with respect to the system size. Many approaches focus on introducing approximations into conventional quantum chemistry methods to reduce scaling orders or eliminate computationally complex terms.

Fragment-based quantum chemistry may be considered as one of the many approaches to reduced scaling quantitatively accurate quantum chemistry. In these models, a full, realistic chemical system is partitioned into smaller subsets whose calculations are tractable at a certain level of theory. Many methods have been developed using the fragmentation model, and interested readers are directed to a number of excellent reviews for further information on the subject.[1–7] Although existing fragmentation methods have made the simulation of large chemical systems possible, the goal of quantitative accuracy is often not realized. In many cases, this is due to the large number of approximations that are introduced in the development of a given fragmentation method.

The focus of this work is on the development of a novel model of fragment-based electronic structure theory and the resulting hierarchy of variational methods, which may be derived from this model. These methods are developed by employing as few approximations as possible in order to attain the linear scaling needed to make the resultant methods viable for the study of realistic systems while maintaining the accuracy needed to be applicable to the study of real systems.

Published under an exclusive license by AIP Publishing

II. BACKGROUND

A. Non-orthogonal molecular orbital theory

The problem of solving for the variationally minimized set of non-orthogonal molecular orbitals (MOs) has been well-studied and detailed elsewhere,[8–21] so only a brief overview of the conventional non-orthogonal molecular orbital (NOMO) theory is given here.

For a closed-shell system of $2n$ electrons in n corresponding non-orthogonal linearly independent spatial MOs, given by the set $\{\phi_i\}_{i\in n}$, the expression for the energy of the system is given by

$$E = \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{Z}_{ij} h_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} \boldsymbol{Z}_{ij}\boldsymbol{Z}_{kl}\left(2J_{ij}^{kl} - K_{ij}^{kl}\right), \tag{1}$$

where the core Hamiltonian matrix elements, h, are given by

$$h_{ij} = \left(\phi_i\middle|\hat{h}\middle|\phi_j\right), \tag{2}$$

the Coulomb electron repulsion integral (ERI) elements, J, are given by

$$J_{ij}^{kl} = \left(\phi_i\phi_j\middle|r_{12}^{-1}\middle|\phi_k\phi_l\right), \tag{3}$$

and the exchange ERIs, K, are given by

$$K_{ij}^{kl} = \left(\phi_i\phi_j\middle|r_{12}^{-1}\middle|\phi_k\phi_j\right), \tag{4}$$

where r_{12} is the inter-electron distance and the $\boldsymbol{Z}_{ij}$ are elements of the inverse of the MO overlap matrix, $\boldsymbol{S}$, whose elements, $\boldsymbol{S}_{ij}$, are the MO overlap integrals, s_{ij},

$$\boldsymbol{S}_{ij} = (\phi_i|\phi_j) = s_{ij}, \tag{5}$$

$$\boldsymbol{Z}_{ij} = \left(\boldsymbol{S}^{-1}\right)_{ij}. \tag{6}$$

One may then partition the MOs into M disjoint closed-shell subsets or fragments. Thus, for the Ith fragment, there are n_I corresponding MOs, and therefore $2n_I$ electrons, assigned to it. For each fragment, orbitals assigned to the fragment are denoted with a superscript representing the fragment index. For example, for the Ith fragment, the set of MOs assigned to the fragment is given by

$$\{\phi_i\}_{i\in I} = \left\{\phi_i^I\right\}_{i\in I}. \tag{7}$$

The original set of all n MOs may then be written as a union of all fragment MOs,

$$\{\phi_\mu\}_{\mu=1}^{\mu=n} = \bigcup_{I=1}^{M}\left\{\phi_i^I\right\}_{i\in I}. \tag{8}$$

The energy may then be rewritten in the fragmentation representation as

$$E = \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J} \boldsymbol{Z}_{iI,jJ} h_{iI,jJ} + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L=1}^{M}\sum_{l=L}^{n_L} \times \boldsymbol{Z}_{iI,jJ}\boldsymbol{Z}_{kK,lL}\left(2J_{iI,jJ}^{kK,lL} - K_{iI,jJ}^{kK,lL}\right). \tag{9}$$

For the sake of simplicity, it is assumed that all MOs within a given fragment are expressed in the same atomic orbital (AO) basis. Then, without loss of generality, it may be assumed that all MOs assigned to a given fragment are mutually orthonormal,

$$\left(\phi_i^I\middle|\phi_{i'}^J\right) = \delta_{ii'}\,\forall\, i,\ i' \in I;\ I \in M. \tag{10}$$

The MOs in different fragments, however, will not, in general, be orthogonal,

$$\left(\phi_i^I\middle|\phi_j^J\right) = s_{iI,jJ}\,\forall\, i \in I,\ j \in J;\ I \neq J. \tag{11}$$

Utilizing the identity

$$\delta(\boldsymbol{SZ}) = \delta\boldsymbol{I}, \tag{12}$$

one can derive the functional derivative of the total energy expression with respect to an arbitrary variation of a particular MO.

Following manipulation of the resultant equations, one obtains the general expression for the optimal MOs,

$$(1-\hat{p})\widehat{\boldsymbol{F}}\phi = 0, \tag{13}$$

where $\widehat{\boldsymbol{F}}$ is the non-orthogonal form of the Fock operator,

$$\widehat{\boldsymbol{F}} = h + \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{Z}_{ij}\left(2\hat{J}_{ij} - \widehat{K}_{ij}\right), \tag{14}$$

and $\hat{p}$ is a projection operator onto the space spanned by all occupied MOs of the total system,

$$\hat{p} = \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{Z}_{ij}|\phi_i\rangle\langle\phi_j|. \tag{15}$$

The expression in Eq. (13) may be transformed in a number of ways, depending on what constraints will be imposed on the orbitals, to generate various eigenvalue equations for the optimal orbitals. One of the earliest working methods using this approach was that developed by Stoll *et al.*, who used the following form of the eigenvalue equation to determine the optimal orbitals for each subsystem:[15–17]

$$\left[\widehat{\boldsymbol{F}} - \left(\widehat{\boldsymbol{F}}\hat{p}_{(I)} + \hat{p}_{(I)}^{\dagger}\widehat{\boldsymbol{F}}\right) + \hat{p}_{(I)}^{\dagger}\widehat{\boldsymbol{F}}\hat{p}_{(I)}\right]\phi_i^I = \epsilon_i^I\phi_i^I. \tag{16}$$

In Eq. (16), the superscript dagger † indicates the transpose. Without going into detail, it is sufficient to note that the projection operator in Eq. (16) is fragment-specific and the Fock operator is the total non-orthogonal Fock operator for the whole system, given in Eq. (14), rather than the conventional Fock operator of orthonormal MO theory. Utilizing this formalism, it is possible to obtain a set of fragment wave functions, each of which may be expanded in a unique basis set.

III. THEORY

A. Expanded non-orthogonal molecular orbital theory

This work focuses on the development of a theory for closed-shell systems of $2n$ electrons, with n corresponding MOs. Extension

Published under an exclusive license by AIP Publishing

of the methods derived herein to open-shell and multi-reference NOMO methods is straightforward and will be presented in a future publication.

1. Expansion of the molecular orbital overlap matrix

First, for convenience, the inverse MO overlap matrix, $\boldsymbol{Z}$, in Eq. (6) is transformed into an alternative representation. To begin, an auxiliary MO overlap matrix, $\boldsymbol{S}'$, is defined as follows:

$$\boldsymbol{S}' = \boldsymbol{S} - \boldsymbol{I}, \tag{17}$$

where $\boldsymbol{S}$ is the MO overlap matrix given in Eq. (5) and $\boldsymbol{I}$ is the identity matrix. The $\boldsymbol{S}'$ matrix is simply the original MO overlap matrix with diagonal elements set equal to zero. Given the new matrix, the $\boldsymbol{Z}$ matrix is rewritten in its power series expansion form as follows:

$$\boldsymbol{Z} = \left(\boldsymbol{I} + \boldsymbol{S}'\right)^{-1} = \boldsymbol{I} - \boldsymbol{S}' + \left(\boldsymbol{S}'\right)^2 - \left(\boldsymbol{S}'\right)^3 + \cdots = \boldsymbol{I} - \boldsymbol{\Delta Z}. \tag{18}$$

The last equality in Eq. (18) is written in order to isolate terms that depend on the interfragment MO overlap. The exact NOMO energy, as it was given in Eq. (9), may then be written in the form

$$\begin{aligned} = 2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}(\boldsymbol{I}-\boldsymbol{\Delta Z})_{iI,jJ}h_{iI,jJ} + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L=1}^{M}\sum_{l\in L}^{n_L} \\ \times (\boldsymbol{I}-\boldsymbol{\Delta Z})_{iI,jJ}(\boldsymbol{I}-\boldsymbol{\Delta Z})_{kK,lL}\left(2J_{iI,jJ}^{kK,lL} - K_{iI,jJ}^{kK,lL}\right). \end{aligned} \tag{19}$$

Expanding the matrix products in Eq. (19) produces the following expression for the total energy:

$$\begin{aligned} E = {} & 2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\boldsymbol{I}_{iI,jJ}h_{iI,jJ} - 2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\boldsymbol{\Delta Z}_{iI,jJ}h_{iI,jJ} \\ & + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L=1}^{M}\sum_{l\in L}^{n_L}\boldsymbol{I}_{iI,jJ}\boldsymbol{I}_{kK,lL}\left(2J_{iI,jJ}^{kK,lL} - K_{iI,jJ}^{kK,lL}\right) \\ & - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L=1}^{M}\sum_{l\in L}^{n_L}\boldsymbol{I}_{iI,jJ}\boldsymbol{\Delta Z}_{kK,lL}\left(2J_{iI,jJ}^{kK,lL} - K_{iI,jJ}^{kK,lL}\right) \\ & - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L=1}^{M}\sum_{l\in L}^{n_L}\boldsymbol{\Delta Z}_{iI,jJ}\boldsymbol{I}_{kK,lL}\left(2J_{iI,jJ}^{kK,lL} - K_{iI,jJ}^{kK,lL}\right) \\ & + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L=1}^{M}\sum_{l\in L}^{n_L}\boldsymbol{\Delta Z}_{iI,jJ}\boldsymbol{I}_{kK,lL}\left(2J_{iI,jJ}^{kK,lL} - K_{iI,jJ}^{kK,lL}\right). \end{aligned} \tag{20}$$

Using the properties of the identity matrix, the first, third, fourth, and fifth terms may be simplified by restricting summations

$$\begin{aligned} E = {} & 2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}h_{iI,iI} - 2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\boldsymbol{\Delta Z}_{iI,jJ}h_{iI,jJ} + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left(2J_{iI,iI}^{kK,kK} - K_{iI,iI}^{kK,kK}\right) - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{k=K}^{n_K}\sum_{L=1}^{M}\sum_{l\in L}^{n_L}\boldsymbol{\Delta Z}_{kK,lL}\left(2J_{iI,iI}^{kK,kK} - K_{iI,iI}^{kK,lL}\right) \\ & - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\boldsymbol{\Delta Z}_{iI,jJ}\left(2J_{iI,iI}^{kK,kK} - K_{iI,jJ}^{kK,kK}\right) + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L=1}^{M}\sum_{l\in L}^{n_L}\boldsymbol{\Delta Z}_{iI,jJ}\boldsymbol{\Delta Z}_{kK,lL}\left(2J_{iI,jJ}^{kK,lL} - K_{iI,jJ}^{kK,lL}\right). \end{aligned} \tag{21}$$

The energy expression in Eq. (21) can be written as a sum of the conventional orthonormal MO energy expression and the so-called exchange-repulsion correction term

$$E = E_0 + E_{ex-rep}, \tag{22}$$

where $E_{ex\text{-}rep}$ collects all terms that are not included in the part of the energy expression that is identical to the conventional orthonormal MO energy expression, namely, those terms that contain elements of the $\boldsymbol{\Delta Z}$ matrix, as shown in the following equation:

$$\begin{aligned} E_{ex-rep} = {} & -2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\boldsymbol{\Delta Z}_{iI,jJ}h_{iI,jJ} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L=1}^{M}\sum_{l\in L}^{n_L}\boldsymbol{\Delta Z}_{kK,lL} \\ & \times \left(2J_{iI,iI}^{kK,lL} - K_{iI,iI}^{kK,lL}\right) - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\boldsymbol{\Delta Z}_{iI,jJ} \\ & \times \left(2J_{iI,jJ}^{kK,kK} - K_{iI,jJ}^{kK,kK}\right) + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L=1}^{M}\sum_{l\in L}^{n_L} \\ & \times \boldsymbol{\Delta Z}_{iI,jJ}\boldsymbol{\Delta Z}_{kK,lL}\left(2J_{iI,jJ}^{kK,lL} - K_{iI,jJ}^{kK,iL}\right). \end{aligned} \tag{23}$$

In order to develop a systematic approach to the formulation of a hierarchy of approximate NOMO models, the $\boldsymbol{\Delta Z}$ matrix is first rewritten in a factorized form

$$\boldsymbol{\Delta Z} = \boldsymbol{S}'\boldsymbol{T}, \tag{24}$$

where $\boldsymbol{S}'$ is the previously defined auxiliary MO overlap matrix, and therefore, the $\boldsymbol{T}$ matrix is given by

$$\boldsymbol{T} = \boldsymbol{I} - \boldsymbol{S}' + \left(\boldsymbol{S}'\right)^2 - \left(\boldsymbol{S}'\right)^3 + \cdots = \boldsymbol{Z}. \tag{25}$$

Although the $\boldsymbol{T}$ matrix is equal to the expansion of the original inverse MO overlap matrix, $\boldsymbol{Z}$, the $\boldsymbol{T}$ matrix notation will be used in place of the $\boldsymbol{Z}$ matrix in the following sections (Secs. III A 2, III B 2, and III B 3, as well as Appendix sections 2, 3, 4, and 5) in order to avoid confusion when approximations are introduced.

Using the expression for $\boldsymbol{\Delta Z}$ in Eq. (24), the factorized exchange-repulsion energy is then given by

Published under an exclusive license by AIP Publishing

$$E_{ex-rep} = -2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\left(\boldsymbol{S}'\boldsymbol{T}\right)_{iI,jJ}h_{iI,jJ} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L=1}^{M}\sum_{l\in L}^{n_L}\left(\boldsymbol{S}'\boldsymbol{T}\right)_{kK,lL}\left(2J_{iI,iI}^{kK,lL} - K_{iI,iI}^{kK,lL}\right)$$
$$- \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_j}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left(\boldsymbol{S}'\boldsymbol{T}\right)_{iI,jJ}\left(2J_{iI,jJ}^{kK,kK} - K_{iI,jJ}^{kK,kK}\right)$$
$$+ \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L=1}^{M}\sum_{l\in L}^{n_L}\left(\boldsymbol{S}'\boldsymbol{T}\right)_{iI,jJ}\left(\boldsymbol{S}'\boldsymbol{T}\right)_{kK,lL}\left(2J_{iI,jJ}^{kK,lL} - K_{iI,jJ}^{kK,lL}\right). \tag{26}$$

Expanding the matrix products,

$$E_{ex-rep} = -2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_p}\boldsymbol{S}'_{iI,pP}\boldsymbol{T}_{pP,jJ}h_{iI,jJ} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L=1}^{M}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R}\boldsymbol{S}'_{kK,rR}\boldsymbol{T}_{rR,lL}\left(2J_{iI,iI}^{kK,lL} - K_{iI,iI}^{kK,lL}\right)$$
$$- \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{P=1}^{M}\sum_{p\in P}^{n_p}\boldsymbol{S}'_{iI,pP}\boldsymbol{T}_{pP,jJ}\left(2J_{iI,jJ}^{kK,kK} - K_{iI,jJ}^{kK,kK}\right) + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L=1}^{M}\sum_{l\in L}^{n_L}\sum_{P=1}^{M}\sum_{p\in P}^{n_p}\boldsymbol{S}'_{iI,pP}\boldsymbol{T}_{pP,jJ}$$
$$\times \sum_{R=1}^{M}\sum_{r\in R}^{n_R}\boldsymbol{S}'_{kK,rR}\boldsymbol{T}_{rR,lL}\left(2J_{iI,jJ}^{kK,lL} - K_{iI,jJ}^{kK,iL}\right). \tag{27}$$

For the sake of clarity in comparing later results with the initial equations in the derivation and because fragment indices are still, at this point, arbitrary, in the following, the J and P indices are swapped, as are the L and R indices:

$$E_{ex-rep} = -2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_p}\boldsymbol{S}'_{iI,jJ}\boldsymbol{T}_{jJ,pP}h_{iI,pP} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L=1}^{M}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R}\boldsymbol{S}'_{kK,lL}\boldsymbol{T}_{lL,rR}\left(2J_{iI,iI}^{kK,rR} - K_{iI,iI}^{kK,rR}\right)$$
$$- \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P}\boldsymbol{S}'_{iI,jJ}\boldsymbol{T}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left(2J_{iI,pP}^{kK,kK} - K_{iI,pP}^{kK,kK}\right) + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J=1}^{M}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P}\boldsymbol{S}'_{iI,jJ}\boldsymbol{T}_{jJ,pP}$$
$$\times \sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L=1}^{M}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R}\boldsymbol{S}'_{kK,lL}\boldsymbol{T}_{lL,rR}\left(2J_{iI,lL}^{kK,rR} - K_{iI,lL}^{kK,rR}\right). \tag{28}$$

Using the properties of the $\boldsymbol{S}'$ matrix given in Eq. (17), Eq. (28) may now be simplified by limiting the summations over I, J and K, L fragment index pairs to unique pairs only [note that $\boldsymbol{S}_{iI,jJ} = s_{iI,jJ}$, as shown in Eq. (5)],

$$E_{ex-rep} = -2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J}s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_p}\boldsymbol{T}_{jJ,pP}h_{iI,pP} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{l\in L}^{n_L}s_{kK,lL}\sum_{R=1}^{M}\sum_{r\in R}^{n_R}\boldsymbol{T}_{lL,rR}\left(2J_{iI,iI}^{kK,rR} - K_{iI,iI}^{kK,rR}\right)$$
$$- \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J}s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P}\boldsymbol{T}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left(2J_{iI,pP}^{kK,kK} - K_{iI,pP}^{kK,kK}\right) + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J}s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P}\boldsymbol{T}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}s_{kK,lL}$$
$$\times \sum_{R=1}^{M}\sum_{r\in R}^{n_R}\boldsymbol{T}_{lL,rR}\left(2J_{iI,pP}^{kK,rR} - K_{iI,pP}^{kK,rR}\right). \tag{29}$$

This expression is the final equation for the *exact* exchange-repulsion energy in the expanded NOMO model.

2. Truncation of the inverse molecular orbital overlap matrix expansion

If the expansion of the $\boldsymbol{T}$ matrix is truncated at some order, n, then the exchange-repulsion energy is approximated by

Published under an exclusive license by AIP Publishing

$$
\begin{aligned}
E_{ex-rep}^{(n)} &\approx E_{\text{ex-rep}} \\
&= E^{(n)} - E_0 \\
&= -2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} \boldsymbol{T}_{jJ,pP}^{(n)} h_{iI,pP} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} \boldsymbol{T}_{lL,rR}^{(n)}\left(2J_{iI,iI}^{kK,rR} - K_{iI,iI}^{kK,rR}\right) - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J} s_{iI,jJ} \\
&\quad \times \sum_{P=1}^{M}\sum_{p\in P}^{n_P} \boldsymbol{T}_{jJ,pP}^{(n)}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left(2J_{iI,pP}^{kK,kK} - K_{iI,pP}^{kK,kK}\right) + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} \boldsymbol{T}_{jJ,pP}^{(n)}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} \boldsymbol{T}_{lL,rR}^{(n)}\left(2J_{iI,pP}^{kK,rR} - K_{iI,pP}^{kK,rR}\right).
\end{aligned} \tag{30}
$$

The matrix $\boldsymbol{T}^{(n)}$ is defined as

$$
\boldsymbol{T}^{(n)} = \sum_{m=0}^{n-1} (-1)^m \left(\boldsymbol{S}'\right)^m,\ n \geq 1, \tag{31}
$$

from which it immediately follows that the complete $\boldsymbol{T}$ matrix, as shown in Eq. (25), can be expressed concisely in the form

$$
\boldsymbol{T} = \sum_{m=0}^{\infty} (-1)^m \left(\boldsymbol{S}'\right)^m. \tag{32}
$$

Using the expression given in Eq. (31), $E_{ex-rep}^{(n)}$ may be rewritten as

$$
\begin{aligned}
E_{ex-rep}^{(n)} = \sum_{m=0}^{n-1} (-1)^m \Bigg\{ & -2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} \boldsymbol{S}'^{m}_{jJ,pP} h_{iI,pP} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\neq K}^{M-1}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} \boldsymbol{S}'^{m}_{lL,rR}\left(2J_{iI,iI}^{kK,rR} - K_{iI,iI}^{kK,rR}\right) \\
& - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} \boldsymbol{S}'^{m}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left(2J_{iI,pP}^{kK,kK} - K_{iI,pP}^{kK,kK}\right) + \sum_{m'=0}^{n-1} (-1) m' \Bigg[\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} \boldsymbol{S}'^{m}_{jJ,pP} \\
& \times \sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\neq K}^{M-1}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} \boldsymbol{S}'^{m'}_{lL,rR}\left(2J_{iI,pP}^{kK,rR} - K_{iI,pP}^{kK,rR}\right) \Bigg] \Bigg\}.
\end{aligned} \tag{33}
$$

From this expression for $E_{ex-rep}^{(n)}$, it is possible to derive a general expression for the unique terms in the exchange-repulsion energy, which are contributed by a given order in the expansion, denoted as $\delta E^{(n)}$,

$$
\delta E_{ex-rep}^{(n)} = E_{ex-rep}^{(n)} - E_{ex-rep}^{(n-1)}. \tag{34}
$$

The unique terms in the exchange-repulsion energy that are added at a given order, n, are thus given by

$$
\begin{aligned}
\delta E_{ex-rep}^{(n)} = (-1)^{n-1} \Bigg\{ & -2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} \boldsymbol{S}'^{n-1}_{jJ,pP} h_{iI,pP} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\neq K}^{M-1}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} \boldsymbol{S}'^{n-1}_{lL,rR} \\
& \times \left[\left(2J_{iI,iI}^{kK,rR} - K_{iI,iI}^{kK,rR}\right) - \sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} \boldsymbol{T}_{jJ,pP}^{(n-1)}\left(2J_{iI,pP}^{kK,rR} - K_{iI,pP}^{kK,rR}\right) \right] \\
& - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} \boldsymbol{S}'^{n-1}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left[\left(2J_{iI,pP}^{kK,kK} - K_{iI,pP}^{kK,kK}\right) - \sum_{L\neq K}^{M-1}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} \boldsymbol{T}_{lL,rR}^{(n-1)}\left(2J_{iI,pP}^{kK,rR} - K_{iI,pP}^{kK,rR}\right) \right] \Bigg\} \\
& + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} \boldsymbol{S}'^{n-1}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\neq K}^{M-1}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} \boldsymbol{S}'^{n-1}_{lI,rR}\left(2J_{iI,pP}^{kK,rR} - K_{iI,pP}^{kK,rR}\right).
\end{aligned} \tag{35}
$$

Published under an exclusive license by AIP Publishing

B. Approximate expanded non-orthogonal molecular orbital theory

Although the direct computation of the $\boldsymbol{Z}$ matrix is not itself computationally complex, requiring only the calculation of the MO overlap matrix and its inverse, the inclusion of the full $\boldsymbol{Z}$ matrix, and by extension higher orders of the $\boldsymbol{T}$ matrix, introduces terms in the exchange-repulsion energy that involve all unique fragment quartets and all resulting unique MO quartet ERIs. Computing the full set of these integrals has the exact same computational complexity as performing the corresponding conventional HF or NOMO theory calculation. Thus, in order to further develop the model, an appropriate approximation is needed to eliminate some elements of the $\boldsymbol{T}$ matrix at each order of the expansion.

1. Partial neglect of interfragment overlap

In order to reduce the number of $\boldsymbol{T}$ matrix elements that must be included at a given order, an approximation, referred to as the *partial neglect of interfragment overlap* (PNIO) approximation, is introduced herein and subsequently employed in the derivation of a simplified model. The PNIO *concept* has been utilized in various formulations in many fragment-based models and methods for the purpose of improving computational tractability.[22,23]

The basis for the PNIO approximation is as follows: For a sufficiently large system of many fragments, each with relatively well-localized MOs, there will, in general, be fragment pairs for which the overlap between each and every unique pair of MOs (i, k) in those two fragments (I, K) is vanishingly small. In other words, there will exist fragment pairs for which the following relationship holds:

$$\phi_i^I \phi_k^K \approx 0 \; a.e. \Rightarrow s_{iI,kK} \approx 0 \; \forall \; i \in I, \; k \in K. \tag{36}$$

Such fragment pairs will be referred to as *non-overlapping*. Conversely, fragment pairs for which Eq. (36) does not hold will be referred to as *overlapping*. For a given fragment index, I, the subset of fragment indices for fragments that are overlapping with I is denoted as I_{ovl}. That is, I_{ovl} is a set of fragment indices, $\{J \neq I\}$, for which one or more MO overlap integrals between MOs in each fragment is non-vanishing,

$$s_{iI,jJ} \neq 0 \Rightarrow J \in \boldsymbol{I}_{ovl}. \tag{37}$$

The size of this set will be denoted $M_{\boldsymbol{I}_{ovl}}$. Using this approximation, the structure of the $\boldsymbol{S}'$ matrix takes the form

$$\boldsymbol{S}'_{aA,bB} = \begin{cases} 0, & B \notin A_{ovl}, \\ s_{aA,bB}, & B \in A_{ovl} \leftrightarrow A \in B_{ovl}, \\ 0, & B = A. \end{cases} \tag{38}$$

Clearly, the PNIO approximation may also be considered to be an assumption of a sparse MO overlap matrix. In practical applications, the set of non-overlapping fragment pairs may be chosen as those for which all MO overlap integrals fall below some numerical cutoff threshold, s_{cutoff},

$$s_{iI,kK} < s_{cutoff} \; for \; all \; i \in I, \; k \in K. \tag{39}$$

For a simpler implementation, the set of non-overlapping fragments may also be determined directly from the set of AO overlap integrals between fragments, since it will be the case that if all AO integrals fall below some chosen AO threshold, then all MO integrals must also fall below a corresponding MO threshold.

The PNIO approximation is related to those approximations that form the basis of fragment (or locality) based correlation methods, such as the Domain Local Pair-Natural Orbitals (DLPNOs), Cluster-in-Molecule (CIM), and Divide–Expand–Consolidate (DEC) methods. Along with a great deal of other related methods, these local correlation methods have been shown to be capable of calculating the total correlation energy for large molecular systems at a quantitative level of accuracy while maintaining linear scaling with respect to the size of the system.[24–38]

2. Approximate expanded non-orthogonal molecular orbital energy

It is now possible to formulate an approximate form of the exchange-repulsion energy, given in Eq. (29), which accounts for the strongest interactions between overlapping fragments and neglects those vanishing interactions between non-overlapping fragments. By introducing the PNIO approximation, the computational complexity associated with calculating the exchange-repulsion energy at some order of truncation is significantly reduced, and the scaling of calculations, with respect to the total system size, is made to be linear at any arbitrary order of truncation.

First, the PNIO approximation is applied to the I, J and K, L fragment index pairs,

$$\begin{aligned} E_{ex-rep} \approx & -2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} \boldsymbol{T}_{jJ,pP} h_{iI,pP} \\ & - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} \boldsymbol{T}_{lL,rR}\left(2J_{iI,iI}^{kK,rR} - K_{iI,iI}^{kK,rR}\right) \\ & - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} \boldsymbol{T}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left(2J_{iI,pP}^{kK,kK} - K_{iI,pP}^{kK,kK}\right) \\ & + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} \boldsymbol{T}_{jJ,pP} \\ & \times \sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} \boldsymbol{T}_{lL,rR}\left(2J_{iI,pP}^{kK,rR} - K_{iI,pP}^{kK,rR}\right). \end{aligned} \tag{40}$$

This general expression for the approximated exchange-repulsion energy, prior to choosing a truncation order for the $\boldsymbol{T}$ matrix, utilizes only the restriction of summations over fragment pairs to those that are considered overlapping under the criteria given in Eq. (36). For the truncated exchange-repulsion energy, $E_{ex-rep}^{(n)}$, the PNIO approximation is applied to each new fragment index pair that appears at each increasing order n. By proceeding

Published under an exclusive license by AIP Publishing

in this manner, the increase in the computational complexity with each increasing order is linear with respect to the total size of the system.

The expressions for the approximate exchange-repulsion energy, which are obtained for the first few orders of truncation, n, are given below. Extension of the model to higher orders is straightforward and follows the same pattern used to obtain the following expressions;

For the $n = 1$ (first-order) case, the equation is easily simplified by replacing all instances of $\boldsymbol{T}$ with the identity matrix and eliminating the additional summation indices. Thus, $E^{(1)}_{ex-rep}$ is given by

$$E^{(1)}_{ex-rep} = -2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}h_{iI,jJ} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\left(2J^{kK,lL}_{iI,iI} - K^{kK,lL}_{iI,iI}\right)$$
$$- \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left(2J^{kK,kK}_{iI,jJ} - K^{kK,kK}_{iI,jJ}\right) + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\left(2J^{kK,lL}_{iI,jJ} - K^{kK,lL}_{iI,jJ}\right). \tag{41}$$

As will be shown in Sec. IV C 2, this expression for the first-order exchange-repulsion energy closely resembles that of previously described methods. Although relatively inexpensive to calculate, this expression alone accounts for a substantial percentage of the total exchange-repulsion energy.

For the $n = 2$ (second-order) case, additional applications of the PNIO approximation are needed to maintain tractability by avoiding the addition of new summations over the full range of M fragment indices. In this instance, the approximation is applied to the new J, P and L, R fragment index pairs. The unique terms for the exchange-repulsion energy at this order [i.e., $\delta E^{(2)}_{ex-rep} = E^{(2)}_{ex-rep} - E^{(1)}_{ex-rep}$] are given by

$$\begin{aligned}\delta E^{(2)}_{ex-rep} &= E^{(2)}_{ex-rep} - E^{(1)}_{ex-rep}\\ &= 2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P\in J_{ovl}}^{M_{J_{ovl}}}\sum_{p\in P}^{n_P} s_{jJ,pP}h_{iI,pP} + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R\in L_{ovl}}^{M_{L_{ovl}}}\sum_{r\in R}^{n_R} s_{iL,rR}\left[\left(2J^{kK,rR}_{iI,iI} - K^{kK,rR}_{iI,iI}\right) - \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\left(2J^{kK,rR}_{iI,jJ} - K^{kK,rR}_{iI,jJ}\right)\right]\\ &+ \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P\in J_{ovl}}^{M_{J_{ovl}}}\sum_{p\in P}^{n_P} s_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left[\left(2J^{kK}_{iI,pP} - K^{kK,kK}_{iI,pP}\right) - \sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\left(2J^{kK,lL}_{iI,pP} - K^{kK,lL}_{iI,pP}\right)\right] + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P\in J_{ovl}}^{M_{J_{ovl}}}\sum_{p\in P}^{n_P} s_{jJ,pP}\\ &\times \sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R\in L_{ovl}}^{M_{L_{ovl}}}\sum_{r\in R}^{n_R} s_{lL,rR}\left(2J^{kK,rR}_{iI,pP} - K^{kK,rR}_{iI,pP}\right).\end{aligned} \tag{42}$$

As can be seen in Eq. (42), all terms in the second-order exchange-repulsion correction, aside from the last (fourth), have the signs reversed relative to the first-order correction. Therefore, these first three terms will generally have the effect of reducing the magnitude of the corresponding terms in $E^{(1)}_{ex-rep}$.

Next, consider the expression for the unique terms (i.e., $\delta E^{(3)}_{ex-rep} = E^{(3)}_{ex-rep} - E^{(2)}_{ex-rep} - E^{(1)}_{ex-rep}$) in the exchange-repulsion energy for the $n = 3$ (third-order) case,

$$\begin{aligned}\delta E^{(3)}_{ex-rep} &= E^{(3)}_{ex-rep} - E^{(2)}_{ex-rep} - E^{(1)}_{ex-rep}\\ &= -2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P\in J_{ovl}}^{M_{J_{ovl}}}\sum_{p\in P}^{n_P} s_{jJ,pP}\sum_{Q\in P_{ovl}}^{M_{P_{ovl}}}\sum_{q\in Q}^{n_Q} s_{pP,qQ}h_{iI,qQ} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R\in L_{ovl}}^{M_{L_{ovl}}}\sum_{r\in R}^{n_R} s_{lL,rR}\sum_{\boldsymbol{S}\in R_{ovl}}^{M_{R_{ovl}}}\sum_{s\in \boldsymbol{S}}^{n_{\boldsymbol{S}}} s_{rR,s\boldsymbol{S}}\\ &\times \left\{\left(2J^{kK,s\boldsymbol{S}}_{iI,iI} - K^{kK,s\boldsymbol{S}}_{iI,iI}\right) - \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\left[\left(2J^{kK,s\boldsymbol{S}}_{iI,jJ} - K^{kK,s\boldsymbol{S}}_{iI,jJ}\right) - \sum_{P\in J_{ovl}}^{M_{J_{ovl}}}\sum_{p\in P}^{n_P} s_{jJ,pP}\left(2J^{kK,s\boldsymbol{S}}_{iI,pP} - K^{kK,s\boldsymbol{S}}_{iI,pP}\right)\right]\right\}\\ &- \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P\in J_{ovl}}^{M_{J_{ovl}}}\sum_{p\in P}^{n_P} s_{jJ,pP}\sum_{Q\in P_{ovl}}^{M_{P_{ovl}}}\sum_{q\in Q}^{n_Q} s_{pp,qQ}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left\{\left(2J^{kK,kK}_{iI,qQ} - K^{kK,kK}_{iI,qQ}\right) - \sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\right.\\ &\times \left.\left[\left(2J^{kK,kK}_{iI,qQ} - K^{kK,kK}_{iI,qQ}\right) - \sum_{R\in L_{ovl}}^{M_{L_{ovl}}}\sum_{r\in R}^{n_R} s_{lL,rR}\left(2J^{kK,kK}_{iI,qQ} - K^{kK,kK}_{iI,qQ}\right)\right]\right\} + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P\in J_{ovl}}^{M_{J_{ovl}}}\sum_{p\in P}^{n_P} s_{jJ,pP}\sum_{Q\in P_{ovl}}^{M_{P_{ovl}}}\sum_{q\in Q}^{n_Q} s_{pP,qQ}\\ &\times \sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R\in L_{ovl}}^{M_{L_{ovl}}}\sum_{r\in R}^{n_R} s_{lL,rR}\sum_{\boldsymbol{S}\in R_{ovl}}^{M_{R_{ovl}}}\sum_{s\in \boldsymbol{S}}^{n_{\boldsymbol{S}}} s_{rR,s\boldsymbol{S}}\left(2J^{kK,s\boldsymbol{S}}_{iI,qQ} - K^{kK,s\boldsymbol{S}}_{iI,qQ}\right).\end{aligned} \tag{43}$$

Published under an exclusive license by AIP Publishing

Higher-order corrections follow the same pattern of including one factor of $\boldsymbol{S}'$ for each instance of the I, J and K, L fragment index pairs and applying the PNIO approximation to new fragment indices that appear as a consequence. The derivation of the last three expressions is given in detail in the Appendix.

3. General Fock operator for the expanded non-orthogonal molecular orbital method

Analogous to the conventional NOMO theory, in order to define the Fock operator that determines the optimal orbitals for each fragment, it is necessary to first define two new operators. The first is a mean-field embedding Fock operator, $\widehat{\boldsymbol{F}}^{(n)}$, which is equivalent for *all* fragments,

$$\widehat{\boldsymbol{F}}^{(n)} = \hat{h} + \sum_{K=1}^{M}\sum_{k\in K}^{n_K}\Bigg[\left(2\hat{J}_{kK,kK} - \widehat{K}_{kK,kK}\right) - \sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kL,lL}\sum_{R(n)}\sum_{r\in R}^{n_R}\boldsymbol{T}^{(n)}_{lL,rR}\left(2\hat{J}_{kK,rR} - \widehat{K}_{kK,rR}\right)\Bigg], \tag{44}$$

where $\hat{J}_{kK,kK}$ and $\widehat{K}_{kK,kK}$ are the conventional Coulomb and exchange operators and $\hat{J}_{kK,lL}$ and $\widehat{K}_{kK,lL}$ are the two-fragment analogs of these operators,

$$\hat{J}_{kK,lL} = \int \phi_k^{K*}(r_2) r_{12}^{-1} \phi_l^L(r_2) dr_2, \tag{45}$$

$$\widehat{K}_{kK,lL} = \int \phi_k^{K*}(r_2) r_{12}^{-1} \phi_l^L(r_2) dr_2 \hat{P}_{12}(r_1, r_2), \tag{46}$$

and $\hat{P}_{12}$ is the permutation operator acting on electrons 1 and 2. The summation limits on the fragment index R in Eq. (44) are not specified, as they are dependent on the choice of the truncation order, n. The nature of this dependency will become clear when the Fock operators for the first few orders of truncation are given in Sec. IV A.

In addition to the embedding Fock operator $\widehat{\boldsymbol{F}}(n)$ defined in Eq. (44), it is necessary to define for each fragment (I) a fragment-specific operator that projects onto some part of the occupied subspace of the full system, which is overlapping with the occupied subspace spanned by the MOs in the Ith fragment, dependent on the order n. Specifically, this operator, $\hat{P}^{(n)}_{\boldsymbol{I}_{ovl}}$ spans the (potentially non-orthogonal) subspace that contains fragments that overlap with a given fragment I and is given by

$$\begin{aligned}\hat{p}^{(n)}_{\boldsymbol{I}_{ovl}} &= \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\left(\sum_{j\in J}^{n_J}\hat{p}_{jJ}\right)\left(1 - \hat{p}^{(n-1)}_{J_{ovl}}\right);\ \hat{p}^{(0)}_{J_{ovl}} = 0 \\ &= \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\hat{p}_J\left(1 - \hat{p}^{(n-1)}_{J_{ovl}}\right) \\ &= \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P(n)}\sum_{p\in P}^{n_P}\boldsymbol{T}^{(n)}_{jJ,pP}\left|\phi_j^J\right\rangle\left\langle\phi_p^P\right|,\end{aligned} \tag{47}$$

where

$$\hat{p}_{jJ} = \left|\phi_j^J\right\rangle\left\langle\phi_j^J\right| \tag{48}$$

and

$$\hat{p}_J = \sum_{j\in J}^{n_J}\hat{p}_{jJ}. \tag{49}$$

Using the two operators $\widehat{\boldsymbol{F}}^{(n)}$ and $\hat{p}^{(n)}_{\boldsymbol{I}_{ovl}}$, it is now possible to define the fragment-specific Fock operator that determines the optimal set of MOs for a given fragment. Denoting this Fock operator for the Ith fragment as $\widehat{\boldsymbol{F}}^{(n)}_I$, the operator is defined as

$$\begin{aligned}\widehat{\boldsymbol{F}}^{(n)}_I = \widehat{\boldsymbol{F}}^{(n)} &- \left[\widehat{\boldsymbol{F}}^{(n)}\left(\hat{p}^{(n)}_{\boldsymbol{I}_{ovl}}\right)^{\dagger} + \hat{p}^{(n)}_{\boldsymbol{I}_{ovl}}\widehat{\boldsymbol{F}}^{(n)}\right] \\ &- \sum_{m=1}^{n}\sum_{a=1}^{m-1}\left[\hat{p}^{(a)}_{\boldsymbol{I}_{ovl}}\widehat{\boldsymbol{F}}^{(n)}\left(\hat{p}^{(m-a)}_{\boldsymbol{I}_{ovl}}\right)^{\dagger}\right].\end{aligned} \tag{50}$$

The corresponding eigenvalue equation that determines the MOs for the Ith fragment, after the usual unitary transformation of the pseudo-eigenvalue equation, is given by

$$\widehat{\boldsymbol{F}}^{(n)}_I \phi^I_i = \epsilon^I_i \phi^I_i. \tag{51}$$

The set of coupled Hartree–Fock (HF) equations for the fragments may be solved iteratively using the two-level self-consistent field (SCF) procedure, whereby individual HF equations are solved in a static embedding potential of the other $M-1$ fragments, and then a global SCF iteration is performed to update the fragment-density-dependent embedding potential. The SCF is converged when all fragment wave functions have converged and the embedding potential has therefore also converged. Although other iterative schemes are possible, this procedure has proven to be efficient and well-behaved in a number of previously developed methods.[39]

IV. DISCUSSION

A. Unique fragmentation methods

The ENMO/n model offers a general formulation for the development of various fragmentation methods, depending on the level of truncation chosen in the expansion of the $\boldsymbol{T}$ matrix. Below, the first few possible methods are outlined, and their distinct characteristics are highlighted.

1. ENMO/1: The first-order expanded non-orthogonal molecular orbital method

If one truncates the $\boldsymbol{T}$ matrix expansion at the first order (i.e., $\boldsymbol{T} \approx \boldsymbol{T}^{(1)} = \boldsymbol{I}$), the ENMO/1 or First-order Non-Orthogonal Molecular Orbital (FNMO) model is obtained. The method derived from this model represents the complete variational optimization of all coupled fragment HF equations defined by the pair of operators $\widehat{\boldsymbol{F}}^{(1)}$ and $\hat{p}^{(1)}_{\boldsymbol{I}_{ovl}}$,

$$\widehat{\boldsymbol{F}}^{(1)} = \hat{h} + \sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left[\left(2\hat{J}_{kK,kK} - \widehat{K}_{kK,kK}\right) - \sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\left(2\hat{J}_{kK,lL} - \widehat{K}_{kK,lL}\right)\right] \tag{52}$$

and

Published under an exclusive license by AIP Publishing

$$\hat{P}^{(1)}_{\boldsymbol{I}_{ovl}} = \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{\boldsymbol{I}_{ovl}}} \sum_{j\in J}^{n_J} \hat{p}_{jj} \tag{53}$$

to obtain MOs for each fragment and thus for the full system. In some respects, ENMO/1 may be considered a partially symmetry-adapted embedding analog of the electrostatically embedded fragmentation methods, such as the fragment molecular orbital (FMO)[40] and explicit polarization (X-Pol)[41] methods. In particular, the ENMO/1 Fock operator involves a one-particle mean-field embedding potential composed of all other fragments in the total system.

The computational procedure of the ENMO/1 (FNMO) method is similar to those of other embedded methods, with variations arising due to the symmetry-adapted nature of the embedding potential. Figure 1 presents an outline of the general algorithm for computing the ENMO/1 fragment wave functions and energies. The elements $C_{\mu\nu}$ are AO coefficients, and the interfragment embedding operator, $\widehat{V}_1^{embed}$, is defined as

$$\widehat{V}_1^{embed} = \sum_{K=1}^{M} \widehat{V}_K^{(1)} = \sum_{K=1}^{M} \sum_{k\in K}^{n_K} \left[\left(2\hat{J}_{kK,kK} - \widehat{K}_{kK,kK}\right) - \sum_{L\in K_{ovl}}^{M_{K_{ovl}}} \sum_{l\in L}^{n_L} s_{kK,lL} \left(2\hat{J}_{kK,lL} - \widehat{K}_{kK,lL}\right) \right]. \tag{54}$$

At the local SCF level, shown in Fig. 2, the computation of each fragment wave function is distinctly different from other embedded methods. The distinction is specifically that for each local SCF iteration in the ENMO/1 method, the terms in the Fock matrix (both

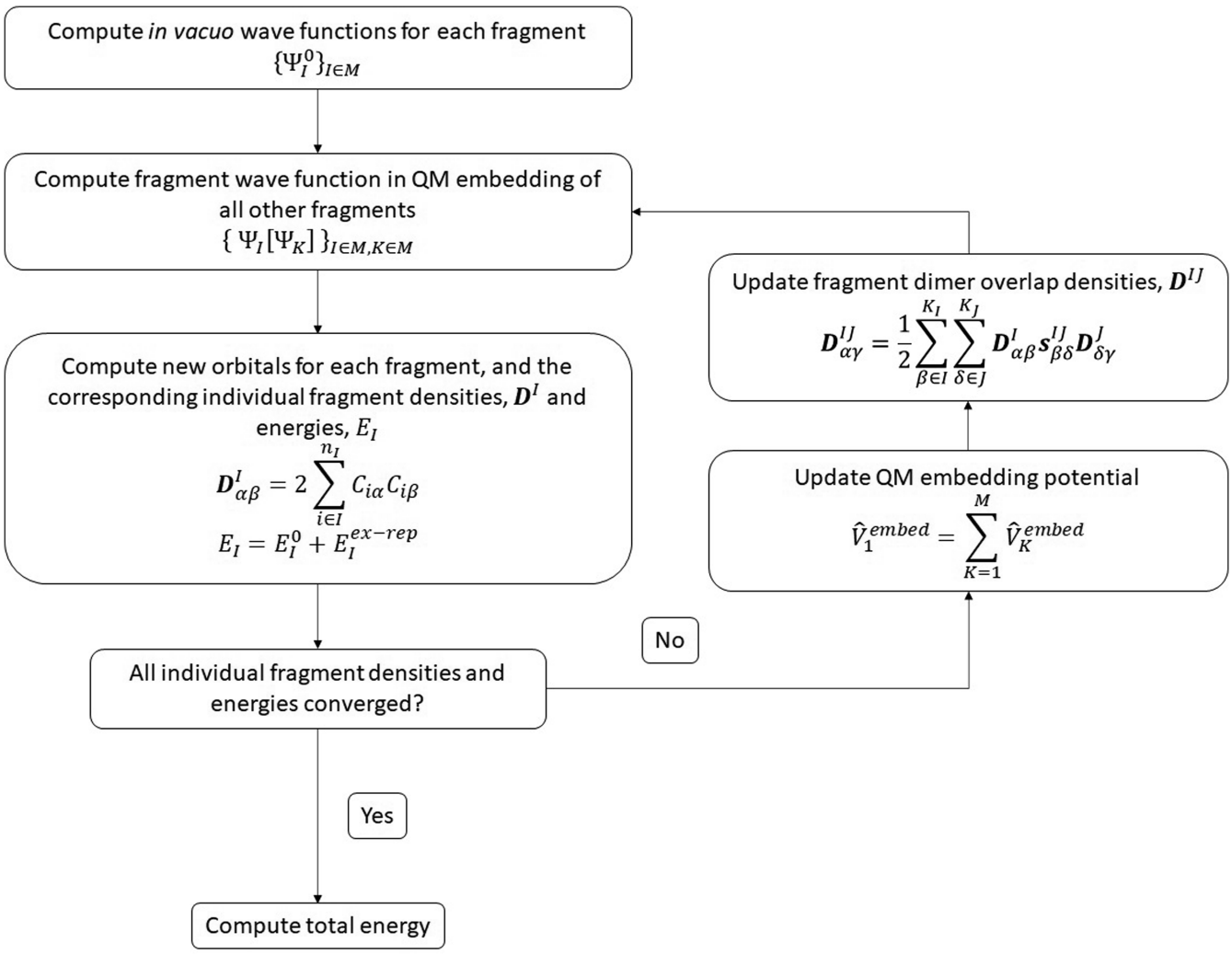


FIG. 1. Global SCF algorithm for computing the ENMO/1 fragment wave functions and energy in a two-level iterative scheme.

Published under an exclusive license by AIP Publishing

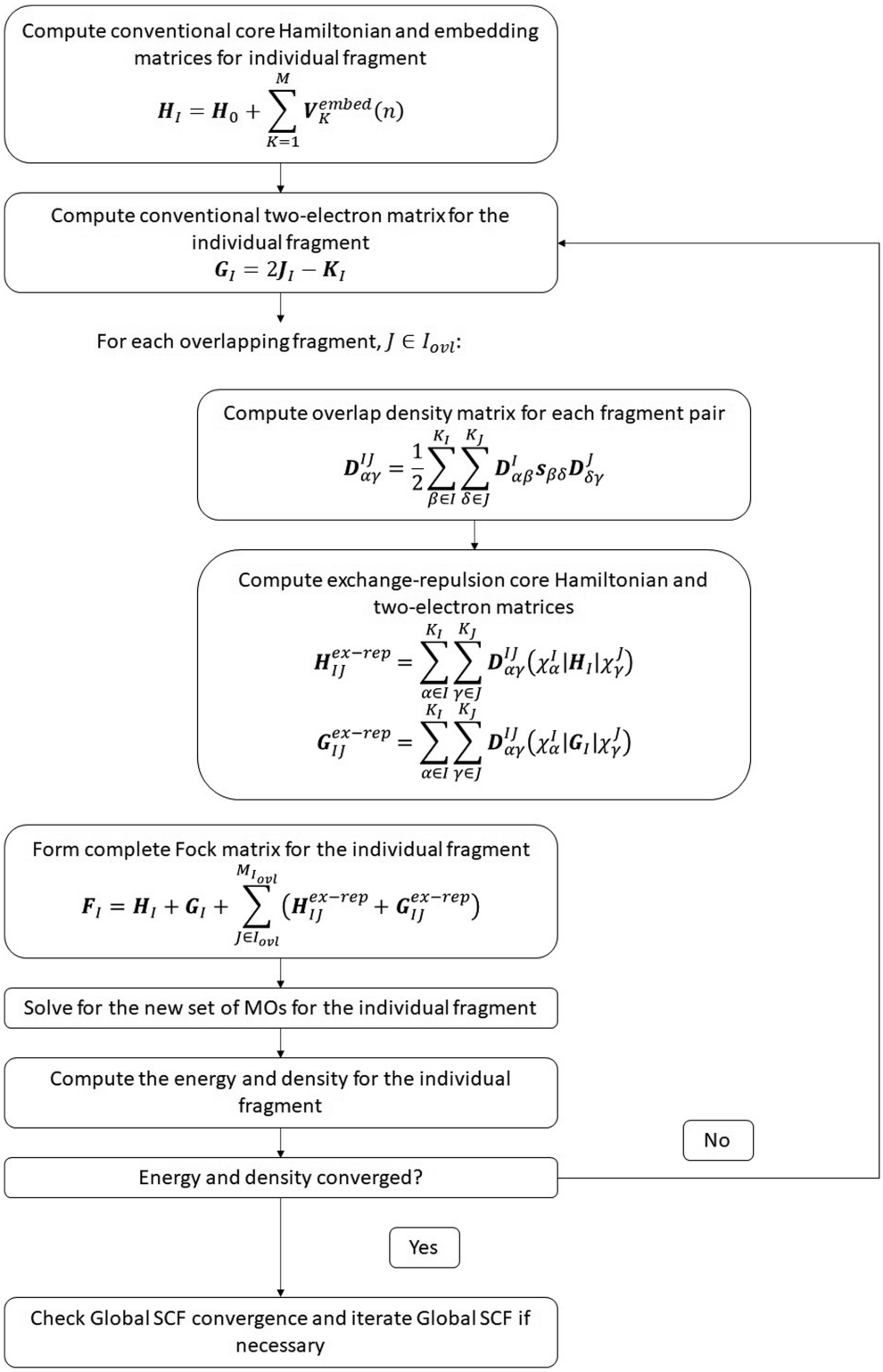


FIG. 2. Local SCF algorithm for computing each of the ENMO/1 (also known as FNMO) fragment wave functions and energy in a two-level iterative scheme.

Published under an exclusive license by AIP Publishing

the core Hamiltonian and the two-electron matrices) that depend on overlap with the MOs assigned to the fragment being studied must be recomputed because these MOs change each iteration. While this recomputation of elements in the two matrices increases the computational complexity of computing each term in the Fock matrix, the overall scaling with respect to the total size of the system is still minimized, as will be discussed in greater detail in Sec. IV D.

$$\widehat{F}^{(2)} = \hat{h} + \sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left\{\left(2\hat{J}_{kK,kK} - \widehat{K}_{kK,kK}\right) - \sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\left[\left(2\hat{J}_{kK,lL} - \widehat{K}_{kK,lL}\right) - \sum_{R\in L_{ovl}}^{M_{L_{ovl}}}\sum_{r\in R}^{n_R} s_{lL,rR}\left(2\hat{J}_{kK,rR} - \widehat{K}_{kK,rR}\right)\right]\right\} \tag{55}$$

and

$$\begin{aligned}\hat{p}^{(2)}_{\boldsymbol{I}_{ovl}} &= \sum_{J\in\boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\hat{p}_{jJ}\left(1 - \sum_{P\in J_{ovl}}^{M_{J_{ovl}}}\sum_{p\in P}^{n_P}\hat{p}_{pP}\right)\\ &= \sum_{J\in\boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\left(|\phi_j^J\rangle\langle\phi_j^J| - \sum_{P\in J_{ovl}}^{M_{K_{ovl}}}\sum_{p\in P}^{n_P} s_{jJ,pP}|\phi_j^J\rangle\langle\phi_p^P|\right)\\ &= \sum_{J\in\boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}|\phi_j^J\rangle\left(\langle\phi_j^J| - \sum_{P\in J_{ovl}}^{M_{K_{ovl}}}\sum_{p\in P}^{n_P} s_{jJ,pP}\langle\phi_p^P|\right)\end{aligned} \tag{56}$$

to obtain the MOs for each fragment and thus the full system.

The computational procedure for the ENMO/2 method is similar to that for the ENMO/1 method, with variations in the local and global SCF iterations due to the presence of products of fragment pair overlap density matrices in the mean-field embedding operator, as well as in the fragment-specific projection operator. The global SCF algorithm is shown schematically in Fig. 3:

In Fig. 3, the only new algorithmic step is the computation of fragment trimer overlap density matrices at each global SCF iteration. In addition, the embedding operator, $\widehat{V}_2^{embed}$, now includes additional Coulomb and exchange operators

2. ENMO/2: The second-order expanded non-orthogonal molecular orbital method

If one instead truncates the $\boldsymbol{T}$ matrix expansion at the second order (i.e., $\boldsymbol{T} \approx \boldsymbol{T}^{(2)} = \boldsymbol{I} - \boldsymbol{S}'$), the ENMO/2 model is obtained. In analogy with ENMO/1, ENMO/2 represents the complete variational optimization of all coupled fragment HF equations defined by the pair of operators

$$\begin{aligned}\widehat{V}_2^{embed} &= \widehat{V}_1^{embed} + \delta\widehat{V}_{(2)}^{embed}\\ &= \sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left\{\left(2\hat{J}_{kK,kK} - \widehat{K}_{kK,kK}\right) - \sum_{L\in K}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\right.\\ &\quad\left.\times\left[\left(2\hat{J}_{kK,kK} - \widehat{K}_{kK,lL}\right) - \sum_{R\in L_{ovl}}^{M_{L_{ovl}}}\sum_{r\in R}^{n_R} s_{lL,rR}\left(2\hat{J}_{kK,lL} - \widehat{K}_{kK,rR}\right)\right]\right\}.\end{aligned} \tag{57}$$

There are also only minor algorithmic differences between the ENMO/1 and ENMO/2 methods at the local SCF level. Specifically, the ENMO/2 local SCF algorithm requires additional matrix multiplications to form overlapping fragment trimer terms. The modified ENMO/2 algorithm for the local SCF is shown in Fig. 4:

3. ENMO/3: The third-order expanded non-orthogonal molecular orbital method

Truncation of the $\boldsymbol{T}$ matrix expansion at the third order [i.e., $\boldsymbol{T} \approx \boldsymbol{T}^{(3)} = \boldsymbol{I} - \boldsymbol{S}' + (\boldsymbol{S}')^2$] produces the ENMO/3 model. Once again, the method derived from this model represents the complete variational optimization of all coupled fragment HF equations defined by the pair of operators

$$\widehat{F}^{(3)} = \hat{h} + \sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left\{\left(2\hat{J}_{kK,kK} - \widehat{K}_{kK,kK}\right) - \sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\left[\left(2\hat{J}_{kK,lL} - \widehat{K}_{kK,lL}\right) - \sum_{R\in L_{ovl}}^{M_{L_{ovl}}}\sum_{r\in R}^{n_R} s_{lL,rR}\left[\left(2\hat{J}_{kK,rR} - \widehat{K}_{kK,rR}\right) - \sum_{\boldsymbol{S}\in R_{ovl}}^{M_{R_{ovl}}}\sum_{s\in\boldsymbol{S}}^{n_S} s_{rR,s\boldsymbol{S}}\left(2\hat{J}_{kK,s\boldsymbol{S}} - \widehat{K}_{kK,s\boldsymbol{S}}\right)\right]\right]\right\} \tag{58}$$

and

$$\begin{aligned}\hat{p}^{(3)}_{\boldsymbol{I}_{ovl}} &= \sum_{J\in\boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\hat{p}_{jJ}\left[1 - \sum_{P\in J_{ovl}}^{M_{J_{ovl}}}\sum_{p\in P}^{n_P}\hat{p}_{pP}\left(1 - \sum_{Q\in P_{ovl}}^{M_{P_{ovl}}}\sum_{q\in Q}^{n_Q}\hat{p}_{qQ}\right)\right] = \sum_{J\in\boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\left[|\phi_j^J\rangle\langle\phi_j^J| - \sum_{P\in J_{ovl}}^{M_{J_{ovl}}}\sum_{p\in P}^{n_P} s_{jJ,pP}\left(|\phi_j^J\rangle\langle\phi_p^P| - \sum_{Q\in P_{ovl}}^{M_{P_{ovl}}}\sum_{q\in Q}^{n_Q} s_{pP,qQ}|\phi_j^J\rangle\langle\phi_q^Q|\right)\right]\\ &= \sum_{J\in\boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}|\phi_j^J\rangle\left[\langle\phi_j^J| - \sum_{P\in J_{ovl}}^{M_{J_{ovl}}}\sum_{p\in P}^{n_P} s_{jJ,pP}\left(\langle\phi_p^P| - \sum_{Q\in P_{ovl}}^{M_{Q_{ovl}}}\sum_{q\in Q}^{n_Q} s_{pP,qQ}\langle\phi_q^Q|\right)\right]\end{aligned} \tag{59}$$

to obtain the MOs for each fragment and thus the full system.

Published under an exclusive license by AIP Publishing

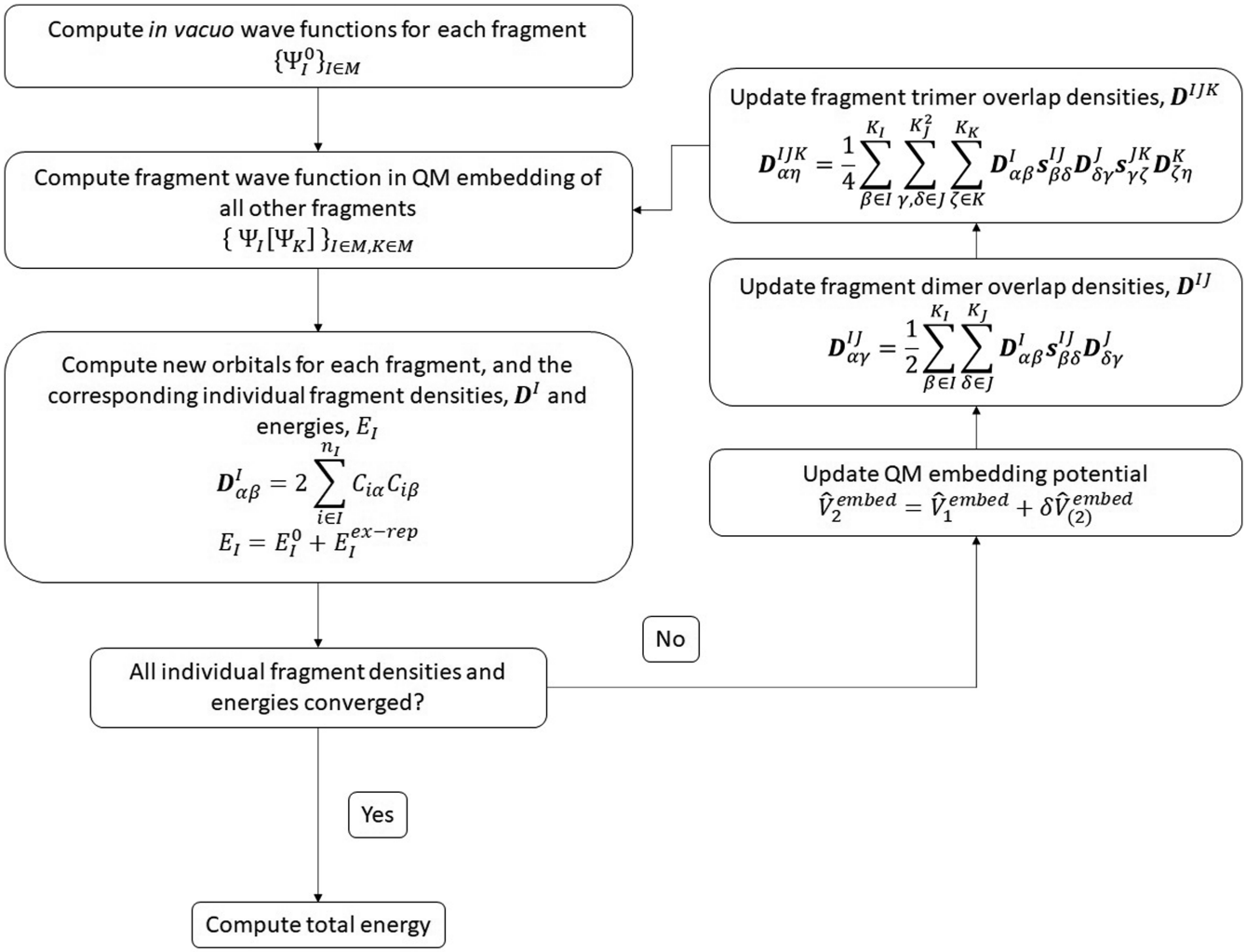


FIG. 3. Global SCF algorithm for computing the ENMO/2 fragment wave functions and energy in a two-level iterative scheme.

The computational procedure of the ENMO/3 method expands on the procedure for the ENMO/2 method by requiring the formation of fragment tetramer overlap density matrices. The general algorithms for the global and local SCF cycles follow straightforwardly from Figs. 1–4.

The only new algorithmic step is the computation of fragment tetramer overlap density matrices at each global SCF iteration. In addition, the embedding operator, $\widehat{V}_3^{embed}$, is expanded to include additional Coulomb and exchange operators,

$$\widehat{V}_3^{embed} = \widehat{V}_2^{embed} + \delta\widehat{V}_{(3)}^{embed} = \sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left\{\left(2\hat{J}_{kK,kK} - \widehat{K}_{kK,kK}\right) - \sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\left[\left(2\hat{J}_{kK,lL} - \widehat{K}_{kK,lL}\right) - \sum_{R\in L_{ovl}}^{M_{L_{ovl}}}\sum_{r\in R}^{n_R} s_{lL,rR}\left\{\left(2\hat{J}_{kK,rR} - \widehat{K}_{kK,rR}\right) - \sum_{S\in R_{ovl}}^{M_{R_{ovl}}}\sum_{s\in S}^{n_S} s_{rR,sS}\left(2\hat{J}_{kK,sS} - \widehat{K}_{kK,sS}\right)\right\}\right]\right\}. \quad (60)$$

Published under an exclusive license by AIP Publishing

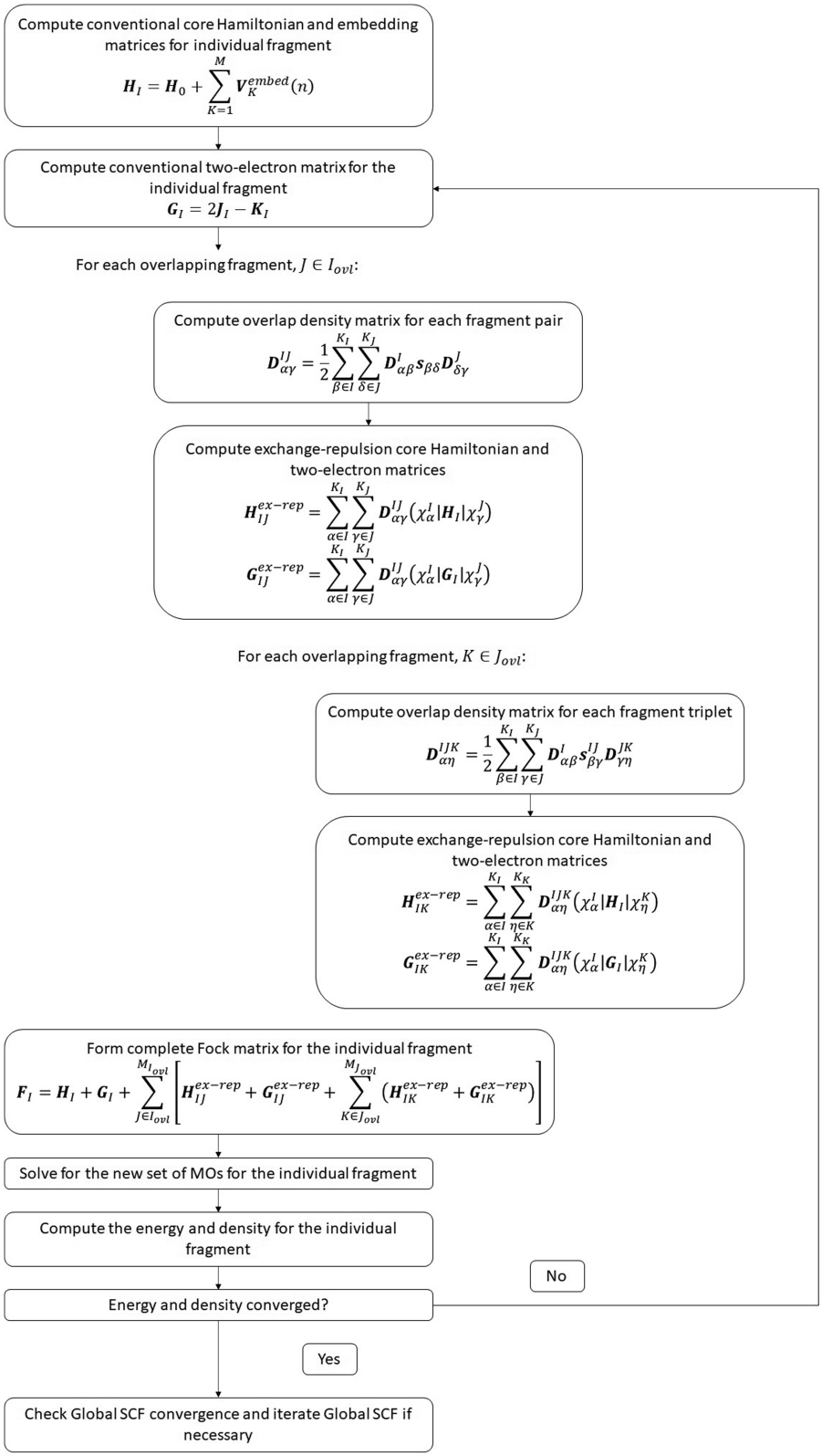


FIG. 4. Local SCF algorithm for computing each of the ENMO/2 fragment wave functions, and their corresponding energies, in a two-level iterative scheme.

Published under an exclusive license by AIP Publishing

At the local SCF level, the algorithm for the ENMO/3 method is extended from the ENMO/2 method. In addition to forming fragment trimer overlapping density matrices, each iteration also requires the formation of fragment tetramer overlapping density matrices. Similarly, the global SCF iterations also require this computation.

The necessary modifications to the global and local SCF algorithms for the extension of the ENMO/n model to the development of methods that include higher-order exchange-repulsion corrections (i.e., $n > 3$) follow the algorithms outlined above in a straightforward manner. At each increasing order, additional matrix multiplications are required to form fragment N-mer overlap density matrices.

B. Fragment basis sets

Two choices for defining basis sets for each fragment appear to be the most appropriate, although others are certainly possible. The first is the simple explicitly localized basis set approach, in which basis function centers are limited to atomic centers assigned to a given fragment. This is the approach taken in the Absolutely Localized Molecular Orbital (ALMO) method, along with other related methods.[18–21] It is valuable for isolating the Coulomb and exchange effects from the so-called charge transfer interactions. The drawback is that the quantitative results omit charge transfer interaction energies, and these may be critical for accurately describing a particular system. It is possible to add approximate charge-transfer terms onto the explicitly localized energy expression.[21,51–53]

A second approach is to assign basis functions to a given fragment based on its own atomic centers as well as those of the overlapping fragments. Since the charge-transfer interaction will, in the zero-overlap limit of the PNIO approximation, be restricted to overlapping fragments only, this approach can accurately capture the charge-transfer effects without the need to add secondary corrections. The use of so-called shared basis functions has been utilized in the Group Molecular Orbital (GMO) method and has shown potential for substantial accuracy.[44] This approach does introduce certain complexities into performing calculations on a given system for multiple points on the potential energy surface (PES). This is because the set of overlapping fragments for a given fragment may change over the PES, and at the boundaries of such changes, it is possible for discontinuities to appear. These discontinuities, however, may be minimized by proper treatment of fragment overlap and the use of sufficiently tight overlapping thresholds.

A second issue with the shared basis functions approach is with regard to the derivation of analytic derivatives. This is due to the fact that the set of basis function coefficients to be determined in the typical HF response calculations may grow or shrink at certain points. This is not an intractable problem, but it will require more complex algorithms to account for such changes in comparison to the explicitly localized basis functions approach.

Although it adds certain complexities, the use of shared basis functions is the preferred implementation for the methods presented here, as they offer greater accuracy and do not require perturbative corrections for certain many-body interaction energies. The implementation of such codes will be the focus of future work.

C. Comparisons to existing methods

The ENMO/n methods share a number of features with existing fragmentation methods, while also seeking to improve upon certain shortcomings in these methods. In the following, comparisons are drawn between the ENMO/n methods and two existing classes of fragmentation methods in order to highlight these similarities and differences.

1. Comparison to electrostatically embedded methods

Many of the most popular existing fragmentation methods include a mean-field electrostatic embedding potential, which may be included in a variational formulation[40,41] or in a non-iterative formulation.[43] There are good justifications for each approach and methods based on each are applicable to different systems and have corresponding benefits and drawbacks.

The ENMO/n model improves on common electrostatic embedding fragmentation formalisms in three primary ways through the partial symmetry adaptation of the interfragment embedding potential. First, the inclusion of exchange-repulsion effects penalizes variational collapse when nearly linearly dependent basis functions exist between different fragments. This allows more systematic improvement of the accuracy of calculations through the use of increasing basis set sizes, whereas existing methods, specifically those that are based on self-consistent electrostatically embedded fragments, can exhibit random errors when certain basis sets are used, namely, those with diffuse functions.

The second improvement is in the intrinsic treatment of interfragment bonds. Electrostatically embedded methods must use various restricted-variational-subspace formulations, some of which are, on their own, *ad hoc*,[44] to handle fragmented systems with interfragment covalent bonding. In the methods discussed in the current work, the partial accounting for exchange-repulsion between fragments once again penalizes variational collapse, thereby stabilizing bonds between fragments. It is also still possible to combine the various restricted-variational-subspace methodologies with the ENMO/n methods when retaining the character of the interfragment bonds is critical.

Finally, and most importantly, the partially symmetry-adapted interfragment embedding potential accounts for part of the many-body exchange-repulsion energy at any order of truncation due to the mean-field nature of the potential. Therefore, it does not require the addition of terms via a many-body expansion (MBE) in order to capture such interfragment interaction energies. Because the exchange-repulsion interaction is a component of the self-consistent embedding, the many-body polarization (at the MO level of theory) is tempered by its effects, and thus, certain phenomena due to variational collapse, such as the so-called *polarization collapse*, are less likely to occur due to the increased constraints on the fragment wave functions.[45]

2. Comparison to the Huzinaga method

The equations of the ENMO/1 method resemble, to a certain degree, those of the Huzinaga method[46] and others based upon it.[42,47–50] In particular, the *forms* of the fragment-specific Fock

Published under an exclusive license by AIP Publishing

operators appear, at first glance, to be the same. The Huzinaga Fock operator has the form

$$\widehat{\boldsymbol{F}}_I^{\boldsymbol{H}} = \widehat{\boldsymbol{F}} - \left(\widehat{\boldsymbol{F}}\hat{p}_{(I)} + \hat{p}_{(I)}\widehat{\boldsymbol{F}}\right), \tag{61}$$

where $\widehat{\boldsymbol{F}}$ is the Fock operator in the orthonormal representation. The projection operator, $\hat{p}_{(I)}$, is defined as

$$\begin{aligned}\hat{p}_{(I)} &= \sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J}\hat{p}_{jJ} \\ &= \sum_{J\neq 1}^{M-1}\sum_{j\in J}^{n_J}|\phi_j^J\rangle\langle\phi_j^J|,\end{aligned} \tag{62}$$

and M is again the total number of fragments in the system. One important difference between the methods is that the fragment-specific projection operator in the ENMO/n methods is limited to overlapping fragments, while the analogous projection operator in the Huzinaga method runs over all other fragments in the system. The limitation to overlapping fragments has previously been applied in other methods, such as the GMO method.[42]

The major difference between the methods arises from the difference between the mean-field embedding potential component of the Fock operators. In order to focus solely on these differences, it is first assumed that the PNIO approximation has been applied to the projection operators in both methods, as this assumption only changes the limits on the summation indices for those terms that are shared by both methods. For the additional terms in the exchange-repulsion energy that are unique to the ENMO/n methods and therefore the corresponding fragment Fock operators as well, this assumption has no effect. Then, the ENMO/1 and Huzinaga projection operators for a given fragment are equivalent. The difference between the ENMO/1 and Huzinaga Fock operators for a given fragment, I, denoted ΔF_I, is therefore given by

$$\begin{aligned}\boldsymbol{\Delta F}_I &= \widehat{\boldsymbol{F}}^{(1)} - \widehat{\boldsymbol{F}} \\ &= \left(\widehat{\boldsymbol{F}}^{(1)} - \widehat{\boldsymbol{F}}\right) - \left[\left(\widehat{\boldsymbol{F}}^{(1)} - \widehat{\boldsymbol{F}}\right)\hat{p}_{\boldsymbol{I}_{ovl}} + \hat{p}_{\boldsymbol{I}_{ovl}}\left(\widehat{\boldsymbol{F}}^{(1)} - \widehat{\boldsymbol{F}}\right)\right].\end{aligned} \tag{63}$$

Now, letting

$$\begin{aligned}\delta\widehat{\boldsymbol{F}} &= \widehat{\boldsymbol{F}}^{(1)} - \widehat{\boldsymbol{F}} \\ &= -\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\left(2\hat{J}_{kK,lL} - \widehat{K}_{kK,lL}\right),\end{aligned} \tag{64}$$

the expression for ΔF_I may be rewritten as

$$\begin{aligned}\boldsymbol{\Delta F}_I &= \delta\widehat{\boldsymbol{F}} - \left(\delta\widehat{\boldsymbol{F}}\hat{p}_{\boldsymbol{I}_{ovl}} + \hat{p}_{\boldsymbol{I}_{ovl}}\delta\widehat{\boldsymbol{F}}\right) \\ &= -\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\left(2\hat{J}_{kK,lL} - \widehat{K}_{kK,lL}\right) \\ &\quad + \sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\left[\left(2\hat{J}_{kK,lL} - \widehat{K}_{kK,lL}\right)\hat{p}_{\boldsymbol{I}_{ovl}}\right. \\ &\quad \left. + \hat{p}_{\boldsymbol{I}_{ovl}}\left(2\hat{J}_{kK,lL} - \widehat{K}_{kK,lL}\right)\right].\end{aligned} \tag{65}$$

The corresponding difference in the total energies, which is the difference in the approximate exchange-repulsion energies calculated by the two methods, again assuming the overlapping fragment approximation has been applied, is given by

$$\begin{aligned}\boldsymbol{\Delta}E_{ex-rep} &= E_{ENMO/1} - E_{\boldsymbol{H}} \\ &= \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{\boldsymbol{I}_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\left(2J_{iI,lL}^{kK,lL} - K_{iI,jJ}^{kK,lL}\right).\end{aligned} \tag{66}$$

It has been noted previously that the analogous terms in the exchange-repulsion energy difference between the Huzinaga and conventional NOMO methods, which involve the $\boldsymbol{Z}$ matrix elements rather than MO overlap integrals directly, vanish when the S^2, or single-exchange, approximation is strictly applied not only in the expansion of the inverse MO overlap matrix but also to all resulting terms in the total energy expression.[48] In order to determine the effect of the RHS of Eq. (66) on the total energy expression, note that the zeroth-order energy term in the power series expansion given in Eq. (33) neglects all exchange-repulsion effects. Given that the complete sum of the higher-order terms will always be positive, as exchange is a repulsive interaction in non-orthogonal formulations, one may conclude that the zeroth-order term alone will always be lower in energy than the variationally bound NOMO energy. Thus, the additional terms in the Huzinaga energy expression will raise the energy. More explicitly,

$$\begin{aligned}E_{\boldsymbol{H}}^{ex-rep} &= E_{\boldsymbol{H}} - E_0 \\ &= -2\left[\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{\boldsymbol{I}_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}h_{iI,jJ}\right. \\ &\quad \left. + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\left(2J_{iI,iI}^{kK,lL} - K_{iI,iI}^{kK,lL}\right)\right] \geq 0,\end{aligned} \tag{67}$$

where the equality in the final line of Eq. (67) only holds when all MOs in the total system are mutually orthogonal. In the ENMO/1 method, the effect of Eq. (66) on the total exchange-repulsion energy can be seen by closer inspection. It is immediately apparent that the RHS of Eq. (67) is a first-order correction for the non-orthogonal MOs interacting with the mean-field embedding. This embedding, however, is not by itself completely correct as it is composed of potentially overlapping regions between MOs in different fragments. Therefore, there is an over-correction in the exchange-repulsion due to a double-counting of the non-orthogonal component of the embedding subspace between each pair of overlapping MOs. In the ENMO/1 method, the extra term in the exchange-repulsion energy correction (compared to the Huzinaga method), given in Eq. (66), is a first-order correction to the mean-field embedding. This can be seen more clearly if one formulates the exchange-repulsion terms in the ENMO/1 energy expression in an alternative representation by combining the second and third terms of Eq. (41) and moving the fourth term inside the resulting expression,

Published under an exclusive license by AIP Publishing

$$E^{(1)}_{ex-rep} = -2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{\boldsymbol{I}_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}h_{iI,jJ} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\Bigg[\left(2J^{kK,lL}_{iI,iI} - K^{kK,lL}_{iI,iI}\right) - \frac{1}{2}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{\boldsymbol{I}_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}$$
$$\times\left(2J^{kK,lL}_{iI,jJ} - K^{kK,lL}_{iI,jJ}\right)\Bigg] - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{\boldsymbol{I}_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\Bigg[\left(2J^{kK,kK}_{iI,jJ} - K^{kK,kK}_{iI,jJ}\right) - \frac{1}{2}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\left(2J^{kK,lL}_{iI,jJ} - K^{kK,lL}_{iI,jJ}\right)\Bigg]$$
$$= -2\Bigg\{\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{\boldsymbol{I}_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}h_{iI,jJ} + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{\boldsymbol{I}_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\Bigg[\left(2J^{kK,kK}_{iI,jJ} - K^{kK,kK}_{iI,jJ}\right) - \frac{1}{2}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\left(2J^{kK,lL}_{iI,jJ} - K^{kK,lL}_{iI,jJ}\right)\Bigg]\Bigg\}. \tag{68}$$

Equation (68) shows that the exchange-repulsion energy for a pair of overlapping MOs interacting with the mean-field embedding potential is partially corrected by the subtraction of doubly counted embedding terms due to overlapping MOs between fragments, given in the last term in brackets, to account for the non-orthogonality of the MOs that comprise this embedding. Of course, this correction is only a first approximation, and higher-order terms, such as those that appear in ENMO/2 or ENMO/3, improve this correction. Although the complete correction for such overestimation is only obtained in the exact NOMO formulation (i.e., the ENMO/n limit), the first-order approximation in ENMO/1 is expected to capture a significant part of this interaction, as all higher-order corrections will consist of products of 3-or-more MO overlap integrals.

Although this is, in general, an improvement, as it reduces the overestimation of the exchange-repulsion energy predicted by the Huzinaga method, it remains to be seen if the parallelity of the Huzinaga method (the relatively constant error in the total energy between the exact NOMO and Huzinaga methods along a given PES) is retained in the ENMO/1 method.

D. Theoretical computational scaling

In order to study the scaling of calculations that employ the ENMO/n methods at various orders, it is useful to define a set of constants. First, let the system under study be composed of M fragments. The number of fragments that overlap with a given fragment, $M_{\boldsymbol{I}_{ovl}}$, is chosen to be equal for each fragment, and this constant is denoted M_{ovl},

$$M_{1_{ovl}} = M_{2_{ovl}} = \cdots = M_{M_{ovl}} = M_{ovl}. \tag{69}$$

Generally, the number of overlapping fragments will vary for each fragment in actual calculations. However, this variation can be ignored for the present purpose of investigating ideal scaling behaviors. If it is assumed that the value of M_{ovl} is chosen as the largest value for the set of fragments in a total system, then the true prefactor in a given calculation will be smaller than that predicted by the theoretical expression.

For simplicity, it is assumed that the number of basis functions for a given fragment, K_I, is equal for all fragments. Thus,

$$K_1 = K_2 = \cdots = K_M. \tag{70}$$

Therefore, the number of unique basis functions for the total system is given by

$$K = K_1 + K_2 + \cdots + K_M = MK_M. \tag{71}$$

With these quantities defined, one can study the theoretical scaling of the ENMO/n methods, at various levels of truncation, with respect to the parameters M, M_{ovl}, K_M, and M. To determine the theoretical scaling for various ENMO/n methods, it is necessary to first rewrite the expression for an arbitrary individual fragment Fock matrix, $\boldsymbol{F}_I^{(n)}$, in the AO basis,

$$\left(\boldsymbol{F}_I^{(n)}\right)_{\mu I,\nu I} = \left(\chi^I_\mu\middle|\widehat{\boldsymbol{F}}^{(n)}_I\middle|\chi^I_\nu\right)$$
$$= \left(\chi^I_\mu\middle|\widehat{\boldsymbol{F}}^{(n)}\middle|\chi^I_\nu\right) - \left(\chi^I_\mu\middle|\left[\widehat{\boldsymbol{F}}^{(n)}\left(\hat{p}^{(n)}_{\boldsymbol{I}_{ovl}}\right)^\dagger + \hat{p}^{(n)}_{\boldsymbol{I}_{ovl}}\widehat{\boldsymbol{F}}^{(n)}\right]\middle|\chi^I_\nu\right)$$
$$- \sum_{m=1}^{n}\sum_{a}^{n-1}\left(\chi^I_\mu\middle|p^{(a)}_{\boldsymbol{I}_{ovl}}\boldsymbol{F}^{(n)}\left(p^{(m-a)}_{\boldsymbol{I}_{ovl}}\right)^\dagger\middle|\chi^I_\nu\right), \tag{72}$$

where χ^I_μ denotes the μth AO basis function assigned to the Ith fragment. The full expression for the ENMO/1 Fock operator, in the AO basis, is given in the Subsection 1 of the Appendix for reference.

The complexity of computing the total Fock matrix scales with respect to the various parameters that affect the summation indices in the respective Fock matrix expressions in the AO basis. For example, the scaling of computing the core Hamiltonian matrix (in big O notation) is given by

$$\boldsymbol{H}_I^{core} = \mathcal{O}\left(\sum_{\mu\in I}^{K_I}\left(\boldsymbol{H}_I^{core}\right)_{\mu\nu}\right)$$
$$= \mathcal{O}\left(K_I^2\right). \tag{73}$$

Of course, eliminating redundant terms can reduce the scaling by some constant prefactor, but the exponential scaling with respect to the particular parameter [K_I in Eq. (73)] is unaffected. The scaling of the computation of the full Fock matrix for each fragment, utilizing the constants given in Eqs. (69)–(71), may be written approximately, focusing on the leading (largest) term in the scaling expression, as

$$\boldsymbol{F}_I^{(1)} \approx \mathcal{O}\left(KM_{ovl}^2K_M^4\right). \tag{74}$$

This approach may be continued for higher-order ENMO/n Fock operators in order to derive big O scaling expressions for these methods. For the second-order (ENMO/2) case, the leading term in the scaling expression is

Published under an exclusive license by AIP Publishing

$$\boldsymbol{F}_I^{(2)} \approx \mathcal{O}\left(KM_{ovl}^4 K_M^4\right). \tag{75}$$

Finally, for the third-order (ENMO/3) case, the leading term in the scaling expression is

$$\boldsymbol{F}_I^{(3)} \approx \mathcal{O}\left(KM_{ovl}^6 K_M^4\right). \tag{76}$$

From Eqs. (74)–(76), it is apparent that increasing the order of the ENMO/n method from n to $n + 1$ increases the total scaling expression primarily by the addition of a term of the form

$$\boldsymbol{F}_I^{(n+1)} \approx \mathcal{O}\left(KM_{ovl}^{2n} K_M^4\right). \tag{77}$$

For computational tractability and applicability to the study of large, complex chemical systems, the most important parameter is the size of the system, since the size impacts both the computer time demand and the memory requirements. The size parameter is generally chosen as the total number of basis functions, K, as this value is directly proportional to the number of floating-point operations required in a given computation. Below, the approximate theoretical scaling of the ENMO/n methods, for n = 1, 2, 3, with respect to the value of K is shown. For the plot in Fig. 5, the number of basis functions assigned to each fragment, K_M, is chosen as 100 and the number of overlapping fragments assigned to each fragment, M_{ovl}, is chosen as 2. These values are, of course, arbitrary, and any choice would produce the same set of linearly scaling plots, with only the prefactors (slopes) being affected.

Figure 5 illustrates the linear-scaling nature of the theoretical computational complexity for each of the ENMO/n methods with respect to the total number of basis functions, K. The size of the system in terms of the number of fragments, M, which can be an equally useful metric compared to K as it is proportional to the total number of atoms in the system, varies from 100 to 102 400. In the plot, subsequent points along each line represent a doubling of the fragment count from the previous point.

A critical feature of the plots that is also worth noting is the relative size of the slopes of each line. Because the scaling prefactor grows exponentially with respect to the number of overlapping fragments assigned to each fragment, M_{ovl}, as well as the truncation order of the inverse MO overlap matrix expansion, n, the ratio of the slopes for each line also grows exponentially, and so the computational cost is dramatically affected by the choice of M_{ovl}. This ratio, for two ENMO/n methods with truncation orders, n and $n + 1$, is M_{ovl}^2.

The increasing computational complexity due to changes in the choice of value for M_{ovl} is shown more clearly in Fig. 6. The theoretical prefactor scaling is shown for each of the ENMO/n methods outlined in the text for M_{ovl} values of 2, 4, 6, 8, 10, and 12.

The log–log plot of scaling prefactors in Fig. 6 further demonstrates the rapid growth in computational complexity as a result of increasing the value of M_{ovl} for a given ENMO/n method. Together with Fig. 5, it is clear that for many cases, the limiting factor in computational feasibility when applying the ENMO/n methods is the value of M_{ovl}, rather than the total size of the system. It is therefore critical in a given calculation to choose M_{ovl} as the smallest value possible, which retains quantitative accuracy. Future work will focus on determining this value on the fly for a given calculation in order to choose the appropriate method to be employed. Although this scaling is exponential for any ENMO/n method, it is important to note that the value of M_{ovl} will be dependent on the *chemical nature* of the system (e.g., delocalization of MOs) and not directly on the *physical size* of the system. Of course, there are systems, such as those with periodicity, where delocalization may be intrinsic and the size of M_{ovl} will grow directly in proportion with M. These types of systems, however, are not well-suited for study by the use of *any* of the fragmentation methods that are specifically based on the assumption of fragment-localized MOs.

The final parameter that influences the computational complexity of each of the ENMO/n methods is the number of basis functions assigned to each fragment, K_M. This parameter,

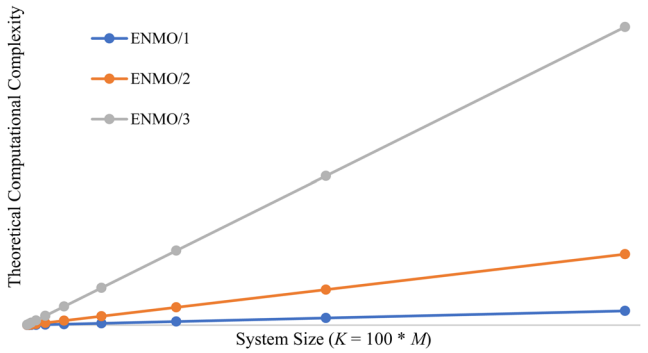


FIG. 5. Theoretical scaling of the ENMO/1, ENMO/2, and ENMO/3 methods with respect to the total number of basis functions.

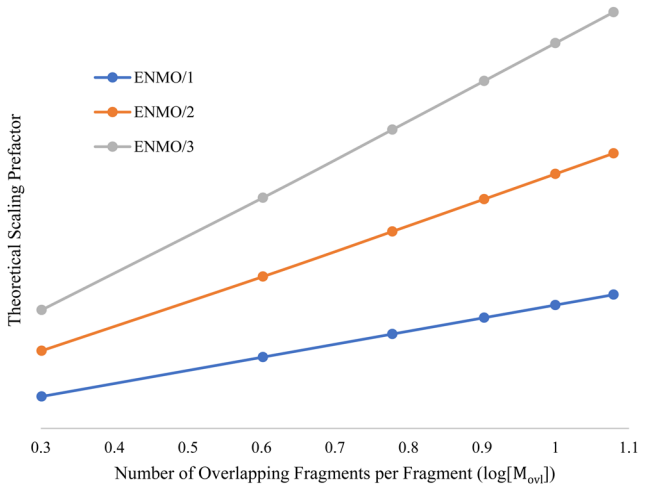


FIG. 6. Theoretical scaling prefactor of ENMO/1, ENMO/2, and ENMO/3 with respect to the number of overlapping fragments assigned to each fragment, M_{ovl} (log–log plot).

Published under an exclusive license by AIP Publishing

however, has a constant quartic contribution to the scaling prefactor at each value of n. Therefore, the K_M parameter only affects the computational complexity of a given ENMO/n calculation indirectly through its effect on the total number of basis functions, K. Similarly, the total number of fragments, M, is proportional to the total number of basis functions, and thus, each ENMO/n method scales linearly with respect to the number of fragments in the total system.

V. CONCLUSIONS AND FUTURE WORK

An approximate formulation of the NOMO model of electronic structure theory has been derived, and from it, a hierarchy of *ab initio* fragment-based quantum chemistry methods, referred to as the ENMO/n methods, have been developed. These methods incorporate interfragment exchange-repulsion interactions that are not accounted for in many existing methods and refine the previously utilized embedding potential of certain other methods.

The outlined methods seek to reproduce the results of conventional MO theory at a substantially reduced computational cost. Consequently, the methods neglect critical interfragment interactions that are accounted for in post-SCF methods. This omission may be remedied in two straightforward ways. For some calculations, post-SCF correlation corrections may be easily included through the use of a many-body expansion in terms of correlated N-mers, as has been done in existing fragmentation methods, such as the FMO and Effective Fragment Molecular Orbital (EFMO) methods.[29,34,35] This MBE-based approach to post-SCF corrections has been shown to be highly accurate in existing fragmentation methods, and the same can be expected from the ENMO/n methods. For each truncation order of the ENMO/n method, it is also possible to develop a novel formulation of the Symmetry-adapted Perturbation Theory (SAPT) for intermolecular interactions by using the ENMO/n fragment wave functions as a reference state, an approach that has been previously studied using electrostatically embedded fragment methods.[51–53] The development and application of these correlation-consistent extensions to the outlined methods will be investigated in future work. Note that certain correlation effects may be accounted for through the use of a Kohn–Sham Density Functional Theory (KS-DFT) based on the ENMO/n method, and this too will be studied.

Note also that the approximations employed in the derivation of the ENMO/n methods, specifically the expansion of the $\boldsymbol{Z}$ matrix followed by the application of the PNIO approximation to fragment pairs, can be applied in other non-orthogonal MO methods. The approach may, for example, be applied to single-molecule systems by partitioning MOs into subgroups and applying the PNIO approximation to certain pairs of subgroups.

Finally, as mentioned in the Introduction, extension of the outlined methods to multi-reference states will allow for a much wider variety of chemical systems of interest to be studied. The derivation of the ENMO/n equations for such states will be developed and used to study model systems of large chemical systems with one or more fragments, which require the use of multi-reference states.

ACKNOWLEDGMENTS

This work was supported by the Department of Energy GAMESS Exascale Computing Project under Project No. 17-SC-20-SC, administered by the Ames Laboratory, which is operated by Iowa State University under Contract No. DE-AC02-07CH11338. B.M.W. acknowledges a fellowship from the Molecular Sciences Software Institute (MolSSI) under NSF Grant No. ACI-1547580.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

APPENDIX: EXPLICIT DERIVATIONS OF EXPRESSIONS REFERENCED WITHIN THE TEXT

1. Matrix eigenvalue equations for the ENMO/1 method

Here, the AO basis matrix equations necessary for solving for each fragment wave function in a system will be given for the FNMO (ENMO/1) method. For a particular fragment, I, the basis functions associated with MOs assigned to the fragment will be denoted χ_μ^I, with the number of these basis functions given by K_I. The mean-field embedding component of the Fock operator, F, in the AO basis, is then given by

$$\boldsymbol{F} = \boldsymbol{H}^{Core} + \sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{\alpha\in K}^{K_K}\sum_{\beta\in K}^{K_K} C_{k\alpha}C_{k\beta} \times \left\{\left(2J_{\alpha\beta} - K_{\alpha\beta}\right) - \sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L}\sum_{\gamma\in L}^{K_L}\sum_{\delta\in L}^{K_L} C_{l\gamma}C_{l\delta}s_{\beta K,\delta L}\left(2J_{\alpha\gamma} - K_{\beta\gamma}\right)\right\}. \tag{A1}$$

In order to make the definition of later quantities more readily apparent, this expression will be expanded into a less concise form

$$\boldsymbol{F} = \boldsymbol{H}^{Core} + \sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{\alpha\in K}^{K_K}\sum_{\beta\in K}^{K_K} 2C_{k\alpha}C_{k\beta}\left(J_{\alpha\beta} - \frac{1}{2}K_{\alpha\beta}\right) - \sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{\alpha\in K}^{K_K}\sum_{\beta K}^{K_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L}\sum_{\gamma\in L}^{K_L}\sum_{\delta\in L}^{K_L} 2C_{k\alpha}C_{k\beta}C_{l\gamma}C_{l\delta}s_{\beta K,\delta L}\left(J_{\alpha\gamma} - \frac{1}{2}K_{\alpha\gamma}\right). \tag{A2}$$

Now, density matrices for fragment monomers and dimers are defined, which will be denoted D^K and D^{KL} for the monomer K and the K, L fragment pair, respectively. The density matrices are defined as

$$\boldsymbol{D}_{\alpha\beta}^K = 2\sum_{k\in K}^{n_K} C_{k\alpha}C_{k\beta} \tag{A3}$$

and

Published under an exclusive license by AIP Publishing

$$\boldsymbol{D}^{KL}_{\alpha\gamma} = \sum_{k\in K}^{n_K}\sum_{\alpha\in K}^{K_K}\sum_{\beta\in K}^{K_K}\sum_{l\in L}^{n_L}\sum_{\gamma\in L}^{K_L}\sum_{\delta\in L}^{K_L} C_{k\alpha}C_{k\beta}s_{\beta K,\delta L}C_{l\gamma}C_{l\delta}$$
$$= \sum_{\beta\in K}^{K_K}\sum_{\delta\in L}^{K_L}\boldsymbol{D}^{K}_{\alpha\beta}s^{KL}_{\beta\delta}\boldsymbol{D}^{L}_{\gamma\delta}. \tag{A4}$$

Utilizing the monomer and dimer density matrices allows the Fock operator to be written in the simplified form

$$\boldsymbol{F} = \boldsymbol{H}^{Core} + \sum_{K=1}^{M}\sum_{\alpha\in K}^{K}\sum_{\beta\in K}^{K_K}\boldsymbol{D}_{\alpha\beta}\left(J_{\varepsilon\beta} - \frac{1}{2}K_{\alpha\beta}\right)$$
$$- \frac{1}{2}\sum_{K=1}^{M}\sum_{\alpha\in K}^{K_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{\gamma\in L}^{K_L}\boldsymbol{D}^{KL}_{\alpha\gamma}\left(J_{\alpha\gamma} - \frac{1}{2}K_{\alpha\gamma}\right). \tag{A5}$$

It is useful to note that each pair of overlapping fragments appears twice in the third term of Eq. (A5). Limiting the summation to unique overlapping fragment pairs, the total number of which will be denoted M^{total}_{ovl}, gives

$$\boldsymbol{F} = \boldsymbol{H}^{Core} + \sum_{K=1}^{M}\sum_{\alpha\in K}^{K_K}\sum_{\beta\in K}^{K_K}\boldsymbol{D}_{\alpha\beta}\left(J_{\alpha\beta} - \frac{1}{2}K_{\alpha\beta}\right)$$
$$- \sum_{K,L}^{M^{total}_{ovl}}\sum_{\alpha\in K}^{K_K}\sum_{\gamma\in L}^{K_L}\boldsymbol{D}^{KL}_{\alpha\gamma}\left(J_{\alpha\gamma} - \frac{1}{2}K_{\alpha\gamma}\right), \tag{A6}$$

from which it follows that the mean-field embedding Fock matrix, in the Ith basis, is expressed as

$$\boldsymbol{F}_{\mu\nu} = \boldsymbol{H}^{core}_{\mu\nu} + \sum_{K=1}^{M}\sum_{\alpha\in K}^{K_K}\sum_{\beta\in K}^{K_K}\boldsymbol{D}^{K}_{\alpha\beta}\left[(\mu\nu|\alpha\beta) - \frac{1}{2}(\mu\beta|\alpha\nu)\right]$$
$$- \sum_{K,L}^{M^{total}_{ovl}}\sum_{\alpha\in K}^{K_K}\sum_{\gamma\in L}^{K_L}\boldsymbol{D}^{KL}_{\alpha\gamma}\left[(\mu\nu|\alpha\gamma) - \frac{1}{2}(\mu\gamma|\alpha\nu)\right]. \tag{A7}$$

In order to express the unique terms in the Ith fragment Fock operator, $\widehat{\boldsymbol{F}}_I$, begin by noting that the projection operator onto the overlapping subspace, for the Ith fragment, is given by

$$\hat{p}_{\boldsymbol{I}_{ovl}} = \sum_{J\in\boldsymbol{I}_{ovl}}^{M_{\boldsymbol{I}_{ovl}}}\sum_{j\in J}^{n_J}\sum_{\lambda\in J}^{K_J}\sum_{\sigma\in J}^{K_J} C_{j\lambda}C_{j\sigma}|\chi^J_\lambda\rangle\langle\chi^J_\sigma|$$
$$= \frac{1}{2}\sum_{J\in\boldsymbol{I}_{ovl}}^{M_{\boldsymbol{I}_{ovl}}}\sum_{\lambda\in J}^{K_J}\sum_{\sigma\in J}^{K_J}\boldsymbol{D}^{J}_{\lambda\sigma}|\chi^J_\lambda\rangle\langle\chi^J_\sigma|. \tag{A8}$$

Simple expressions for the two terms $\boldsymbol{F}\hat{p}_{\boldsymbol{I}_{ovl}}$ and $\hat{p}_{\boldsymbol{I}_{ovl}}\boldsymbol{F}$ may now be given as

$$(\chi^I_\mu|\boldsymbol{F}\hat{p}_{\boldsymbol{I}_{ovl}}|\chi^I_\nu) = \frac{1}{2}\sum_{J\in\boldsymbol{I}_{ovl}}^{M_{\boldsymbol{I}_{ovl}}}\sum_{\lambda\in J}^{K_J}\sum_{\sigma\in J}^{K_J}\boldsymbol{D}^{J}_{\lambda\sigma}(\chi^I_\mu|\boldsymbol{F}|\chi^J_\lambda)(\chi^J_\sigma|\chi^I_\nu)$$
$$= \frac{1}{2}\sum_{J\in\boldsymbol{I}_{ovl}}^{M_{\boldsymbol{I}_{ovl}}}\sum_{\lambda\in J}^{K_J}\sum_{\sigma\in J}^{K_J}\boldsymbol{D}^{J}_{\lambda\sigma}\boldsymbol{F}_{\mu I,\lambda J}s^{IJ}_{\sigma J,\nu I} \tag{A9}$$

and

$$(\chi^I_\mu|\hat{p}_{\boldsymbol{I}_{ovl}}\boldsymbol{F}|\chi^I_\nu) = \frac{1}{2}\sum_{J\in\boldsymbol{I}_{ovl}}^{M_{\boldsymbol{I}_{ovl}}}\sum_{\lambda\in J}^{K_J}\sum_{\sigma\in J}^{K_J}\boldsymbol{D}^{J}_{\lambda\sigma}(\chi^I_\mu|\chi^J_\lambda)(\chi^J_\sigma|\boldsymbol{F}|\chi^I_\nu)$$
$$= \frac{1}{2}\sum_{J\in\boldsymbol{I}_{ovl}}^{M_{\boldsymbol{I}_{ovl}}}\sum_{\lambda\in J}^{K_J}\sum_{\sigma\in J}^{K_J}\boldsymbol{D}^{J}_{\lambda\sigma}s^{IJ}_{\mu I,\lambda J}\boldsymbol{F}_{\sigma J,\nu I}. \tag{A10}$$

In Eqs. (A9) and (A10), the $s^{IJ}_{\sigma J,\nu I}$ and $s^{IJ}_{\mu I,\lambda J}$ matrices are K_I by K_J matrices whose elements are each an AO overlap integral between an AO assigned to the Ith fragment and an AO assigned to the Jth fragment. In order to further reduce the scaling of the algorithm, the corresponding intermediate matrices, given by $\frac{1}{2}\sum_{\sigma\in J}^{K_J}\boldsymbol{D}^{J}_{\lambda\sigma}s^{IJ}_{\sigma J,\nu I}$ and $\frac{1}{2}\sum_{\lambda\in J}^{K_J}\boldsymbol{D}^{J}_{\lambda\sigma}s^{IJ}_{\mu J,\lambda I}$, respectively, may be precomputed. Let $\tilde{\boldsymbol{D}}^{IJ}_{\lambda J,\nu I}$ and $\tilde{\boldsymbol{D}}^{IJ}_{\sigma J,\mu I}$ represent the resulting matrices, respectively. Note that these are partial transformations and can be used to compute the dimer overlap matrices given in Eq. (A4). Using the $\tilde{\boldsymbol{D}}^{IJ}_{\lambda J,\nu I}$ and $\tilde{\boldsymbol{D}}^{IJ}_{\sigma J,\mu I}$ quantities, Eq. (A7) may be rewritten in the final form

$$(\boldsymbol{F}_I)_{\mu\nu} = \boldsymbol{H}^{core}_{\mu\nu} + \sum_{K=1}^{M}\sum_{\alpha\in K}^{K_K}\sum_{\beta\in K}^{K_K}\boldsymbol{D}^{K}_{\alpha\beta}\left[(\mu\nu|\alpha\beta) - \frac{1}{2}(\mu\beta|\alpha\nu)\right] - \sum_{K,L}^{M^{total}_{ovl}}\sum_{\alpha\in K}^{K_K}\sum_{\gamma\in L}^{K_L}\boldsymbol{D}^{KL}_{\alpha\gamma}\left[(\mu\nu|\alpha\gamma) - \frac{1}{2}(\mu\gamma|\alpha\nu)\right] - \sum_{J\in\boldsymbol{I}_{ovl}}^{M_{\boldsymbol{I}_{ovl}}}\sum_{\lambda\in J}^{K_J}\tilde{\boldsymbol{D}}^{IJ}_{\lambda J,\nu I}$$
$$\times\left\{\left[\boldsymbol{H}^{core}_{\mu I,\lambda J} + \sum_{K=1}^{M}\sum_{\alpha\in K}^{K_K}\sum_{\beta\in K}^{K_K}\boldsymbol{D}^{K}_{\alpha\beta}\left[(\mu\lambda|\alpha\beta) - \frac{1}{2}(\mu\beta|\alpha\lambda)\right] - \sum_{K,L}^{M^{total}_{ovl}}\sum_{\alpha\in K}^{K_K}\sum_{\gamma\in L}^{K_L}\boldsymbol{D}^{KL}_{\alpha\gamma}\left[(\mu\lambda|\alpha\gamma) - \frac{1}{2}(\mu\gamma|\alpha\lambda)\right]\right]\right\}$$
$$+ \sum_{J\in\boldsymbol{I}_{ovl}}^{M_{\boldsymbol{I}_{ovl}}}\sum_{\delta\in J}^{K_J}\tilde{\boldsymbol{D}}^{IJ}_{\sigma J,\mu I}\left[\boldsymbol{H}^{core}_{\sigma J,\nu I} + \sum_{K=1}^{M}\sum_{\alpha\in K}^{K_K}\sum_{\beta\in K}^{K_K}\boldsymbol{D}^{K}_{\alpha\beta}\left[(\sigma\nu|\alpha\beta) - \frac{1}{2}(\sigma\beta|\alpha\nu)\right] - \sum_{K,L}^{M^{total}_{ovl}}\sum_{\alpha\in K}^{K_K}\sum_{\gamma\in L}^{K_L}\boldsymbol{D}^{KL}_{\alpha\gamma}\left[(\sigma\nu|\alpha\gamma) - \frac{1}{2}(\sigma\gamma|\alpha\nu)\right]\right]. \tag{A11}$$

2. Derivation of the unique terms in the *n*th-order exchange-repulsion energy

The unique terms in the approximate exchange-repulsion energy correction in the ENMO/n method, denoted $\delta E^{(n)}_{ex-rep}$, may be obtained by simply taking the difference between the ENMO/n and ENMO/$n-1$ exchange-repulsion energies,

Published under an exclusive license by AIP Publishing

$$
\begin{aligned}
E^{(n)}_{ex-rep} - E^{(n-1)}_{ex-rep} &= \delta E^{(n)}_{ex-rep} \\
&= \sum_{m=0}^{n-1}(-1)^m\Bigg\{-2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in I_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s'^{m}_{jJ,pP}h_{iI,pP} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in L_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} \boldsymbol{S}'^{m}_{lL,rR}\left(2J^{kK,rR}_{iI,iI} - K^{kK,rR}_{iI,iI}\right) \\
&\quad - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P}\boldsymbol{S}'^{m}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left(2J^{kK,kK}_{iI,pP} - K^{kK,kK}_{iI,pP}\right) + \sum_{m'=0}^{n-1}(-1)m'\Bigg[\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P}\boldsymbol{S}'^{m}_{jJ,pP} \\
&\quad \times \sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R=1}^{M}\sum_{r\in R}^{n_R}\boldsymbol{S}'^{m'}_{lL,rR}\left(2J^{kK,rR}_{iI,lL} - K^{kK,kK}_{iI,lL}\right)\Bigg]\Bigg\} \\
&\quad - \sum_{m=0}^{n-2}(-1)^m\Bigg\{-2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P}\boldsymbol{S}'^{m}_{jJ,pP}h_{iI,pP} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R=1}^{M}\sum_{r\in R}^{n_R}\boldsymbol{S}'^{m}_{lL,rR}\left(2J^{kK,rR}_{iI,iI} - K^{kK,rR}_{iI,iI}\right) \\
&\quad - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in I_{ovl}}^{M-1}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P}\boldsymbol{S}'^{m}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left(2J^{kK,kK}_{iI,pP} - K^{kK,kK}_{iI,pP}\right) + \sum_{m'=0}^{n-2}(-1)m' \\
&\quad \times\Bigg[\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P}\boldsymbol{S}'^{m}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\neq K}^{M-1}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R=1}^{M}\sum_{r\in R}^{n_R}\boldsymbol{S}'^{m'}_{lL,rR}\left(2J^{kK,kK}_{iI,pP} - K^{kK,rR}_{iI,lL}\right)\Bigg]\Bigg\}.
\end{aligned}
\tag{A12}
$$

Eliminating common terms between $E^{(n)}_{ex-rep}$ and $E^{(n-1)}_{ex-rep}$, Eq. (A12) simplifies to

$$
\begin{aligned}
\delta E^{(n)}_{ex-rep} &= (-1)^{n-1}\Bigg\{-2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P}\boldsymbol{S}'^{n-1}_{jJ,pP}h_{iI,pP} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\neq K}^{M-1}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R=1}^{M}\sum_{r\in R}^{n_R}\boldsymbol{S}'^{n-1}_{lL,rR}\left(2J^{kK,rR}_{iI,iI} - K^{kK,rR}_{iI,iI}\right) \\
&\quad - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P}\boldsymbol{S}'^{n-1}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left(2J^{kK,kK}_{iI,pP} - K^{kK,kK}_{iI,pP}\right) + \sum_{m'=0}^{n-2}(-1)m' \\
&\quad \times\Bigg[\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P}\boldsymbol{S}'^{n-1}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\neq K}^{M-1}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R=1}^{M}\sum_{r\in R}^{n_R}\boldsymbol{S}'^{m'}_{lL,rR}\left(2J^{kK,rR}_{iI,lL} - K^{kK,rR}_{iI,lL}\right)\Bigg] \\
&\quad + \sum_{m=0}^{n-2}(-1)^m\Bigg[\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P}\boldsymbol{S}'^{m}_{jJ,pP}\times\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\neq K}^{M-1}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R=1}^{M}\sum_{r\in R}^{n_R}\boldsymbol{S}'^{n-1}_{lL,rR}\left(2J^{kK,rR}_{iI,lL} - K^{kK,rR}_{iI,lL}\right)\Bigg] \\
&\quad + (-1)^{n-1}\Bigg[\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\in P}^{n_P}\boldsymbol{S}'^{n-1}_{jJ,pP}\times\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\neq K}^{M-1}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R=1}^{M}\sum_{r\in R}^{n_R}\boldsymbol{S}'^{n-1}_{lL,rR}\left(2J^{kK,rR}_{iI,lL} - K^{kK,rR}_{iI,lL}\right)\Bigg]\Bigg\}.
\end{aligned}
\tag{A13}
$$

Moving the third, fourth, and fifth terms out of the braced expression then leads to the expression

$$
\begin{aligned}
\delta E^{(n)}_{ex-rep} &= (-1)^{n-1}\Bigg\{-2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S}'^{n-1}_{jJ,pP}h_{iI,pP} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\neq K}^{M-1}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{S}'^{n-1}_{lL,rR}\left(2J^{kK,rR}_{iI,iI} - K^{kK,rR}_{iI,iI}\right) - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S}'^{n-1}_{jJ,pP} \\
&\quad \times\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left(2J^{kK,kK}_{iI,pP} - K^{kK,kK}_{iI,pP}\right)\Bigg\} + (-1)^{2(n-1)}\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S}'^{n-1}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\neq K}^{M-1}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{S}'^{n-1}_{iL,lL}\left(2J^{kK,rR}_{iL,lL} - K^{kK,rR}_{iL,lL}\right) \\
&\quad + (-1)^{n-1}\sum_{m'=0}^{n-2}(-1)m'\Bigg[\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S}'^{n-1}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\neq K}^{M-1}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{S}'^{m'}_{lL,rR}\left(2J^{kK,rR}_{iI,lL} - K^{kK,rR}_{iI,lL}\right)\Bigg] \\
&\quad + (-1)^{n-1}\sum_{m=0}^{n-2}(-1)^m\Bigg[\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S}'^{\mu}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\neq K}^{M-1}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{S}'^{(n-1)}_{lL,rR}\left(2J^{kK,rR}_{iI,lL} - K^{kK,rR}_{iI,lL}\right)\Bigg].
\end{aligned}
\tag{A14}
$$

Published under an exclusive license by AIP Publishing

The factor of $(-1)^{2(n-1)}$ may be eliminated, as it is always equal to unity. The last two terms in Eq. (A14) may be rewritten in terms of the truncated $\boldsymbol{T}$ matrix expansion as

$$\begin{aligned}
\delta E^{(n)}_{ex-rep} = (-1)^{n-1}\Bigg\{ & -2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S'}^{n-1}_{jJ,pP}h_{iI,pP} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\neq K}^{M-1}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{S'}^{n-1}_{lL,rR}\left(2J^{kK,rR}_{iI,iI} - K^{kK,rR}_{iI,iI}\right) \\
& - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S'}^{n-1}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left(2J^{kK,kK}_{iI,pP} - K^{kK,kK}_{iI,pP}\right)\Bigg\} + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S'}^{n-1}_{jJ,pP} \\
& \times \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{L\neq K}^{M-1}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{S'}^{n-1}_{lL,rR}\left(2J^{kK,rR}_{iI,lL} - K^{kK,rR}_{iI,lL}\right) + (-1)^{n-1}\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S'}^{n-1}_{jJ,pP} \\
& \times \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{L\neq K}^{M-1}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{T}^{n-1}_{lL,rR}\left(2J^{kK,rR}_{iI,lL} - K^{kK,rR}_{iI,lL}\right) + (-1)^{n-1}\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{T}^{n-1}_{jJ,pP} \\
& \times \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{L\neq K}^{M-1}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{S}^{n-1}_{lL,rR}\left(2J^{kK,rR}_{iI,lL} - K^{kK,rR}_{iI,lL}\right).
\end{aligned}$$

Moving the terms with elements of the $\boldsymbol{T}$ matrix in their summations into the expression in braces, Eq. (A15) may be rewritten as

$$\begin{aligned}
\delta E^{(n)}_{ex-rep} = (-1)^{n-1}\Bigg\{ & -2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S'}^{n-1}_{jJ,pP}h_{iI,pP} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\neq K}^{M-1}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{S'}^{n-1}_{lL,rR} \\
& \times \left[\left(2J^{kK,rR}_{iI,iI} - K^{kK,rR}_{iI,iI}\right) - \sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{T}^{(n-1)}_{jJ,pP}\left(2J^{kK,rR}_{iI,lL} - K^{kK,rR}_{iI,lL}\right)\right] \\
& - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S'}^{n-1}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left[\left(2J^{kK,kK}_{iI,pP} - K^{kK,kK}_{iI,pP}\right) - \sum_{L\neq K}^{M-1}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{T}^{(n-1)}_{lL,rR}\left(2J^{kK,rR}_{iI,lL} - K^{kK,rR}_{iI,lL}\right)\right]\Bigg\} \\
& + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\neq I}^{M-1}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S'}^{n-1}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\neq K}^{M-1}\sum_{l=L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{S'}^{n-1}_{lL,rR}\left(2J^{kK,rR}_{iI,lL} - K^{kK,rR}_{iI,lL}\right).
\end{aligned} \tag{A15}$$

Thus, the expression given in the main text in Eq. (35) is obtained for the unique terms in the exchange-repulsion component of NOMO energy at the nth-order. For the ENMO/n methods specifically, the PNIO approximation is also applied to all fragment pairs, giving the following simplified expression for $\delta E^{(n)}_{ex-rep}$, which involves restricted summations over fragment pairs:

$$\begin{aligned}
\delta E^{(n)}_{ex-rep} = (-1)^{n-1}\Bigg\{ & -2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S'}^{n-1}_{jJ,pP}h_{iI,pP} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R=1}^{M}\sum_{r\in R}^{n_R}\boldsymbol{S'}^{n-1}_{lL,rR} \\
& \times \left[\left(2J^{kK,rR}_{iI,iI} - K^{kK,rR}_{iI,iI}\right) - \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{T}^{(n-1)}_{jJ,pP}\left(2J^{kK,rR}_{iI,pP} - K^{kK,rR}_{iI,pP}\right)\right] - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_j}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S'}^{n-1}_{jJ,pP} \\
& \times \sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left[\left(2J^{kK,kK}_{iI,pP} - K^{kK,kK}_{iI,pP}\right) - \sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{T}^{(n-1)}_{lL,rR}\left(2J^{kK,rR}_{iI,pP} - K^{kK,rR}_{iI,pP}\right)\right]\Bigg\} \\
& + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_j}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S'}^{n-1}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{S'}^{n-1}_{lL,rR}\left(2J^{kK,rR}_{iI,pP} - K^{kK,rR}_{iI,pP}\right).
\end{aligned} \tag{A16}$$

Published under an exclusive license by AIP Publishing

3. Derivation of the first-order exchange-repulsion energy

As $E^{(0)}_{ex-rep}$ is zero by definition, the first-order exchange-repulsion energy, $E^{(1)}_{ex-rep}$, is equivalently given by $\delta E^{(1)}_{ex-rep}$,

$$
\begin{aligned}
E^{(1)}_{ex-rep} &= \delta E^{(1)}_{ex-rep} \\
&= (-1)^0 \Bigg\{ -2 \sum_{I=1}^{M} \sum_{i\in I}^{n_I} \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}} \sum_{j\in J}^{n_J} \sum_{P=1}^{M} \sum_{p\in P}^{n_P} s_{iI,jJ} \boldsymbol{S'}^{0}_{jJ,pP} h_{iI,pP} - \sum_{I=1}^{M} \sum_{i\in I}^{n_I} \sum_{K=1}^{M} \sum_{k\in K}^{n_K} \sum_{L\in K_{ovl}}^{M_{K_{ovl}}} \sum_{l\in L}^{n_L} \sum_{R=1}^{M} \sum_{r\in R}^{n_R} s_{kK,lL} \boldsymbol{S'}^{0}_{lL,rR} \\
&\times \left[\left(2J^{kK,rR}_{iI,iI} - K^{kK,rR}_{iI,iI} \right) - \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}} \sum_{j\in J}^{n_J} \sum_{P=1}^{M} \sum_{p\in P}^{n_P} s_{iI,jJ} \boldsymbol{T}^{(0)}_{jJ,pP} \left(2J^{kK,rR}_{iI,pP} - K^{kK,rR}_{iI,pP} \right) \right] \\
&- \sum_{I=1}^{M} \sum_{i\in I}^{n_I} \sum_{J\in I_{ovl}}^{M_{I_{ovl}}} \sum_{j\in J}^{n_J} \sum_{P=1}^{M} \sum_{p\in P}^{n_P} s_{iI,jJ} \boldsymbol{S'}^{0}_{jJ,pP} \sum_{K=1}^{M} \sum_{k\in K}^{n_K} \left[\left(2J^{kK,kK}_{iI,pP} - K^{kK,kK}_{iI,pP} \right) - \sum_{L\in K_{ovl}}^{M_{K_{ovl}}} \sum_{l\in L}^{n_L} \sum_{R=1}^{M} \sum_{r\in R}^{n_R} s_{kK,lL} \boldsymbol{T}^{(0)}_{lL,rR} \left(2J^{kK,rR}_{iI,pP} - K^{kK,rR}_{iI,pP} \right) \right] \Bigg\} \\
&+ \sum_{I=1}^{M} \sum_{i\in I}^{n_I} \sum_{J\in I_{ovl}}^{M_{I_{ovl}}} \sum_{j\in J}^{n_J} \sum_{P=1}^{M} \sum_{p\in P}^{n_P} s_{iI,jJ} \boldsymbol{S'}^{0}_{jJ,pP} \sum_{K=1}^{M} \sum_{k\in K}^{n_K} \sum_{L\in K_{ovl}}^{M_{K_{ovl}}} \sum_{l=L}^{n_L} \sum_{R=1}^{M} \sum_{r\in R}^{n_R} s_{kK,lL} \boldsymbol{S'}^{0}_{lL,rR} \left(2J^{kK,rR}_{iI,pP} - K^{kK,rR}_{iI,pP} \right).
\end{aligned}
\tag{A17}
$$

As $\boldsymbol{T}^{(0)}_{aA,bB}$ was previously defined as zero for all aA, bB pairs in Eq. (31), terms that include this factor vanish. Additionally, all instances of the zeroth power of the $\boldsymbol{S'}$ matrix may be replaced by the identity matrix. This simplifies Eq. (A17) to the form

$$
\begin{aligned}
\delta E^{(1)}_{ex-rep} &= -2 \sum_{I=1}^{M} \sum_{i\in I}^{n_I} \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}} \sum_{j\in J}^{n_J} \sum_{P=1}^{M} \sum_{p\in P}^{n_P} s_{iI,jJ} \boldsymbol{I}_{jJ,pP} h_{iI,pP} - \sum_{K=1}^{M} \sum_{k\in K}^{n_K} \sum_{K=1}^{M} \sum_{k\in K}^{n_K} \sum_{L\in K_{ovl}}^{M_{K_{ovl}}} \sum_{l=L}^{n_L} \sum_{R=1}^{M} \sum_{r\in R}^{n_R} s_{kK,lL} \boldsymbol{I}_{lL,rR} \left(2J^{kK,rR}_{iI,iI} - K^{kK,rR}_{iI,iI} \right) \\
&- \sum_{I=1}^{M} \sum_{i\in I}^{n_I} \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}} \sum_{j\in J}^{n_J} \sum_{P=1}^{M} \sum_{p\in P}^{n_P} s_{iI,jJ} \boldsymbol{I}_{jJ,pP} \sum_{K=1}^{M} \sum_{k\in K}^{n_K} \left(2J^{kK,kK}_{iI,pP} - K^{kK,rR}_{iI,pP} \right) + \sum_{I=1}^{M} \sum_{i\in I}^{n_I} \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}} \sum_{j\in J}^{n_J} \sum_{P=1}^{M} \sum_{p\in P}^{n_P} s_{iI,jJ} \boldsymbol{I}_{jJ,pP} \\
&\times \sum_{K=1}^{M} \sum_{k\in K}^{n_K} \sum_{L\in K_{ovl}}^{M_{K_{ovl}}} \sum_{l\in L}^{n_L} \sum_{R=1}^{M} \sum_{r\in R}^{n_R} s_{kK,lL} \boldsymbol{I}_{lL,rR} \left(2J^{kK,rR}_{iI,pP} - K^{kK,rR}_{iI,pP} \right).
\end{aligned}
\tag{A18}
$$

Using the properties of the identity matrix, Eq. (A18) may be simplified to the final expression for $\delta E^{(1)}_{ex-rep}$ by eliminating all terms for which $J \neq P$ or $L \neq R$ and thus replacing all instances of the indices P or R with J or L, respectively,

$$
\begin{aligned}
\delta E^{(1)}_{ex-rep} &= -2 \sum_{I=1}^{M} \sum_{i\in I}^{n_I} \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}} \sum_{j\in J}^{n_J} s_{iI,jJ} h_{iI,jJ} - \sum_{I=1}^{M} \sum_{i\in I}^{n_I} \sum_{K=1}^{M} \sum_{k\in K}^{n_K} \sum_{L\in K_{ovl}}^{M_{K_{ovl}}} \sum_{l\in L}^{n_L} s_{kK,lL} \left(2J^{kK,lL}_{iI,iI} - K^{kK,lL}_{iI,iI} \right) \\
&- \sum_{I=1}^{M} \sum_{i\in I}^{n_I} \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}} \sum_{j\in J}^{n_J} s_{iI,jJ} \sum_{K=1}^{M} \sum_{k\in K}^{n_K} \left(2J^{kK,kK}_{iI,jJ} - K^{kK,kK}_{iI,jJ} \right) + \sum_{I=1}^{M} \sum_{i\in I}^{n_I} \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}} \sum_{j\in J}^{n_J} s_{iI,jJ} \\
&\times \sum_{I=1}^{M} \sum_{k\in K}^{n_K} \sum_{L\in K_{ovl}}^{M_{K_{ovl}}} \sum_{l\in L}^{n_L} s_{kK,lL} \left(2J^{kK,lL}_{iI,jJ} - K^{kK,lL}_{iI,iI} \right).
\end{aligned}
\tag{A19}
$$

Published under an exclusive license by AIP Publishing

4. Derivation of the unique terms in the second-order exchange-repulsion energy

Unlike the expression for the first-order exchange-repulsion energy, the expression for the unique terms in the second-order exchange-repulsion energy, $\delta E^{(2)}_{ex-rep}$, will not involve vanishing $\boldsymbol{T}^{(n)}$ matrix elements or replacement of $\boldsymbol{S}'$ matrix elements with those of the identity matrix. As such, it will be necessary to apply the PNIO approximation to new pairs of fragment indices. The general expression for $\delta E^{(2)}_{ex-rep}$ is given by

$$\begin{aligned}
\delta E^{(2)}_{ex-rep} = (-1)\Bigg\{ & -2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S}'^{1}_{jJ,pP}h_{iI,pP} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in \boldsymbol{K}_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{S}'^{1}_{lL,rR} \\
& \times\Bigg[\left(2J^{kK,rR}_{iI,iI}-K^{kK,rR}_{iI,iI}\right) - \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{T}^{(1)}_{jJ,pP}\left(2J^{kK,rR}_{iI,pP}-K^{kK,rR}_{iI,pP}\right)\Bigg] - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S}'^{1}_{jJ,pP} \\
& \times\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\Bigg[\left(2J^{kK,kK}_{iI,pP}-K^{kK,kK}_{iI,pP}\right) - \sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{T}^{(1)}_{lL,rR}\left(2J^{kK,rR}_{iI,pP}-K^{kK,rR}_{iI,pP}\right)\Bigg]\Bigg\} \\
& + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S}'^{1}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{S}'^{1}_{lL,rR}\left(2J^{kK,rR}_{iI,pP}-K^{kK,rR}_{iI,pP}\right).
\end{aligned} \tag{A20}$$

The instances of $\boldsymbol{T}^{(1)}_{aA,bB}$ may be rewritten as

$$\begin{aligned}
\boldsymbol{T}^{(1)}_{aA,bB} &= (-1)^0(S\prime)^0_{aA,bB} \\
&= \boldsymbol{I}_{aA,bB} \\
&= \delta_{aA,bB}.
\end{aligned} \tag{A21}$$

Making this substitution in Eq. (A20), as well as multiplying all terms in braces by the negative factor, and replacing all instances of $\boldsymbol{S}'^{1}_{aA,bB}$ with the corresponding MO overlap integrals simplify the expression to

$$\begin{aligned}
\delta E^{(2)}_{ex-rep} = & 2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}s_{jJ,pP}h_{iI,pP} + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}s_{lL,rR} \\
& \times\Bigg[\left(2J^{kK,rR}_{iI,iI}-K^{kK,rR}_{iI,iI}\right) - \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\delta_{jJ,pP}\left(2J^{kK,rR}_{iI,pP}-K^{kK,rR}_{iI,pP}\right)\Bigg] + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}s_{jJ,pP} \\
& \times\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\Bigg[\left(2J^{kK,kK}_{iI,pP}-K^{kK,kK}_{iI,pP}\right) - \sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\delta_{lL,rR}\left(2J^{kK,rR}_{iI,pP}-K^{kK,rR}_{iI,pP}\right)\Bigg] \\
& + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}s_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}s_{lL,rR}\left(2J^{kK,rR}_{iI,pP}-K^{kK,rR}_{iI,pP}\right).
\end{aligned} \tag{A22}$$

First, the terms involving Kronecker delta factors are again simplified by eliminating all pairs for which $J \neq P$ or $L \neq R$,

$$\begin{aligned}
\delta E^{(2)}_{ex-rep} = & 2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}s_{jJ,pP}h_{iI,pP} + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}s_{lL,rR} \\
& \times\Bigg[\left(2J^{kK,rR}_{iI,iI}-K^{kK,rR}_{iI,iI}\right) - \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\left(2J^{kK,rR}_{iI,pP}-K^{kK,rR}_{iI,pP}\right)\Bigg] + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}s_{jJ,pP} \\
& \times\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\Bigg[\left(2J^{kK,kK}_{iI,pP}-K^{kK,kK}_{iI,pP}\right) - \sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\left(2J^{kK,lL}_{iI,pP}-K^{kK,lL}_{iI,pP}\right)\Bigg] + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}s_{jJ,pP} \\
& \times\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}s_{lL,rR}\left(2J^{kK,rR}_{iI,pP}-K^{kK,rR}_{iI,pP}\right).
\end{aligned} \tag{A23}$$

Published under an exclusive license by AIP Publishing

Now, the PNIO approximation is applied to the new J, P and L, R fragment index pairs in all of the summations in Eq. (A23). Doing so produces the final expression for $\delta E^{(2)}_{ex-rep}$,

$$\begin{aligned}\delta E^{(2)}_{ex-rep} = {} & 2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P\in J_{ovl}}^{M_{J_{ovl}}}\sum_{p\in P}^{n_P} s_{iI,jJ}s_{jJ,pP}h_{iI,pP} + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L}\sum_{R\in L_{ovl}}^{M_{L_{ovl}}}\sum_{r\in R}^{n_R} s_{kK,lL}s_{lL,rR} \\ & \times\left[\left(2J^{kK,rR}_{iI,iI} - K^{kK,rR}_{iI,iI}\right) - \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\left(2J^{kK,rR}_{iI,jJ} - K^{kK,rR}_{iI,jJ}\right)\right] + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P\in J_{ovl}}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}s_{jJ,pP} \\ & \times\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left[\left(2J^{kK,kK}_{iI,pP} - K^{kK,kK}_{iI,pP}\right) - \sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\left(2J^{kK,lL}_{iI,pP} - K^{kK,lL}_{iI,pP}\right)\right] + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P\in J_{ovl}}^{M_{J_{ovl}}}\sum_{p\in P}^{n_P} s_{iI,jJ}s_{jJ,pP} \\ & \times\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L}\sum_{R\in L_{ovl}}^{M_{L_{ovl}}}\sum_{r\in R}^{n_R} s_{kK,lL}s_{lL,rR}\left(2J^{kK,rR}_{iI,pP} - K^{kK,rR}_{iI,pP}\right). \end{aligned} \tag{A24}$$

5. Derivation of the unique terms in the third-order exchange-repulsion energy

As in the derivation of the second-order exchange-repulsion energy, the expression for the unique terms in the third-order exchange-repulsion energy, $\delta E^{(3)}_{ex-rep}$, will also involve non-vanishing $\boldsymbol{T}^{(n)}$ matrix elements. As such, it will again be necessary to apply the PNIO approximation to new pairs of fragment indices. The general expression for $\delta E^{(3)}_{ex-rep}$ is given by

$$\begin{aligned}\delta E^{(3)}_{ex-rep} = {} & (-1)^2\Bigg\{-2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S'}^2_{jJ,pP}h_{iI,pP} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{S'}^2_{lL,rR} \\ & \times\left[\left(2J^{kK,rR}_{iI,iI} - K^{kK,rR}_{iI,iI}\right) - \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{T}^{(2)}_{jJ,pP}\left(2J^{kK,rR}_{iI,pP} - K^{kK,rR}_{iI,pP}\right)\right] - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S'}^2_{jJ,pP} \\ & \times\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left[\left(2J^{kK,kK}_{iI,pP} - K^{kK,kK}_{iI,pP}\right) - \sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{T}^{(2)}_{lL,rR}\left(2J^{kK,rR}_{iI,pP} - K^{kK,rR}_{iI,pP}\right)\right]\Bigg\} \\ & + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S'}^2_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{S'}^2_{lL,rR}\left(2J^{kK,rR}_{iI,pP} - K^{kK,rR}_{iI,pP}\right). \end{aligned} \tag{A25}$$

The instances of $\boldsymbol{T}^{(2)}_{aA,bB}$ may be rewritten as

$$\begin{aligned}\boldsymbol{T}^{(2)}_{aA,bB} &= (-1)^0(S\prime)^0_{aA,bB} + (-1)^1(S\prime)^1_{aA,bB} \\ &= \boldsymbol{I}_{aA,bB} - S\prime_{aA,bB} \\ &= \delta_{aA,bB} - s_{aA,bB}. \end{aligned} \tag{A26}$$

Making this substitution in Eq. (A25) produces the expression

$$\begin{aligned}\delta E^{(3)}_{ex-rep} = {} & -2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S'}^2_{jJ,pP}h_{iI,pP} + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{S'}^2_{lL,rR} \\ & \times\left[\left(2J^{kK,rR}_{iI,iI} - K^{kK,rR}_{iI,iI}\right) - \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}(\delta_{jJ,pP} - s_{jJ,pP})\left(2J^{kK,rR}_{iI,pP} - K^{kK,rR}_{iI,pP}\right)\right] + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S'}^2_{jJ,pP} \\ & \times\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left[\left(2J^{kK,kK}_{iI,pP} - K^{kK,kK}_{iI,pP}\right) - \sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}(\delta_{lL,rR} - s_{lL,rR})\left(2J^{kK,rR}_{iI,pP} - K^{kK,rR}_{iI,pP}\right)\right] \\ & + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J}\sum_{P=1}^{M}\sum_{p\in P}^{n_P} s_{iI,jJ}\boldsymbol{S'}^2_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L}\sum_{R=1}^{M}\sum_{r\in R}^{n_R} s_{kK,lL}\boldsymbol{S'}^2_{lL,rR}\left(2J^{kK,rR}_{iI,pP} - K^{kK,rR}_{iI,pP}\right). \end{aligned} \tag{A27}$$

Simplifying Eq. (A27) by limiting summations over Kronecker delta and MO overlap integral elements gives

Published under an exclusive license by AIP Publishing

$$\delta E^{(3)}_{ex-rep} = -2\sum_{I=1}^{M}\sum_{i\epsilon I}^{n_I}\sum_{J\epsilon \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\epsilon J}^{n_J}\sum_{P=1}^{M}\sum_{p\epsilon P}^{n_P} s_{iI,jJ}\boldsymbol{S}'^{2}_{jJ,pP}h_{iI,pP} - \sum_{I=1}^{M}\sum_{i\epsilon I}^{n_I}\sum_{K=1}^{M}\sum_{k\epsilon K}^{n_K}\sum_{L\epsilon K_{ovl}}^{M_{K_{ovl}}}\sum_{l\epsilon L}^{n_L}\sum_{R=1}^{M}\sum_{r\epsilon R}^{n_R} s_{kK,lL}\boldsymbol{S}'^{2}_{lL,rR}$$
$$\times\left\{\left(2J^{kK,rR}_{iI,iI} - K^{kK,rR}_{iI,iI}\right) - \sum_{J\epsilon \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\epsilon J}^{n_J} s_{iI,jJ}\left[\left(2J^{kK,rR}_{iI,jJ} - K^{kK,rR}_{iI,jJ}\right) - \sum_{P=1}^{M}\sum_{p\epsilon P}^{n_P} s_{iJ,pP}\left(2J^{kK,rR}_{iI,pP} - K^{kK,rR}_{iI,pP}\right)\right]\right\}$$
$$- \sum_{I=1}^{M}\sum_{i\epsilon I}^{n_I}\sum_{J\epsilon \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\epsilon J}^{n_J}\sum_{P=1}^{M}\sum_{p\epsilon P}^{n_P} s_{iI,jJ}\boldsymbol{S}^{2}_{jJ,pP}\sum_{K=1}^{M}\sum_{k\epsilon K}^{n_K}\left\{\left(2J^{kK,kK}_{iI,pP} - K^{kK,kK}_{iI,pP}\right) - \sum_{L\epsilon K_{ovl}}^{M_{K_{ovl}}}\sum_{l\epsilon L}^{n_L} s_{kK,lL}\right.$$
$$\left.\times\left[\left(2J^{kK,rR}_{iI,pP} - K^{kK,rR}_{iI,pP}\right) - \sum_{R=1}^{M}\sum_{r\epsilon R}^{n_R} s_{lL,rR}\left(2J^{kK,rR}_{iI,pP} - K^{kK,rR}_{iI,pP}\right)\right]\right\} + \sum_{I=1}^{M}\sum_{i\epsilon I}^{n_I}\sum_{J\epsilon \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\epsilon J}^{n_J}\sum_{P=1}^{M}\sum_{p\epsilon P}^{n_P} s_{iI,jJ}\boldsymbol{S}'^{2}_{jJ,pP}$$
$$\times\sum_{K=1}^{M}\sum_{k\epsilon K}^{n_K}\sum_{L\epsilon K_{ovl}}^{M_{K_{ovl}}}\sum_{l\epsilon L}^{n_L}\sum_{R=1}^{M}\sum_{r\epsilon R}^{n_R} s_{kK,lL}\boldsymbol{S}'^{2}_{lL,rR}\left(2J^{kK,rR}_{iI,pP} - K^{kK,rR}_{iI,pP}\right). \tag{A28}$$

Now, the $\boldsymbol{S}'^{2}_{aA,bB}$ matrices are expanded. In order to maintain consistency with the previous exchange-repulsion expressions, the fact that fragment indices in the $\boldsymbol{S}'^{2}_{jJ,pP}$ matrix expansion are arbitrary will be used to rearrange fragment indices. Specifically, the expansion of the $\boldsymbol{S}'^{2}_{jJ,pP}$ and $\boldsymbol{S}'^{2}_{lL,rR}$ will be rewritten in the form

$$\boldsymbol{S}'^{2}_{jJ,pP} = \sum_{Q=1}^{M}\sum_{q\epsilon Q}^{n_Q}\boldsymbol{S}'_{jJ,qQ}\sum_{P=1}^{M}\sum_{p\epsilon P}^{n_P}\boldsymbol{S}'_{qQ,pP} = \sum_{P=1}^{M}\sum_{p\epsilon P}^{n_P}\boldsymbol{S}'_{jJ,pP}\sum_{Q=1}^{M}\sum_{q\epsilon Q}^{n_Q}\boldsymbol{S}'_{pP,qQ} \tag{A29}$$

and

$$\boldsymbol{S}'^{2}_{lL,rR} = \sum_{S=1}^{M}\sum_{s\epsilon S}^{n_s}\boldsymbol{S}'_{lL,sS}\sum_{R=1}^{M}\sum_{r\epsilon R}^{n_R}\boldsymbol{S}'_{sS,rR} = \sum_{R=1}^{M}\sum_{r\epsilon R}^{n_R}\boldsymbol{S}'_{lL,rR}\sum_{S=1}^{M}\sum_{s\epsilon S}^{n_S}\boldsymbol{S}'_{rR,sS}. \tag{A30}$$

Using the expressions given in Eqs. (A29) and (A30), the expression in Eq. (A28) may be rewritten in the expanded form

$$\delta E^{(3)}_{ex-rep} = -2\sum_{I=1}^{M}\sum_{i\epsilon I}^{n_I}\sum_{J\epsilon \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\epsilon J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\epsilon P}^{n_P} s_{jJ,pP}\sum_{Q=1}^{M}\sum_{q\epsilon Q}^{n_Q} s_{pP,qQ}h_{iI,qQ} - \sum_{I=1}^{M}\sum_{i\epsilon I}^{n_I}\sum_{K=1}^{M}\sum_{k\epsilon K}^{n_K}\sum_{L\epsilon K_{ovl}}^{M_{K_{ovl}}}\sum_{l\epsilon L}^{n_L} s_{kK,lL}\sum_{R=1}^{M}\sum_{r\epsilon R}^{n_R} s_{lL,rR}$$
$$\times\sum_{S=1}^{M}\sum_{s\epsilon S}^{n_s} s_{rR,sS}\left\{\left(2J^{kK,sS}_{iI,iI} - K^{kK,rR}_{iI,iI}\right) - \sum_{J\epsilon \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\epsilon J}^{n_J} s_{iI,jJ}\left[\left(2J^{kK,sS}_{iI,jJ} - K^{kK,sS}_{iI,jJ}\right) - \sum_{P=1}^{M}\sum_{p\epsilon P}^{n_P} s_{jJ,pP}\left(2J^{kK,sS}_{iI,pP} - K^{kK,sS}_{iI,pP}\right)\right]\right\}$$
$$- \sum_{I=1}^{M}\sum_{i\epsilon I}^{n_I}\sum_{J\epsilon \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\epsilon J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\epsilon P}^{n_P} s_{jJ,pP}\sum_{Q=1}^{M}\sum_{q\epsilon Q}^{n_Q} s_{pP,qQ}\sum_{K=1}^{M}\sum_{k\epsilon K}^{n_K}\left\{\left(2J^{kK,kK}_{iI,qQ} - K^{kK,kK}_{iI,qQ}\right) - \sum_{L\epsilon K_{ovl}}^{M_{K_{ovl}}}\sum_{l\epsilon L}^{n_L} s_{kK,lL}\right.$$
$$\left.\times\left[\left(2J^{kK,rR}_{iI,qQ} - K^{kK,rR}_{iI,qQ}\right) - \sum_{R=1}^{M}\sum_{r\epsilon R}^{n_R} s_{lL,rR}\left(2J^{kK,rR}_{iI,qQ} - K^{kK,rR}_{iI,qQ}\right)\right]\right\} + \sum_{I=1}^{M}\sum_{i\epsilon I}^{n_I}\sum_{J\epsilon \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\epsilon J}^{n_J} s_{iI,jJ}\sum_{P=1}^{M}\sum_{p\epsilon P}^{n_P} s_{jJ,pP}$$
$$\times\sum_{Q=1}^{M}\sum_{q\epsilon Q}^{n_Q} s_{jJ,pP}\sum_{K=1}^{M}\sum_{k\epsilon K}^{n_K}\sum_{L\epsilon K_{ovl}}^{M_{K_{ovl}}}\sum_{l\epsilon L}^{n_L} s_{kK,lL}\sum_{R=1}^{M}\sum_{r\epsilon R}^{n_R} s_{lL,rR}\sum_{S=1}^{M}\sum_{s\epsilon S}^{n_s} s_{rR,sS}\left(2J^{kK,sS}_{iI,qQ} - K^{kK,sS}_{iI,qQ}\right). \tag{A31}$$

As was previously done in the derivation of the $\delta E^{(2)}_{ex-rep}$ expression, the PNIO approximation is first applied to the J, P and L, R fragment index pairs. In addition, the PNIO approximation is also applied to the new fragment index pairs, P, Q and R, S. Thus, the final expression for $\delta E^{(3)}_{ex-rep}$, with each of the summations properly restricted to overlapping fragment pairs, is given by

Published under an exclusive license by AIP Publishing

$$\delta E^{(3)}_{ex-rep} = -2\sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P\in J_{ovl}}^{M_{J_{ovl}}}\sum_{p\in P}^{n_P} s_{jJ,pP}\sum_{Q\in P_{ovl}}^{M_{P_{ovl}}}\sum_{q\in Q}^{n_Q} s_{pP,qQ}h_{iI,qQ} - \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R\in L_{ovl}}^{M_{L_{ovl}}}\sum_{r\in R}^{n_R} s_{iL,rR}$$
$$\times\sum_{S\in R_{ovl}}^{M_{R_{ovl}}}\sum_{s\in S}^{n_s} s_{rR,sS}\left\{\left(2J^{kK,sS}_{iI,iI} - K^{kK,rR}_{iI,iI}\right) - \sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\left[\left(2J^{kK,sS}_{iI,jJ} - K^{kK,sS}_{iI,jJ}\right) - \sum_{P\in J_{ovl}}^{M_{J_{ovl}}}\sum_{p\in P}^{n_P} s_{jJ,pP}\left(2J^{kK,sS}_{iI,pP} - K^{kK,sS}_{iI,pP}\right)\right]\right\}$$
$$- \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P\in J_{ovl}}^{M_{J_{ovl}}}\sum_{p\in P}^{n_P} s_{jJ,pP}\sum_{Q\in P_{ovl}}^{M_{P_{ovl}}}\sum_{q\in Q}^{n_Q} s_{pP,qQ}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\left\{\left(2J^{kK,kK}_{iI,qQ} - K^{kK,kK}_{iI,qQ}\right) - \sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\right.$$
$$\left.\times\left[\left(2J^{kK,rR}_{iI,qQ} - K^{kK,rR}_{iI,qQ}\right) - \sum_{R\in L_{ovl}}^{M_{L_{ovl}}}\sum_{r\in R}^{n_R} s_{lL,rR}\left(2J^{kK,rR}_{iI,qQ} - K^{kK,rR}_{iI,qQ}\right)\right]\right\} + \sum_{I=1}^{M}\sum_{i\in I}^{n_I}\sum_{J\in \boldsymbol{I}_{ovl}}^{M_{I_{ovl}}}\sum_{j\in J}^{n_J} s_{iI,jJ}\sum_{P\in J_{ovl}}^{M_{J_{ovl}}}\sum_{p\in P}^{n_P} s_{jJ,pP}$$
$$\times\sum_{Q\in P_{ovl}}^{M_{P_{ovl}}}\sum_{q\in Q}^{n_Q} s_{jJ,pP}\sum_{K=1}^{M}\sum_{k\in K}^{n_K}\sum_{L\in K_{ovl}}^{M_{K_{ovl}}}\sum_{l\in L}^{n_L} s_{kK,lL}\sum_{R\in L_{ovl}}^{M_{L_{ovl}}}\sum_{r\in R}^{n_R} s_{lL,rR}\sum_{S=1}^{M}\sum_{s\in S}^{n_s} s_{rR,sS}\left(2J^{kK,sS}_{iI,qQ} - K^{kK,sS}_{iI,qQ}\right). \quad \text{(A32)}$$

REFERENCES

[1] M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, "Fragmentation methods: A route to accurate calculations on large systems," Chem. Rev. **112**, 632–672 (2012).

[2] S. R. Pruitt, C. Bertoni, K. R. Brorsen, and M. S. Gordon, "Efficient and accurate fragmentation methods," Acc. Chem. Res. **47**, 2786–2794 (2014).

[3] J. M. Herbert, "Fantasy versus reality in fragment-based quantum chemistry," J. Chem. Phys. **151**, 170901 (2019).

[4] M. A. Collins and R. P. A. Bettens, "Energy-based molecular fragmentation methods," Chem. Rev. **115**, 5607–5642 (2015).

[5] K. Raghavachari and A. Saha, "Accurate composite and fragment-based quantum chemical models for large molecules," Chem. Rev. **115**, 5643–5677 (2015).

[6] A. V. Akimov and O. V. Prezhdo, "Large-scale computations in chemistry: A bird's eye view of a vibrant field," Chem. Rev. **115**, 5797–5890 (2015).

[7] *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*, edited by M. S. Gordon (Wiley, Hoboken, NJ, 2017).

[8] P.-O. Löwdin, "Quantum theory of many-particle systems. I. Physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configuration interaction," Phys. Rev. **97**, 1474 (1955).

[9] W. H. Adams, "On the solution of the Hartree-Fock equations in terms of localized orbitals," J. Chem. Phys. **34**, 89 (1961).

[10] W. H. Adams, "Orbital theories of electronic structure," J. Chem. Phys. **37**, 2009 (1962).

[11] W. H. Adams, "Orbital theories of electronic structure. II. Molecularly invariant orbitals," J. Chem. Phys. **42**, 4030 (1965).

[12] T. L. Gilbert, in *Molecular Orbitals in Chemistry, Physics, and Biology*, edited by P. O. Löwdin and B. Pullman (Academic Press, New York, 1964).

[13] O. Matsuoka, "Expansion methods for Adams–Gilbert equations. I. Modified Adams–Gilbert equation and common and fluctuating basis sets," J. Chem. Phys. **66**, 1245 (1977).

[14] Y. Watanabe and O. Matsuoka, "Nonorthogonal molecular orbital method: Single-determinant theory," J. Chem. Phys. **140**, 204111 (2014).

[15] H. Stoll and H. Preuss, "On the direct calculation of localized HF orbitals in molecule clusters, layers and solids," Theor. Chim. Acta **46**, 11–21 (1977).

[16] H. Stoll, G. Wagenblast, and H. Preuss, "The Hartree-Fock theory of local regions in molecules," J. Am. Chem. Soc. **100**(24), 7742–7743 (1978).

[17] H. Stoll, G. Wagenblast, and H. Preuβ, "On the use of local basis sets for localized molecular orbitals," Theor. Chim. Acta **57**, 169–178 (1980).

[18] J. M. Cullen, "An examination of the effects of basis set and charge transfer in hydrogen-bonded dimers with a constrained Hartree–Fock method," Int. J. Quantum Chem. **40**, 193–207 (1991).

[19] E. Gianinetti, M. Raimondi, and E. Tornaghi, "Modification of the Roothaan equations to exclude BSSE from molecular interaction calculations," Int. J. Quantum Chem. **60**, 157–166 (1996).

[20] T. Nagata, O. Takahashi, K. Saito, and S. Iwata, "Basis set superposition error free self-consistent field method for molecular interaction in multi-component systems: Projection operator formalism," J. Chem. Phys. **115**, 3553–3560 (2001).

[21] R. Z. Khaliullin, M. Head-Gordon, and A. T. Bell, "An efficient self-consistent field method for large systems of weakly interacting components," J. Chem. Phys. **124**, 204105 (2006).

[22] A. Cembran, P. Bao, Y. Wang, L. Song, D. G. Truhlar, and J. Gao, "On the interfragment exchange in the X-Pol method," J. Chem. Theory Comput. **6**(8), 2469–2476 (2010).

[23] D. G. Fedorov and K. Kitaura, "Pair interaction energy decomposition analysis," J. Comput. Chem. **28**, 222–237 (2007).

[24] P. Pulay, "Localizability of dynamic electron correlation," Chem. Phys. Lett. **100**(2), 151–154 (1983).

[25] S. Saebo and P. Pulay, "Local configuration interaction: An efficient approach for larger molecules," Chem. Phys. Lett. **113**(1), 13–18 (1985).

[26] S. Saebo and P. Pulay, "Local treatment of electron correlation," Annu. Rev. Phys. Chem. **44**, 213–236 (1993).

[27] C. Hampel and H. J. Werner, "Local treatment of electron correlation in coupled cluster theory," J. Chem. Phys. **104**, 6286 (1996).

[28] J. Yang, G. K.-L. Chan, F. R. Manby, M. Schütz, and H.-J. Werner, "The orbital-specific-virtual local coupled cluster singles and doubles method," J. Chem. Phys. **136**, 144105 (2012).

[29] M. Schwilk, D. Usvyat, and H.-J. Werner, "Communication: Improved pair approximations in local coupled-cluster methods," J. Chem. Phys. **142**, 121102 (2015).

[30] Y. Kurashige, J. Yang, G. K.-L. Chan, and F. R. Manby, "Optimization of orbital-specific virtuals in local Møller-Plesset perturbation theory," J. Chem. Phys. **136**, 124106 (2012).

[31] P. E. Maslen and M. Head-Gordon, "Non-iterative local second order Møller–Plesset theory," Chem. Phys. Lett. **283**, 102–108 (1998).

[32] C. Riplinger and F. Neese, "An efficient and near linear scaling pair natural orbital based local coupled cluster method," J. Chem. Phys. **138**, 034106 (2013).

[33] C. Riplinger, B. Sandhoefer, A. Hansen, and F. Neese, "Natural triple excitations in local coupled cluster calculations with pair natural orbitals," J. Chem. Phys. **139**, 134101 (2013).

[34] S. Li, J. Shen, W. Li, and Y. Jiang, "An efficient implementation of the 'cluster-in-molecule' approach for local electron correlation calculations," J. Chem. Phys. **125**, 074109 (2006).

[35] W. Li and P. Piecuch, "Improved design of orbital domains within the cluster-in-molecule local correlation framework: Single-environment cluster-in-molecule ansatz and its application to local coupled-cluster approach with singles and doubles," J. Phys. Chem. A **114**, 8644–8657 (2010).

[36] M. Ziolkowski, B. Jansik, T. Kjaergaard, and P. Jorgensen, "Linear scaling coupled cluster method with correlation energy based error control," J. Chem. Phys. **133**, 014107 (2010).

Published under an exclusive license by AIP Publishing

[37]K. Kristensen, M. Ziółkowski, B. Jansík, T. Kjærgaard, and P. Jørgensen, "A locality analysis of the divide-expand-consolidate coupled cluster amplitude equations," J. Chem. Theory Comput. **7**, 1677–1694 (2011).
[38]T. Kjaergaard, P. Baudin, D. Bykov, K. Kristensen, and P. Jorgensen, "The divide–expand–consolidate coupled cluster scheme," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **7**(6), e1319 (2017).
[39]Y. Jin, N. Q. Su, X. Xu, and H. Hu, "Self-consistent field for fragmented quantum mechanical model of large molecular systems," J. Comput. Chem. **37**, 321–326 (2016).
[40]K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, "Fragment molecular orbital method: An approximate computational method for large molecules," Chem. Phys. Lett. **313**, 701–706 (1999).
[41]J. Gao, "Toward a molecular orbital derived empirical potential for liquid simulations," J. Phys. Chem. B **101**, 657–663 (1997).
[42]T. Shimazaki, K. Kitaura, D. G. Fedorov, and T. Nakajima, "Group molecular orbital approach to solve the Huzinaga subsystem self-consistent-field equations," J. Chem. Phys. **146**, 084109 (2017).
[43]E. E. Dahlke and D. G. Truhlar, "Electrostatically embedded many-body expansion for large systems, with applications to water clusters," J. Chem. Theory Comput. **3**, 46–53 (2007).
[44]D. G. Fedorov, "The fragment molecular orbital method: Theoretical development, implementation in GAMESS, and applications," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **7**, 1322 (2017).
[45]L. V. Slipchenko and M. S. Gordon, "Damping functions in the effective fragment potential method," Mol. Phys. **107**, 999–1016 (2009).
[46]S. Huzinaga and A. A. Cantu, "Theory of separability of many-electron systems," J. Chem. Phys. **55**, 5543–5549 (1971).
[47]L. Seijo and Z. Barandiarán, "Self-consistent embedded clusters: Building block equations for localized orthogonal orbitals," J. Math. Chem. **10**, 41–56 (1992).
[48]E. Francisco, A. Martín Pendás, and W. H. Adams, "Generalized Huzinaga building-block equations for nonorthogonal electronic groups: Relation to the Adams–Gilbert theory," J. Chem. Phys. **97**, 6504–6508 (1992).
[49]L. Seijo and Z. Barandiarán, "Parallel, linear-scaling building-block and embedding method based on localized orbitals and orbital-specific basis sets," J. Chem. Phys. **121**(14), 6698–6709 (2004).
[50]G. G. Ferenczy and W. H. Adams, "Optimization of selected molecular orbitals in group basis sets," J. Chem. Phys. **130**(13), 134108 (2009).
[51]C. Steinmann, D. G. Fedorov, and J. H. Jensen, "Effective fragment molecular orbital method: A merger of the effective fragment potential and fragment molecular orbital methods," J. Phys. Chem. A **114**(33), 8705–8712 (2010).
[52]S. R. Pruitt, C. Steinmann, J. H. Jensen, and M. S. Gordon, "Fully integrated effective fragment molecular orbital method," J. Chem. Theory Comput. **9**(5), 2235–2249 (2013).
[53]L. D. Jacobson and J. M. Herbert, "An efficient, fragment-based electronic structure method for molecular systems: Self-consistent polarization with perturbative two-body exchange and dispersion," J. Chem. Phys. **134**, 094118 (2011).

Published under an exclusive license by AIP Publishing