Dissertation zur Erlangung des Doktorgrades
der Fakultät Chemie und Pharmazie
der Ludwig–Maximilians–Universität München

Integral Bounds and Rigorous Screening Algorithms for Reduced Scaling in Explicitly Correlated, Semi-Numerical, and Non-Hermitian Quantum Chemistry

Travis Hunter Thompson

aus

Shreveport LA, USA

2020

Dissertation zur Erlangung des Doktorgrades
der Fakultät Chemie und Pharmazie
der Ludwig–Maximilians–Universität München

Integral Bounds and Rigorous Screening Algorithms for Reduced Scaling in Explicitly Correlated, Semi-Numerical, and Non-Hermitian Quantum Chemistry

Travis Hunter Thompson

aus

Shreveport LA, USA

2020

Erklärung
Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Christian Ochsenfeld betreut.

Eidstattliche Versicherung
Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 15.06.2020 _____

(Travis H. Thompson)

Dissertation eingereicht am 29.04.2020

1. Gutachter: Prof. Dr. Christian Ochsenfeld

2. Gutachterin: Prof. Dr. Regina de Vivie-Riedle

Mündlichen Prüfung am 27.05.2020

Acknowledgments

I would like to first thank Prof. Dr. Christian Ochsenfeld for giving me the opportunity to conduct research and obtain my doctoral thesis in his working group and for the high level of freedom and support that I enjoyed while carrying out my work. I would also like to thank Prof. Dr. Regina de Vivie-Riedle for the preparation of the second assessment of the thesis.

I thank my colleagues Dr. Thomas Jagau, Henryk Laqua, and Dr. Jörg Kussmann for their parts in fruitful academic collaborations. Furthermore, I would like to thank Dr. Matthias Beuerle, Dr. Sigurd Vogler, Dr. Arne Lünser, Dr. Asbjörn Burow, and many other past and present members of the Ochsenfeld working group for productive academic discussions and for the organization of various social activities. Thanks as well to those who read and offered valuable suggestions for improving this manuscript.

I would like to especially thank my parents Mary and John, my siblings Wesley and Rebecca, my aunt Betsy and uncle Peter, and other family members for their support, guidance, and love during my doctoral work and throughout my life. Lastly, I would like to thank all of those who have supported me through friendship and financial help along the way.

List of Publications

This is a cumulative dissertation comprising four articles in peer-reviewed journals (**I-IV**). These articles are listed here together with the author's contributions to them.

I **T. H. Thompson** and C. Ochsenfeld,
"Distance-including rigorous upper bounds and tight estimates for two-electron integrals over long- and short-range operators",
J. Chem. Phys. **147**, 144101 (2017),
Contributions by the author: *All of the derivation, all of the implementation, all calculations and writing of the manuscript.*

II **T. H. Thompson** and C. Ochsenfeld,
"Integral partition bounds for fast and effective screening of general one-, two-, and many-electron integrals",
J. Chem. Phys. **150**, 044101 (2019),
Contributions by the author: *All of the derivation, all of the implementation, all calculations and writing of the manuscript.*

III **T. H. Thompson**, C. Ochsenfeld and T.-C. Jagau,
"A Schwarz inequality for complex basis function methods in non-Hermitian quantum chemistry",
J. Chem. Phys. **151**, 184104 (2019),
Contributions by the author: *All of the derivation, part of the implementation, all calculations and most of the writing of the manuscript.*

IV H. Laqua, **T. H. Thompson**, J. Kussmann, and C. Ochsenfeld,
"Highly efficient, linear scaling semi-numerical exact-exchange method for graphic processing units",
J. Chem. Theor. Comput. **16**, 1456-1468 (2020),
Contributions by the author: *Derivation of integral bounds, implementation of integral bounds, writing of part of the manuscript.*

Abstract

The properties of chemical systems can be determined computationally by solving the physical equations that govern them. This typically requires the calculation of a very large number of molecular integrals that take various forms depending on the approximations used. Most of these integrals are negligible and avoiding the calculation of negligible integrals can increase computational efficiency immensely. In this work, novel upper bounds and screening algorithms are developed for this purpose. These bounds and algorithms are applicable to a wide range of quantum chemical theories and can be used in combination with the state-of-the-art integral approximation methods that are at the heart of today's most efficient numerical implementations.

A Schwarz-type bound that captures the decay of four-center two-electron integrals due to the decreased interaction of distant charge distributions is developed and tested in the context of Hartree-Fock and range-separated density functional theory, and on four-center integrals over short-range correlation factors that arise in explicitly correlated theories.

An integral partitioning procedure is developed which leads to extremely flexible upper bounds, integral partition bounds (IPBs), for molecular integrals over any number of electrons, any number of basis function centers, and various combinations of integral operators. The procedure allows for the inexpensive calculation of rigorous extents for charge distributions within these various contexts. The IPBs are completely separable into two-center factors, which capture all the sources of asymptotic decay. This allows for the formulation of scaling-consistent screening algorithms, even for the three- and four-electron integrals that arise, e.g., within explicitly correlated Møller-Plesset perturbation theory (MP2-F12).

The IPBs are used to increase the efficiency and reliability of semi-numerical techniques for calculation of the exchange matrix in Hartree-Fock and hybrid DFT calculations, where real space numerical quadrature is used to approximate electron repulsion integrals. Similarly, a framework for very efficient MP2-F12 working equations based on optimal combinations of resolution-of-the-indentity, density-fitting, and numerical quadrature approximations are given. They reduce the fifth-order formal scaling of the MP2-F12 method to fourth-order, while drastically reducing the cost of the initially most expensive terms. The resulting equations involve various types of sparse integral tensors which can all be treated using IPBs, and asymptotically linear scaling implementations are possible.

Furthermore, a Schwarz bound is developed and implemented for screening the novel four-center two-electron integrals that arise in the treatment of resonance states using the method of complex basis functions within non-Hermitian quantum chemistry. This is a crucial step for increasing the efficiency and reach of the method.

Contents

Chapter 1

Introduction

The ultimate aim of quantum chemistry is to solve the equations of motion governing chemical systems, allowing for the understanding and prediction of chemical phenomena. Because of the small size of the atomic nuclei and electrons involved, quantitative accuracy can only be achieved by a quantum mechanical treatment, in which the state of the system is described by a wave function Ψ, also known as a state vector. Physical observables are represented by linear operators whose eigenvectors correspond to possible state vectors. The eigenvalue of a particular eigenvector is the value of the observable for that state.

In non-relativistic quantum chemistry, the most important operator is the non-relativistic many-body Hamiltonian $\hat{\mathcal{H}}$, which corresponds to the energy of the system. For N_e electrons and N_n nuclei, $\hat{\mathcal{H}}$ can be written in atomic units as

$$\hat{\mathcal{H}} = -\sum_{A}^{N_\mathrm{n}} \frac{\Delta_A}{2m_A} + \sum_{A}^{N_\mathrm{n}} \sum_{B>A}^{N_\mathrm{n}} \frac{Z_A Z_B}{r_{AB}} - \frac{1}{2}\sum_{i}^{N_\mathrm{e}} \Delta_i - \sum_{i}^{N_\mathrm{e}} \sum_{A}^{N_\mathrm{n}} \frac{Z_A}{r_{iA}} + \sum_{i}^{N_\mathrm{e}} \sum_{j>i}^{N_\mathrm{e}} \frac{1}{r_{ij}} \ , \tag{1.1}$$

where Δ_A and Δ_i are the one-particle Laplacians of atoms and electrons, respectively, r_{AB}, r_{iA}, and r_{ij} are the distances between particles, and Z_A and m_A are the nuclear charge and nuclear mass of nucleus A, respectively. The importance of the Hamiltonian arises from its role in the equation of motion in quantum mechanics, the time-dependent Schrödinger equation [1]:

$$i\hbar \frac{\partial}{\partial t} \Psi = \hat{\mathcal{H}} \Psi \ , \tag{1.2}$$

which allows for stationary states determined by the time-independent eigenvalue equation

$$\hat{\mathcal{H}} \Psi = E \Psi \ . \tag{1.3}$$

In the widely used Born-Oppenheimer approximation [2], nuclear and electronic motion are separated and the first and most time consuming step becomes the description of the electronic structure in a field of fixed positively charged nuclei. The electronic states are then determined by the equation

$$\hat{H} \Psi = E \Psi \ , \tag{1.4}$$

where the electronic Hamiltonian $\hat{H}$ takes the form

$$\hat{H} = \sum_{A}^{N_\mathrm{n}} \sum_{B>A}^{N_\mathrm{n}} \frac{Z_A Z_B}{r_{AB}} - \frac{1}{2} \sum_{i}^{N_\mathrm{e}} \Delta_i - \sum_{i}^{N_\mathrm{e}} \sum_{A}^{N_\mathrm{n}} \frac{Z_A}{r_{iA}} + \sum_{i}^{N_\mathrm{e}} \sum_{j>i}^{N_\mathrm{e}} \frac{1}{r_{ij}} \,. \tag{1.5}$$

Despite the simplifications afforded by the Born-Oppenheimer approximation, solving (1.4) requires computationally demanding numerical procedures for all but one-electron systems. On the other hand, obtaining accurate approximations to these solutions is essential to a quantitative theoretical understanding of chemical phenomena and has thus been the focal point of an extensive amount of scientific research since the early 20th century. This has led to large advances in approximation techniques which, combined with massive increases in computational resources, means that today a wide array of chemical problems can be studied theoretically.

This work is concerned with deterministic computational methods in quantum chemistry, which can be grouped into two categories that distinguish between the main object of interest. In so-called wave function methods, one attempts to construct an accurate model of the many-body wave function Ψ for the state of interest. Many-body wave functions are in general complicated, highly-dimensional mathematical objects. They are dependent on $3N_\mathrm{e}$ spatial coordinates and N_e spin coordinates, which account for the two possible states of the intrinsic angular momentum of each electron. In addition, the antisymmetry properties of fermionic states lead to a complex nodal structure.

Alternatively, a large branch of quantum chemistry known as *density functional theory* (DFT) aims to describe chemical systems through the ground state electronic density ρ, which is a particle density defined on $\mathbb{R}^3$, and is constructed by integration over all but the spatial coordinates of one electron of the electronic probability density $|\Psi|^2$. Using the Hohenberg-Kohn theorems [3], one can reduce the problem of determining the ground state energy to finding the exchange-correlation functional of ρ that gives the external potential of a noninteracting system with the same ground state density. Given such a functional, which must be approximated in practice, the Kohn-Sham equations [4] determine the optimal one-electron functions for building the electronic density of the non-interacting system, which can then be used to determine the energy and other properties of the system.

In both wave function and Kohn-Sham DFT approaches, orbitals, i.e., functions representing one-electron states, play an important role. The orbitals are typically defined as expansions in a large one-electron basis set, and the determination of electronic energies requires the calculation of an extremely large number of integrals over such basis functions. While the integrals themselves can be evaluated quite efficiently, their large numbers can quickly exhaust the resources of even the most modern supercomputers as the size of the studied chemical system increases. On the other hand, the inherent sparsity found in molecular calculations means that a very large proportion of the integrals contribute negligibly to the final energies. For this reason, integral bounds can be used in combination with screening algorithms to drastically reduce the computational costs of many quantum chemical methods. These techniques have a long history in quantum chemistry and have allowed for the treatment of very large systems.

The main work of this cumulative thesis is comprised of **Publications I-IV**, where integral bounds and screening algorithms are developed and applied in various electronic structure theories. While this work is motivated by the needs of modern wave function based methods, the techniques developed are also applicable to hybrid and range-separated DFT models and applications to theories from both paradigms are presented.

In Chapter 2, a survey of wave function based electronic structure theory is presented with special attention given to those methods that have motivated the main work. The central goal of the present work is to allow for efficient and accurate implementations of so-called *explicitly correlated* methods [5, 6], which introduce explicit multi-electron interactions not present in traditional excitation-based correlation treatments. These methods have seen rapid development in the last two decades and have become very effective at reducing basis set incompleteness errors, which present a difficult problem in traditionally correlated methods. At the same time, this family of methods introduces complicated molecular integrals which require new approaches for developing bounds and screening methods. The R12/F12 explicitly correlated theory [5–8] is discussed in detail, especially in regards to its application to second order Møller-Plesset perturbation theory (MP2). Afterwards, approximation techniques for integrals are discussed which can be used to drastically reduce the computational effort required in quantum chemical calculations. Their application to explicitly correlated MP2-F12 and the exchange matrix of Hartree-Fock theory is outlined. A brief introduction to non-Hermitian quantum chemistry is also given, as this is the subject of **Publication III**.

In **Publication I**, a Schwarz-type inequality is introduced that captures the decay of four-center, two-electron integrals as the distance between electronic charge distributions increases. This decay is mediated by a two-electron operator which is traditionally the Coulombic inverse distance operator r_{12}^{-1}, but, in explicitly correlated and range-separated DFT methods, can also be an exponentially decaying function of the inter-electronic distance r_{12}, leading to much faster distance decay. The inequality is formulated for a large class of operator functions $F(r_{12})$ covering all the relevant use cases. One drawback of the bound is that it does not perform as well for integrals that are negligible due to a *combination* of operator decay and small basis function overlap within the charge distributions, and instead works best when one of these factors dominates. To ameliorate this problem, an integral estimate is introduced that captures both sources of decay simultaneously, but is no longer a rigorous upper bound. In practice, the estimate is close enough to an upper bound that it is still very useful. While these bounds and estimates work well for four-center two-electron integrals, they are not easily generalized to different number of basis function centers or multi-electron integrals. On the other hand they turn out to perform extremely well for the typical operators used in range-separated theories.

Publication II, which can be regarded as the most important and fundamental result of this work, introduces a completely new integral partitioning procedure that leads to very general upper bounds for all types of electronic integrals involving any number of electrons and basis function centers. A hierarchy of bounds is obtained that can be used to formulate scaling-consistent screening algorithms with optimal memory space requirements and time complexity. The screening techniques can be used to treat any integral in explicitly

correlated methods and thus pave the way for efficient and low-scaling formulations of the required algorithms. In addition, the generality of the bounds developed here make them useful beyond the intended application to explicitly correlated methods, and in **Publication IV**, they are used to rigorously screen the integrals that arise in the semi-numerical treatment of Fock exchange. This simplifies the previously used algorithm and allows for an extremely efficient calculation of the exchange-matrix on graphical processing units (GPUs).

Publication III is a collaboration with the Jagau group that focuses on the problems that arise in complex basis function methods, a subset of non-Hermitian quantum chemistry. In non-Hermitian quantum chemistry, an unbounded similarity transformation of the electronic Hamiltonian $\hat{H}$ allows one to treat electronic resonance states, i.e., metastable states with regards to auto-ionization, using the same methods developed for bound states with only slight modifications. In the method of complex basis functions, the complex scaling of the Hamiltonian that results from the similarity transformation is approximated by using a partially complex-scaled one-electron basis set. This introduces new integrals which could not be treated directly using the screening techniques employed in the corresponding bound state calculations. The integral bound introduced in this work overcomes the problems associated with the non-Hermiticity of the integral tensor and thus allows for the application of the usual screening procedures that are essential for treating large molecular systems.

In Chapter 3, the bounds developed in **Publication II** are extended to the case of integrals over local molecular orbitals, which require a numerical, instead of analytical, calculation for some of the factors involved.

Throughout this work both explicit and implicit summation notation are used. Summation over indices is implied whenever they appear on only one side of an equation. "Physicist's" integral notation [9] is used for molecular integrals over any number of electrons, e.g., for any two real-valued n-electron functions g and h and an n-electron operator $\hat{O}_{1\cdots n}$,

$$\langle g|\hat{O}_{1\cdots n}|h\rangle = \int d\boldsymbol{x}_1 \ldots \int d\boldsymbol{x}_n \, g(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)\hat{O}_{1\cdots n}h(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \ ,$$

and for $2n$ real-valued one-electron functions $g_1, \ldots, g_n, h_1, \ldots, h_n$,

$$\langle g_1 \cdots g_n|\hat{O}_{1\cdots n}|h_1 \cdots h_n\rangle = \int d\boldsymbol{x}_1 \ldots \int d\boldsymbol{x}_n \, g_1(\boldsymbol{x}_1) \cdots g_n(\boldsymbol{x}_n)\hat{O}_{1\cdots n}h_1(\boldsymbol{x}_1) \cdots h_n(\boldsymbol{x}_n) \ .$$

"Chemist's" integral notation [9] is frequently used for two-electron integrals, e.g., for any two one-electron functions p and q and a multiplicative two-electron operator $\hat{G}_{12}$,

$$(p|\hat{G}_{12}|q) = \int d\boldsymbol{x}_1 \int d\boldsymbol{x}_2 \, p(\boldsymbol{x}_1)\hat{G}(\boldsymbol{x}_1, \boldsymbol{x}_2)q(\boldsymbol{x}_2) \ .$$

This notation is also used for two-electron integrals over complex-scaled basis functions in Section 2.6, where p and q may be complex valued, but integral notation without implicit complex-conjugation is used. This is because the complex parts of the basis functions are generally not complex-conjugated in the method of complex basis functions, since they originate from analytic continuation and not from the use of a complex wave function [10].

Chapter 2

Theory

2.1 Foundations

The numerical solution of equation (1.4) starts with the explicit definition of an approximate many-body wave function Φ, which is motivated by the underlying physics and mathematical structure. In non-relativistic electronic structure theory, the one-particle Hilbert space $\mathfrak{H}$ is spanned by an orthonormal basis of real square-integrable spin orbitals $\phi_\alpha; \alpha \in \mathbb{N}$. These functions depend on the spatial and spin coordinates $\boldsymbol{x} = (\boldsymbol{r}, \omega)$ of a single electron and are orthonormal with respect to the scalar product:

$$\langle \phi_\alpha | \phi_\beta \rangle = \int \phi_\alpha(\boldsymbol{x}_1)\phi_\beta(\boldsymbol{x}_1)d\boldsymbol{x}_1 = \delta_{\alpha\beta} \ . \tag{2.1}$$

They are separable in to spatial and spin functions

$$\phi_\alpha = \varphi_\alpha(\boldsymbol{r}_1)s_\alpha(\omega) \ , \tag{2.2}$$

where s_α is one of two possible orthogonal spin functions $s_\alpha \in \{s^+, s^-\}$:

$$\int s^+(\omega)s^+(\omega)d\omega = \int s^-(\omega)s^-(\omega)d\omega = 1 \ , \qquad \int s^+(\omega)s^-(\omega)d\omega = 0 \ . \tag{2.3}$$

From this one-particle complete basis set (CBS), an orthonormal basis for the space of N_e-electron wave functions can be constructed as the set of all *Slater determinants*

$$\Phi_{\alpha_1,\ldots,\alpha_{N_\mathrm{e}}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N_\mathrm{e}}) = \frac{1}{\sqrt{N_\mathrm{e}!}} \begin{vmatrix} \phi_{\alpha_1}(\boldsymbol{x}_1) & \phi_{\alpha_1}(\boldsymbol{x}_2) & \cdots & \phi_{\alpha_1}(\boldsymbol{x}_{N_\mathrm{e}}) \\ \phi_{\alpha_2}(\boldsymbol{x}_1) & \phi_{\alpha_2}(\boldsymbol{x}_2) & \cdots & \phi_{\alpha_2}(\boldsymbol{x}_{N_\mathrm{e}}) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{\alpha_{N_\mathrm{e}}}(\boldsymbol{x}_1) & \phi_{\alpha_{N_\mathrm{e}}}(\boldsymbol{x}_2) & \cdots & \phi_{\alpha_{N_\mathrm{e}}}(\boldsymbol{x}_{N_\mathrm{e}}) \end{vmatrix} \ , \tag{2.4}$$

with $\alpha_1 < \alpha_2 < \ldots < \alpha_{N_\mathrm{e}}$. This determinant form results from the restriction to anti-symmetric elements of the N_e-particle Hilbert space $\mathfrak{H}_{N_\mathrm{e}} = \otimes_{i=1}^{N_\mathrm{e}}\mathfrak{H}$, which is necessary for

describing the physics of electrons. Thus, any N_e-electron wave function Ψ can be written as an infinite sum

$$\Psi = \sum_i c_i \Phi_i \ , \tag{2.5}$$

where i runs over all the Slater determinants in the complete basis.

In numerical practice, an basis of infinite size cannot be realized and one is restricted to the use of a finite set. However, the mathematical structure described above still motivates most approaches. In this setting, the construction of the finite set of spin orbitals becomes crucial and, in molecular calculations, their spatial parts are typically expanded in a fixed set of Gaussian basis functions. The use of Gaussians is motivated by a compromise between capturing the underlying physics and maximizing computational efficiency [11].

2.1.1 Hartree-Fock Theory

In Hartree-Fock (HF) theory, the wave function is approximated as a single Slater determinant Φ_HF of N_e spin orbitals ϕ_i, $i \in \{1, \ldots, N_\mathrm{e}\}$. The spatial parts, φ_i, of these orbitals are typically expanded in a fixed, non-orthogonal, atom-centered Gaussian basis set $\chi_\mu, \mu \in \{1, \ldots, N_\mathrm{b}\}$, known as atomic orbitals (AOs):

$$\varphi_i = \sum_\mu C_{\mu i} \chi_\mu \ , \tag{2.6}$$

and the expansion coefficients $C_{\mu i}$ are determined through the variational minimization of the expectation value

$$E_\mathrm{HF} = \frac{\langle \Phi_\mathrm{HF} | \hat{H} | \Phi_\mathrm{HF} \rangle}{\langle \Phi_\mathrm{HF} | \Phi_\mathrm{HF} \rangle} \ , \tag{2.7}$$

under the constraint of orthonormality, $\langle \phi_i | \phi_j \rangle = \delta_{ij}$. This results in the canonical HF equations

$$\hat{f}\phi_i = \epsilon_i \phi_i, \qquad \hat{f} = \hat{t} + \hat{v} + \hat{j} - \hat{k} \ , \tag{2.8}$$

with the single-particle Fock operator $\hat{f}$ comprised of the kinetic energy operator $\hat{t}$, nuclear attraction operator $\hat{v}$, and mean-field Coulomb and exchange operators $\hat{j}$ and $\hat{k}$:

$$\hat{t}\phi_i(\boldsymbol{x}_1) = -\frac{1}{2}\Delta_{\boldsymbol{r}_1}\phi_i(\boldsymbol{x}_1), \qquad \hat{v}\phi_i(\boldsymbol{x}_1) = -\sum_A^{N_\mathrm{n}} Z_A \frac{\phi_i(\boldsymbol{x}_1)}{r_{1A}} \tag{2.9}$$

$$\hat{j}\phi_i(\boldsymbol{x}_1) = \phi_i(\boldsymbol{x}_1) \sum_j^{N_\mathrm{e}} \int \frac{|\phi_j(\boldsymbol{x}_2)|^2}{r_{12}} d\boldsymbol{x}_2, \qquad \hat{k}\phi_i(\boldsymbol{x}_1) = \sum_j^{N_\mathrm{e}} \phi_j(\boldsymbol{x}_1) \int \frac{\phi_j(\boldsymbol{x}_2)\phi_i(\boldsymbol{x}_2)}{r_{12}} d\boldsymbol{x}_2 \ . \tag{2.10}$$

In the closed shell Roothan-Hall scheme [12], these non-linear equations are expressed in the AO basis set leading to the Matrix eigenvalue problem

$$\boldsymbol{F'C'} = \boldsymbol{C'\epsilon} \ , \tag{2.11}$$

where

$$\boldsymbol{F'} = \boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{F}\boldsymbol{S}^{-\frac{1}{2}} \ , \quad \boldsymbol{C'} = \boldsymbol{S}^{\frac{1}{2}}\boldsymbol{C} \ , \tag{2.12}$$

$\boldsymbol{S}$ is the overlap matrix with elements $S_{\mu\nu} = \langle\mu|\nu\rangle$, $\boldsymbol{F}$ is the Fock matrix with elements $F_{\mu\nu} = \langle\mu|\hat{f}|\nu\rangle$, $\boldsymbol{C}$ is the coefficient matrix made up of the initially unknown expansion coefficients $C_{\mu i}$, and $\boldsymbol{\epsilon}$ is the diagonal matrix of orbital energies ϵ_i. The use of $\boldsymbol{S}^{-\frac{1}{2}}$ has the effect of representing the problem in an orthogonalized AO basis set and is known as Löwdin orthogonalization [13]. Diagonalization of $\boldsymbol{F'}$ leads to a set of N_b spatial orbitals and corresponding orbital energies. The resulting spin orbitals are called molecular orbitals (MOs) or HF orbitals. The N_e molecular orbitals with the lowest orbital energies are the *occupied* MOs used in the HF Slater determinant Φ_HF. Because the operators $\hat{j}$ and $\hat{k}$ depend on the occupied MOs, the HF equations are solved iteratively by constructing the Fock operator from the solutions of the previous iteration, starting with an initial guess, until some criterion for convergence is met. This iterative procedure is known as the self-consistent field (SCF) method. If convergence is reached, the occupied MOs provide the optimal representation of the wave function as a single Slater determinant in the chosen basis. The remaining *virtual* MOs are interpreted as one-particle excited states.

2.1.2 Orbital Spaces

Table 2.1: Orbital spaces and notation used in this work

Description	Indices	Cardinality
Atomic orbital basis (non-orthogonal)	$\mu, \nu, \lambda, \sigma$	$N_\mu = N_\mathrm{b}$
HF occupied orbitals	i, j, k, l, m	$N_i = N_\mathrm{e}$
HF virtual orbitals	a, b, c, d	N_a
Arbitrary HF orbitals	p, q, r, s	N_p
Complete one-electron basis set (CBS)	α, β, γ	∞
Approximate CBS-spanning auxiliary set (non-orthogonal)	P, Q, R, S	N_P
Complement of HF space in CBS, $\mathrm{span}(\{\alpha'\}) = \mathrm{span}(\{\alpha\})\backslash\mathrm{span}(\{p\})$	α', β', γ'	∞
Approximate complement-spanning auxiliary set (CABS) (orthogonal)	p'', q'', r'', s''	$N_{p''}$
Approximate CBS-spanning auxiliary set (orthogonal) $\{p'\} = \{p\} \cup \{p''\}$	p', q', r', s'	$N_{p'}$

The remaining discussion will rely heavily on the distinction between different sets of orbitals and their relation to the HF procedure. Table 2.1 serves as a reference for understanding the notation for the different orbitals and orbital spaces used in this work.

Particularly important for explicitly correlated theories is the distinction between the space spanned by the HF orbitals and the infinitely dimensional complementary space of functions outside of this span. This complementary space is spanned exactly by an infinite

set of orthogonal orbitals $\{\alpha'\}$, and approximately by the finite orthogonal complementary auxiliary basis set (CABS) $\{p''\}$, which is constructed to be orthogonal to the set of HF orbitals $\{p\}$ [14]. The union of the HF and CABS orbitals $\{p'\} = \{p\} \cup \{p''\}$ thus builds an orthogonal set of orbitals which approximately spans the complete one-electron space.

2.1.3 Electron Correlation and Correlated Wave Functions

The HF equations (2.8) represent a mean-field treatment that neglects all but the exchange-type Fermi-correlations between same-spin electrons, which are necessary for an antisymmetric wave function. The electron-electron Coulomb interactions are treated in an averaged manner, so that the Coulomb repulsion is underestimated in regions of $\mathbb{R}^{3N_e}$ where two (or more) electrons become close. While the HF wave function often represents a qualitatively good approximation, an accurate description of the missing Coulomb correlation is required for quantitative accuracy. Traditionally, and in the following discussion, the correlation energy is defined as the difference between the exact non-relativistic energy of the system and the complete basis set HF energy.

Configuration Interaction

A straightforward way to include correlation effects is provided by configuration interaction (CI) wave functions. A CI wave function is defined as an expansion of Slater determinants within the orthonormal HF orbitals. Because the HF Slater determinant typically provides an accurate description of the system, the resulting HF orbitals are seen as an appropriate candidate for a finite Slater determinant expansion. In the case of full configuration interaction (FCI), all possible determinants are included. The resulting approximate wave function can be written as

$$\Phi_{\mathrm{FCI}} = (1 + \sum_{ia} c_i^a \hat{a}_a^\dagger \hat{a}_i + \sum_{\substack{i<j \\ a<b}} c_{ij}^{ab} \hat{a}_a^\dagger \hat{a}_i \hat{a}_b^\dagger \hat{a}_j + \sum_{\substack{i<j<k \\ a<b<c}} c_{ijk}^{abc} \hat{a}_a^\dagger \hat{a}_i \hat{a}_b^\dagger \hat{a}_j \hat{a}_c^\dagger \hat{a}_k + \ldots)\Phi_{\mathrm{HF}} \, . \tag{2.13}$$

Here, the indices $i, j, \ldots$ and $a, b, \ldots$ are used for occupied and virtual MOs respectively. The *annihilation* operator $\hat{a}_i$ removes an electron in the state ϕ_i from a Slater determinant, while the *creation* operator $\hat{a}_a^\dagger$ adds an electron in the state ϕ_a. The resulting determinants are known as singly, doubly, triply, etc., excited determinants and all possible excitations are included. Linear variation can be used to determine the optimal expansion coefficients $(c_i^a, c_{ij}^{ab}, c_{ijk}^{abc}, \ldots)$ and this results in the best possible approximation of the exact wave function within the given one-electron basis. In practice, FCI is computationally unfeasible for all but very small systems, because the number of determinants grows *factorially* with the number of spin orbitals [11]. The computational scaling of the method with respect to the number of electrons can be reduced by restricting the number of allowed excitations, but this leads to the loss of size consistency [11].

Second-Order Møller-Plesset Perturbation Theory

Another approach to electronic correlation is to treat the electron-electron interactions using perturbation theory. In the case of Møller-Plesset perturbation theory, the electronic Hamiltonian (excluding nuclear repulsions) is written as

$$\hat{H} = \hat{H}_0 + \hat{U} \ , \tag{2.14}$$

where

$$\hat{H}_0 = \sum_i^{N_\mathrm{e}} \hat{f}_i, \qquad \hat{U} = \sum_{i<j}^{N_\mathrm{e}} \frac{1}{r_{ij}} - \sum_i^{N_\mathrm{e}} \left(\hat{\jmath}_i - \hat{k}_i\right) \ . \tag{2.15}$$

The unperturbed Hamiltonian $\hat{H}_0$ is the sum of the Fock operators for each electron and, for any Slater determinant $\Phi_{p_1,\ldots,p_{N_\mathrm{e}}}$ made up of a subset of HF orbitals ϕ_{p_k}, $k \in \{1,\ldots,N_\mathrm{e}\}$, one has

$$\hat{H}_0 \Phi_{p_1,\ldots,p_{N_\mathrm{e}}} = (\sum_k^{N_\mathrm{e}} \epsilon_{p_k}) \Phi_{p_1,\ldots,p_{N_\mathrm{e}}} \ , \tag{2.16}$$

which is taken as the zeroth-order (unperturbed) eigenvalue equation, with the ground state being given by the HF determinant of occupied MOs. The operator $\hat{U}$ is treated as a perturbation to (2.16). A Taylor expansion of the ground state energy and wave function in the order of the perturbation leads to first-, second-, and higher-order corrections to the unperturbed ground state energy and wave function, which define approximate solutions to the full eigenvalue problem. The zeroth-order energy is simply the sum of occupied orbital energies and adding the first-order energy correction to this sum gives exactly the HF energy, so that the second-order correction to the energy, E_MP2, is the first to introduce electronic correlation. The first-order correction to the wave function is expanded *approximately* in the full set of excited HF determinants and intermediate normalization [11] leads to the following expression:

$$E_\mathrm{MP2} = \sum_p \frac{|\langle \Phi_\mathrm{HF}| \hat{H} - \sum_k^{N_\mathrm{e}} \hat{f}_k |\Phi_p\rangle|^2}{(\sum_i^{N_\mathrm{e}} \epsilon_i) - E_p} \ , \tag{2.17}$$

where p runs over all the excited determinants Φ_p, with their corresponding zero-order eigenvalues (sums of orbital energies) denoted as E_p. The second term in the numerator vanishes in all cases, while the first is only non-zero in the case of double excitations, so that the expression simplifies to

$$E_\mathrm{MP2} = -\frac{1}{2} \sum_{\substack{ij \\ ab}} \frac{|\langle ij|r_{12}^{-1}||ab\rangle|^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \ , \tag{2.18}$$

where $||ab\rangle$ is an antisymmetrized two-electron state $||ab\rangle = 2^{-1/2}\big(|ab\rangle - |ba\rangle\big)$.

It is noted that the MP2 energy correction can be written as a sum of pair energies $E_\mathrm{MP2} = \sum_{ij} e_{ij}^\mathrm{MP2}$, where the definition of e_{ij}^MP2 is evident from equation (2.18). In Section

(2.2.2), the fact that the second-order pair energies e_{ij} can also be obtained as the minima of the variational equations

$$J_{ij}^2[u_{ij}] = \langle u_{ij}|\hat{f}_1 + \hat{f}_2 - \epsilon_i - \epsilon_j|u_{ij}\rangle + 2\langle u_{ij}|g_{12}||ij\rangle \ , \tag{2.19}$$

is used to introduce explicit correlation to MP2 theory. Here, the J_{ij}^2 are the second-order Hylleraas pair functionals [15], and the u_{ij} represent abstract two-electron states. Variational minimization of J_{ij}^2 with respect to u_{ij} gives the exact second-order correction e_{ij}^{exact}. In the MP2 formalism, e_{ij}^{exact} is only reached in the limit of a complete basis set, where the expansion of the first order wave function in the doubly excited determinants is exact. The standard MP2 pair energies are recovered from the Hylleraas functionals by using the linear ansatz

$$|u_{ij}^{\text{MP2}}\rangle = \sum_{a<b} t_{ij}^{ab}||ab\rangle \ , \tag{2.20}$$

and minimizing the J_{ij}^2 with respect to the amplitudes t_{ij}^{ab}.

Coupled Cluster Theory

The coupled cluster (CC) method is currently the most successful way for accurately treating electron correlation for systems whose ground state is described well by a single determinant. Similar to CI, the CC wave function is expanded in terms of excited determinants, but instead of a linear expansion one uses the exponential form

$$\Phi_{\text{CC}} = \exp(\hat{T})\Phi_{\text{HF}} \ , \quad \hat{T} = \hat{T}_1 + \hat{T}_2 + \dots \ , \tag{2.21}$$

where the excitation operators

$$\hat{T}_1 = \sum_{ia} t_i^a \hat{a}_a^\dagger \hat{a}_i, \quad \hat{T}_2 = \sum_{\substack{i<j \\ a<b}} t_{ij}^{ab} \hat{a}_a^\dagger \hat{a}_i \hat{a}_b^\dagger \hat{a}_j, \quad \dots \ , \tag{2.22}$$

also appear, albeit in a linear fashion, in the CI wave function (2.13). The advantage over CI is that the operator $\hat{T}$ can be truncated to a given order n, i.e., $\hat{T} \approx \hat{T}^{(n)} = \sum_{k=1}^{n} \hat{T}_k$, without sacrificing size consistency, while retaining high accuracy. This is because the truncated wave function still contains all possible excited determinants as a consequence of the exponential form. The amplitudes of the determinants not included in $\hat{T}$ are approximated naturally as sums of products of those that are included, which reduces the computational cost relative to FCI drastically. One trade-off of the CC formalism is that a variational optimization is no longer computationally feasible and, instead, equations for optimizing the amplitudes are obtained by projection of the similarity transformed equation

$$\exp(-\hat{T}^{(n)})\hat{H}\exp(\hat{T}^{(n)})|\Phi_{\text{HF}}\rangle = E_{\text{CC}}|\Phi_{\text{HF}}\rangle \ , \tag{2.23}$$

onto the set of excited determinants up to order n. The CC energy is obtained by projection onto the HF determinant:

$$E_{\text{CC}}^{(n)} = \langle\Phi_{\text{HF}}|\exp(-\hat{T}^{(n)})\hat{H}\exp(\hat{T}^{(n)})|\Phi_{\text{HF}}\rangle \ . \tag{2.24}$$

This procedure leads to tractable equations for the amplitudes but the CC energy is no longer variational and may, therefore, be lower than the exact energy. This drawback is usually compensated for by the high accuracy of the method.

2.2 Explicit Correlation

2.2.1 Basis Set Incompleteness Errors and Electronic Cusps

The correlated methods presented in Section 2.1.3 all suffer from rather large basis set incompleteness errors (BSIE) and prohibitively large one-electron basis sets are needed to reach their full potential. To increase the size of the basis towards completeness, shells of functions with increasing total angular momenta L are added. The nodes of these functions give the basis set more flexibility to describe correlation but the convergence for correlated methods can be shown for atoms to go as $\mathcal{O}((L+1)^{-3})$ [16–20], which is rather poor. At the same time, increasing the size of the basis set causes large increases in computational cost.

The reason for the slow convergence of the correlated energies with the size of the basis set has been known for some time. In 1957, Kato [21] was able to characterize the regularity of exact eigenfunctions of $\hat{H}$, showing that their partial derivatives are singular at the points where two electrons coalesce, and that the dependence of the wave function on the inter-electronic distance r_{12} near the coalescence points is linear. This analysis and further work [22, 23] led to the formulation of the s- and p-wave *cusp conditions*, which state that in regions close to electron coalescence ($r_{12} \approx 0$),

$$\Psi^{(1)} = \frac{r_{12}}{2(s+1)}\Psi^{(0)} + \mathcal{O}(r_{12}^2) \ , \tag{2.25}$$

where, $s = 0$ for singlet and $s = 1$ for triplet states, and $\Psi^{(1)}$ is the first order correction to the zeroth order wave function $\Psi^{(0)}$ in a perturbative treatment in which the electron-electron Coulomb interaction is treated as the perturbation to a bare nucleus Hamiltonian.

The form of the exact electronic cusp for the coalescence of the two electrons of the helium atom is shown in Figure 2.1, where the slow convergence of CI expansions at the coalescence point is demonstrated.

This slow convergence of excitation based correlation energies means that chemical accuracy is almost impossible to reach by simply increasing the size of the one-electron basis set. However, due to the relatively predictable nature of basis set convergence, correlation energy extrapolation techniques [24–26] can be used in combination with Dunning's correlation consistent basis sets [27] to obtain near CBS results. Another, often more efficient, way to reach the CBS limit is to go beyond the restriction to antisymmetric products of one-electron functions. It is well known that the correlation cusp can be represented much more efficiently, i.e., with fewer terms, when basis functions are included that depend inseparably on two electrons. Wave functions that include terms explicitly dependent on inter-particle distances show much faster convergence with respect to the one-electron basis set.

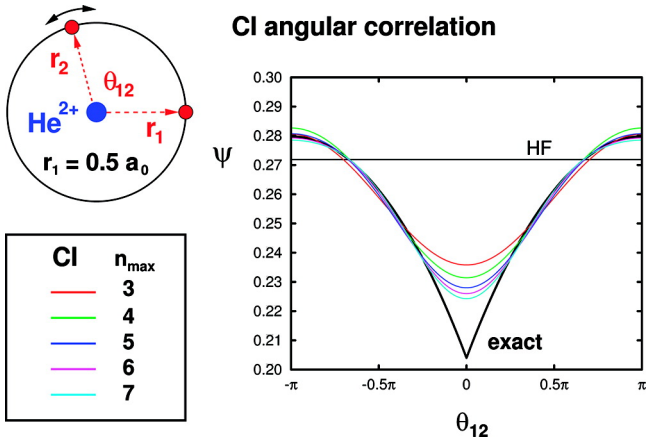


Figure 2.1: Exact angular correlation cusp and convergence of a CI description with basis sets of increasing maximum total angular momentum n_{max}. Reprinted with permission from Ref. [5]. Copyright 2012 American Chemical Society.

2.2.2 Explicitly Correlated Wave Functions

Explicitly correlated wave functions are defined by the inclusion of terms explicitly dependent on electron-electron distances. The first arguments for including such terms were made by Slater in 1928 [28], whose analysis of the wave function of helium led him to suggest multiplying the wave function by the factor $\exp(-\frac{1}{2}r_{12})$ in order to account for the Coulomb singularity at coalescence. Interestingly, this ansatz fulfills Kato's cusp conditions for helium exactly, despite the fact that these would only be formulated many years later, in 1957. The first practical calculations employing explicitly correlated wave functions were published a year later, in 1929, by Hylleraas [29]. Hylleraas had previously failed to satisfactorily reproduce experimental results of the ionization potential of the Helium atom using a CI-type expansion [30], and had discovered that by including odd powers of r_{12} in his expansion, he could reduce the discrepancy with experiment to 0.01 eV using only a very compact six-term wave function. This breakthrough provided strong evidence for the validity of wave mechanics but proved his method proved difficult to extend to polyatomic, many-electron systems, where the inclusion of products of inter-electronic distances leads to highly-dimensional integrals over $3N_e$ variables.

More tractable, but still highly accurate explicitly correlated methods were then developed for studying few-electron atoms. One example is the Hylleraas-configuration interaction (Hylleraas-CI) method [31–35], which has been used for one-center systems with up to ten electrons [36]. In this method the wave function is defined generally as

$$\Psi = \sum_{\nu=0}^{\nu_{\max}} \sum_{n} C_{n,\nu} \sum_{i<j}^{N_e} r_{ij}^{\nu} \Phi_n, \tag{2.26}$$

where n runs over a predetermined set of CI-type Slater determinants $\{\Phi_n\}$, typically $\nu_{\max} = 1$, and the optimal coefficients $C_{n,\nu}$ are determined through variational optimiza-

tion. The restriction to terms containing only one inter-electronic distance means that only up to four-electron integrals are required, but the computational cost of the method remains very high. Another class of explicitly correlated methods is obtained from exponentially correlated Gaussian (ECG) wave functions [37–41], which employ spherical Gaussian geminals in a variational ansatz. The wave function takes the form [5]

$$\Psi = \sum_k c_k \Phi_k \tag{2.27}$$

$$\Phi_k = \hat{\mathcal{A}} X \hat{\mathcal{P}} \exp(-\sum_{i<j} \beta_{ijk} r_{ij}^2) \prod_{i=1}^{N_\mathrm{e}} g_{i_k}(\boldsymbol{r}_i) \ , \tag{2.28}$$

where $\hat{\mathcal{A}}$, X, and $\hat{\mathcal{P}}$ are the N_e-electron antisymmetrizer, spin-function, and symmetry group projector, respectively. The g_{i_k} are Cartesian Gaussian basis functions, which are explicitly correlated through the Gaussian geminals for each electron pair. The use of Gaussians means that the electronic cusp is not captured exactly, but allows for the analytical computation of many-electron integrals. Still, the costs remain high, and the main drawback is the need for non-linear optimization of the geminal exponents.

Also of note is the family of transcorrelated (TC) Hamiltonian methods first proposed by Boys and Handy [42] and recently revisited in Refs. [43–46]. Here, the explicitly correlated terms are folded into the Hamiltonian through a similarity transformation of the form:

$$\hat{H}_\mathrm{TC} = e^{-F} \hat{H} e^{F}, \qquad F = \sum_{i<j} f(\boldsymbol{r}_i, \boldsymbol{r}_j), \tag{2.29}$$

where f is a two-electron correlation function. The function f can be chosen such that the transformed Hamiltonian has no singularities, but a drawback is that $\hat{H}_\mathrm{TC}$ contains three-electron interactions and is a non-Hermitian operator, so that one must abandon the traditional variational approach when solving for its eigenfunctions.

The most successful explicitly correlated methods are based on the so-called R12 method introduced by Kutzelnigg and Klopper [7, 8], which introduces explicitly correlated terms as non-variational corrections. Modern versions of their theory replace the original linear correlation factor r_{12} with a more suitable general r_{12}-dependent correlation factor typically denoted F_{12}. For this reason, these methods will be referred to as F12 methods in this work, as has become standard practice [5, 6]. A breakthrough in Kutzelnigg and Klopper's approach was the realization that resolution-of-the-identity (RI) approximations (see Section 2.3.1) can be used to reduce the resulting three- and four-electron integrals to sums of products of two-electron integrals. The application of their work to the MP2 correlation energy is discussed in the following section (2.2.3), with emphasis placed on the general theory and an RI-free formulation. In Section 2.2.4, the explicitly correlated correction to coupled cluster theory, which is closely related to the MP2-F12 correction, is discussed briefly. The manner in which RI is applied to simplify the resulting F12 terms is discussed in Section 2.5, along with possibilities for reducing the scaling of the method further.

2.2.3 Explicitly Correlated Second-Order Møller-Plesset Perturbation Theory

The most straightforward way to introduce Kutzelnigg and Klopper's R12/F12 method is within MP2 theory. In MP2-F12 theory, the explicitly correlated terms are introduced at the level of the first order wave function correction in the Hylleraas pair functional. Whereas the typical MP2 energy is obtained from the pair functional in equation (2.20), in the MP2-F12 formulation the following pair functional is used:

$$|u_{ij}^{\mathrm{MP2-F12}}\rangle = |u_{ij}^{\mathrm{MP2}}\rangle + |u_{ij}^{\mathrm{F12}}\rangle \tag{2.30}$$

$$= \sum_{a<b} t_{ij}^{ab}||ab\rangle + \sum_{x<y} c_{ij}^{xy}\hat{Q}_{12}F_{12}||(\hat{S}_{xy}xy)\rangle \ . \tag{2.31}$$

The explicitly correlated functionals $|u_{ij}^{\mathrm{F12}}\rangle$ are linear expansions of antisymmetrized geminal states $F_{12}(|\boldsymbol{r}_1 - \boldsymbol{r}_2|)||xy\rangle$, where the geminal generating orbitals, x and y, typically run over the (active) HF occupied space in ground state calculations. The orthogonality of the geminal states with respect to the HF occupied space, and with respect to the double excitations within $|u_{ij}^{\mathrm{MP2}}\rangle$, is assured by the *strong orthogonality operator*

$$\hat{Q}_{12} = (1 - \hat{\mathrm{o}}_1)(1 - \hat{\mathrm{o}}_2)(1 - \hat{\mathrm{v}}_1\hat{\mathrm{v}}_2) \ , \tag{2.32}$$

where $\hat{\mathrm{o}}_n$ and $\hat{\mathrm{v}}_n$ are the projectors of the n-th electron onto the occupied and virtual HF spaces, respectively:

$$\hat{\mathrm{o}}_n \, g(\ldots, \boldsymbol{r}_n, \ldots) = \sum_i \phi_i(\boldsymbol{r}_n)\Big(\int \phi_i(\boldsymbol{r}_n) g(\ldots, \boldsymbol{r}_n, \ldots) d\boldsymbol{r}_n\Big) \tag{2.33}$$

$$\hat{\mathrm{v}}_n \, g(\ldots, \boldsymbol{r}_n, \ldots) = \sum_a \phi_a(\boldsymbol{r}_n)\Big(\int \phi_a(\boldsymbol{r}_n) g(\ldots, \boldsymbol{r}_n, \ldots) d\boldsymbol{r}_n\Big) \ . \tag{2.34}$$

While the projection onto the complement of the doubly occupied space provided by the term $(1 - \hat{\mathrm{o}}_1)(1 - \hat{\mathrm{o}}_2)$ is necessary for orthogonality to the reference configurations, the removal of the double excitations through $(1 - \hat{\mathrm{v}}_1\hat{\mathrm{v}}_2)$ is used to decouple the MP2 and F12 corrections as much as possible.

The rational generator

$$\hat{S}_{xy} = \frac{3}{8} + \frac{1}{8}\hat{P}_{xy} \tag{2.35}$$

$$\hat{P}_{xy}\phi_x(\boldsymbol{r}_1, \sigma_1)\phi_y(\boldsymbol{r}_2, \sigma_2) = \phi_x(\boldsymbol{r}_2, \sigma_1)\phi_y(\boldsymbol{r}_1, \sigma_2), \tag{2.36}$$

which was introduced by Ten-no [47], ensures the simultaneous fulfillment of both the s- and p-wave coalescence conditions.

The optimization of the MP2-F12 parametrization of the Hylleraas pair functional, $J_{ij}^2[u_{ij}^{\mathrm{MP2-F12}}]$, with respect to t_{ij}^{ab} and c_{ij}^{xy} leads to the following weakly coupled amplitude

equations:

$$g_{ij}^{ab} + f_c^a t_{ij}^{cb} + f_d^b t_{ij}^{ad} - t_{kj}^{ab} f_i^k - t_{il}^{ab} f_j^l + C_{xy}^{ab} c_{ij}^{xy} = 0 \tag{2.37}$$

$$V_{ij}^{xy} + C_{ab}^{\dagger xy} t_{ij}^{ab} + B_{vw}^{xy} c_{ij}^{vw} - X_{vw}^{xy} c_{kj}^{vw} f_i^k - X_{vw}^{xy} c_{il}^{vw} f_j^l = 0 \ , \tag{2.38}$$

where v and w are further geminal generating orbitals and the F12 intermediates are defined as

$$B_{vw}^{xy} = \langle (\hat{S}_{xy} xy) | F_{12} \hat{Q}_{12} (\hat{f}_1 + \hat{f}_2) \hat{Q}_{12} F_{12} || (\hat{S}_{vw} vw) \rangle \tag{2.39}$$

$$X_{vw}^{xy} = \langle (\hat{S}_{xy} xy) | F_{12} \hat{Q}_{12} F_{12} || (\hat{S}_{vw} vw) \rangle \tag{2.40}$$

$$V_{ij}^{xy} = \langle (\hat{S}_{xy} xy) | F_{12} \hat{Q}_{12} r_{12}^{-1} || ij \rangle \tag{2.41}$$

$$C_{xy}^{ab} = \langle ab | (\hat{f}_1 + \hat{f}_2) \hat{Q}_{12} F_{12} || (\hat{S}_{xy} xy) \rangle \tag{2.42}$$

$$g_{ij}^{ab} = \langle ab | r_{12}^{-1} || ij \rangle \tag{2.43}$$

$$f_q^p = \langle p | \hat{f} | q \rangle \tag{2.44}$$

For the amplitudes that satisfy equations (2.37) and (2.38), the second-order energy correction is given by

$$E_{\mathrm{MP2-F12}} = g_{ab}^{\dagger ij} t_{ij}^{ab} + V_{xy}^{\dagger ij} c_{ij}^{xy} \tag{2.45}$$

In practice, one usually avoids the need to optimize the geminal amplitudes c_{ij}^{xy} by using the *fixed amplitude ansatz* [47, 48]:

$$c_{ij}^{xy} = \delta_x^i \delta_y^j \ , \tag{2.46}$$

which fulfils the cusp conditions, gives accurate results, and reduces computational demands drastically [5, 6, 48]. Using canonical orbitals and fixed geminal amplitudes, solving (2.37) leads to the corresponding optimal double excitation amplitudes

$$t_{ij}^{ab} = -\frac{g_{ij}^{ab} + C_{ij}^{ab}}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \ . \tag{2.47}$$

The residual of equation (2.38) then becomes:

$$R_{ij}^{xy} = V_{ij}^{xy} - \frac{C_{ab}^{\dagger xy} g_{ij}^{ab} + C_{ab}^{\dagger xy} C_{ij}^{ab}}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} + B_{ij}^{xy} - X_{kj}^{xy} f_i^k - X_{il}^{xy} f_j^l \ , \tag{2.48}$$

and the second-order correction is

$$E_{\mathrm{MP2-F12}} = g_{ab}^{\dagger ij} t_{ij}^{ab} + V_{xy}^{\dagger ij} c_{ij}^{xy} + c_{xy}^{\dagger ij} R_{ij}^{xy} \tag{2.49}$$

$$= g_{ab}^{\dagger ij} t_{ij}^{ab} + V_{ij}^{ij} + R_{ij}^{ij} \tag{2.50}$$

$$= -\frac{g_{ab}^{\dagger ij} g_{ij}^{ab} + 2 g_{ab}^{\dagger ij} C_{ij}^{ab} + C_{ab}^{\dagger ij} C_{ij}^{ab}}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} + 2 V_{ij}^{ij} + B_{ij}^{ij} - X_{kj}^{ij} f_i^k - X_{il}^{ij} f_j^l \ . \tag{2.51}$$

Two approximations that are often employed in the context of MP2-F12 theory are the *generalized Brillouin condition* (GBC), which assumes that the Fock matrix element between an occupied orbital and any function α' outside the span of the HF orbital set vanishes,

$$f_i^{\alpha'} \approx 0 \ , \tag{2.52}$$

and the *extended Brillouin condition* (EBC), which assumes the same for a virtual orbital,

$$f_a^{\alpha'} \approx 0 \ . \tag{2.53}$$

While the GBC is typically a much better approximation than the EBC, both become better as the size of the HF orbital basis set increases.

The EBC is often used to simplify equation (2.51) further, since the elements of C_{ij}^{ab} vanish under the EBC, leading to

$$E_{\mathrm{MP2-F12}}^{\mathrm{EBC}} = -\frac{g_{ab}^{\dagger ij} g_{ij}^{ab}}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} + 2V_{ij}^{ij} + B_{ij}^{ij} - X_{kj}^{ij} f_i^k - X_{il}^{ij} f_j^l \ . \tag{2.54}$$

Equation (2.54) represents one of the simplest explicitly correlated corrections for efficiently approaching complete basis set MP2 results using moderately sized one-electron basis sets. However, already here, a plethora of new integrals are contained in the F12 intermediates that are not present in traditional quantum chemistry.

Approximation-Free Integrals in MP2-F12

For the sake of completeness, the various two-, three-, and four-electron integrals needed in an exact evaluation of the fixed-amplitude F12 correction are now given.

The underlying terms within the V, X, and B intermediates take the form

$$\mathcal{V}_{kl}^{ij} = \langle ij | F_{12} \hat{Q}_{12} r_{12}^{-1} | kl \rangle \tag{2.55}$$

$$\mathcal{X}_{kl}^{ij} = \langle ij | F_{12} \hat{Q}_{12} F_{12} | kl \rangle \tag{2.56}$$

$$\mathcal{B}_{kl}^{ij} = \langle ij | F_{12} \hat{Q}_{12} (\hat{f}_1 + \hat{f}_2) \hat{Q}_{12} F_{12} | kl \rangle \ , \tag{2.57}$$

and, in the fixed amplitude approach, only the diagonal direct-type $\mathcal{V}_{ij}^{ij}$, $\mathcal{X}_{mj}^{ij} f_i^m$, $\mathcal{B}_{ij}^{ij}$, and diagonal exchange-type $\mathcal{V}_{ji}^{ij}$, $\mathcal{X}_{mi}^{ij} f_j^m$, $\mathcal{B}_{ji}^{ij}$ terms are needed.

Inserting the definition of $\hat{Q}_{12}$ gives the following expressions for the direct $\mathcal{V}$ and $\mathcal{X}$ terms,

$$\mathcal{V}_{ij}^{ij} = \mathcal{F}\mathcal{G}_{ij}^{ij} + \mathcal{F}_{kl}^{ij} \mathcal{G}_{ij}^{kl} - \mathcal{F}_{ab}^{ij} \mathcal{G}_{ij}^{ab} - \langle ijk | F_{12} r_{23}^{-1} | kji \rangle - \langle jik | F_{12} r_{23}^{-1} | kij \rangle \tag{2.58}$$

$$\mathcal{X}_{mj}^{ij} f_i^m = \big(\mathcal{F}\mathcal{F}_{mj}^{ij} + \mathcal{F}_{kl}^{ij} \mathcal{F}_{mj}^{kl} - \mathcal{F}_{ab}^{ij} \mathcal{F}_{mj}^{ab} - \langle ijk | F_{12} F_{23} | kjm \rangle - \langle jik | F_{12} F_{23} | kmj \rangle \big) f_i^m \ , \tag{2.59}$$

where

$$\mathcal{G}_{rs}^{pq} = \langle pq|r_{12}^{-1}|rs\rangle \tag{2.60}$$
$$\mathcal{F}_{rs}^{pq} = \langle pq|F_{12}|rs\rangle \tag{2.61}$$
$$\mathcal{FG}_{rs}^{pq} = \langle pq|F_{12}r_{12}^{-1}|rs\rangle \tag{2.62}$$
$$\mathcal{FF}_{rs}^{pq} = \langle pq|F_{12}F_{12}|rs\rangle \tag{2.63}$$

The exchange-type expressions are obtained by exchanging the positions of i and j for $\mathcal{V}$ terms or m and j for $\mathcal{X}$ terms in each *ket*. The last two terms in equations (2.58) and (2.59) contain three-electron integrals that result from the single projectors $\hat{\mathrm{o}}_1$ and $\hat{\mathrm{o}}_2$.

The $\mathcal{B}_{ij}^{ij}$ and $\mathcal{B}_{ji}^{ij}$ integrals are by far the most complicated expressions due to the presence of the Fock operators alongside the projectors in $\hat{Q}_{12}$. These terms contain novel integrals over two, three, and four electrons. To derive these integrals, it is first observed that one only needs to consider expressions coming from the $\hat{f}_1$ operator, since the $\hat{f}_2$ terms can be related to them by

$$\langle ij|F_{12}\hat{Q}_{12}\hat{f}_2\hat{Q}_{12}F_{12}|ij\rangle = \langle ji|F_{12}\hat{Q}_{12}\hat{f}_1\hat{Q}_{12}F_{12}|ji\rangle \ , \tag{2.64}$$

which is obtained by simply switching the electron indices $1 \leftrightarrow 2$ and using the symmetry of $\hat{Q}_{12}$ and F_{12}. The operator $\hat{Q}_{12}\hat{f}_1\hat{Q}_{12}$ within $\mathcal{B}_{ij}^{ij}$ contains the follwing terms:

$$\begin{array}{rrrrr}
+\hat{f}_1 & -\hat{f}_1\hat{\mathrm{o}}_1 & -\hat{f}_1\hat{\mathrm{o}}_2 & +\hat{f}_1\hat{\mathrm{o}}_1\hat{\mathrm{o}}_2 & -\hat{f}_1\hat{\mathrm{v}}_1\hat{\mathrm{v}}_2 \\
-\hat{\mathrm{o}}_1\hat{f}_1 & +\hat{\mathrm{o}}_1\hat{f}_1\hat{\mathrm{o}}_1 & +\hat{\mathrm{o}}_1\hat{f}_1\hat{\mathrm{o}}_2 & -\hat{\mathrm{o}}_1\hat{f}_1\hat{\mathrm{o}}_1\hat{\mathrm{o}}_2 & +\hat{\mathrm{o}}_1\hat{f}_1\hat{\mathrm{v}}_1\hat{\mathrm{v}}_2 \\
-\hat{\mathrm{o}}_2\hat{f}_1 & +\hat{\mathrm{o}}_2\hat{f}_1\hat{\mathrm{o}}_1 & +\hat{\mathrm{o}}_2\hat{f}_1\hat{\mathrm{o}}_2 & -\hat{\mathrm{o}}_2\hat{f}_1\hat{\mathrm{o}}_1\hat{\mathrm{o}}_2 & +\hat{\mathrm{o}}_2\hat{f}_1\hat{\mathrm{v}}_1\hat{\mathrm{v}}_2 \\
+\hat{\mathrm{o}}_1\hat{\mathrm{o}}_2\hat{f}_1 & -\hat{\mathrm{o}}_1\hat{\mathrm{o}}_2\hat{f}_1\hat{\mathrm{o}}_1 & -\hat{\mathrm{o}}_1\hat{\mathrm{o}}_2\hat{f}_1\hat{\mathrm{o}}_2 & +\hat{\mathrm{o}}_1\hat{\mathrm{o}}_2\hat{f}_1\hat{\mathrm{o}}_1\hat{\mathrm{o}}_2 & -\hat{\mathrm{o}}_1\hat{\mathrm{o}}_2\hat{f}_1\hat{\mathrm{v}}_1\hat{\mathrm{v}}_2 \\
-\hat{\mathrm{v}}_1\hat{\mathrm{v}}_2\hat{f}_1 & +\hat{\mathrm{v}}_1\hat{\mathrm{v}}_2\hat{f}_1\hat{\mathrm{o}}_1 & +\hat{\mathrm{v}}_1\hat{\mathrm{v}}_2\hat{f}_1\hat{\mathrm{o}}_2 & -\hat{\mathrm{v}}_1\hat{\mathrm{v}}_2\hat{f}_1\hat{\mathrm{o}}_1\hat{\mathrm{o}}_2 & +\hat{\mathrm{v}}_1\hat{\mathrm{v}}_2\hat{f}_1\hat{\mathrm{v}}_1\hat{\mathrm{v}}_2
\end{array} \tag{2.65}$$

By permutating commuting operators, using the idempotency of projectors, the relations

$$\hat{\mathrm{o}}_n\hat{\mathrm{v}}_n = \hat{\mathrm{v}}_n\hat{\mathrm{o}}_n = \hat{\mathrm{o}}_n\hat{f}_n\hat{\mathrm{v}}_n = \hat{\mathrm{v}}_n\hat{f}_n\hat{\mathrm{o}}_n = 0 \ , \tag{2.66}$$

and removing canceling terms in the third and fourth rows, (2.65) simplifies to:

$$\begin{array}{rrrrr}
+\hat{f}_1 & -\hat{f}_1\hat{\mathrm{o}}_1 & -\hat{f}_1\hat{\mathrm{o}}_2 & +\hat{f}_1\hat{\mathrm{o}}_1\hat{\mathrm{o}}_2 & -\hat{f}_1\hat{\mathrm{v}}_1\hat{\mathrm{v}}_2 \\
-\hat{\mathrm{o}}_1\hat{f}_1 & +\hat{\mathrm{o}}_1\hat{f}_1\hat{\mathrm{o}}_1 & +\hat{\mathrm{o}}_1\hat{\mathrm{o}}_2\hat{f}_1 & -\hat{\mathrm{o}}_1\hat{f}_1\hat{\mathrm{o}}_1\hat{\mathrm{o}}_2 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
-\hat{\mathrm{v}}_1\hat{\mathrm{v}}_2\hat{f}_1 & 0 & 0 & 0 & +\hat{\mathrm{v}}_1\hat{f}_1\hat{\mathrm{v}}_1\hat{\mathrm{v}}_2
\end{array} \tag{2.67}$$

Therefore, $\hat{Q}_{12}\hat{f}_1\hat{Q}_{12}$ can be written as

$$\hat{Q}_{12}\hat{f}_1\hat{Q}_{12} = \hat{f}_1 - \hat{f}_1\hat{\mathrm{o}}_2 \tag{2.68}$$
$$+ \hat{T}(-\hat{f}_1\hat{\mathrm{o}}_1 + \hat{f}_1\hat{\mathrm{o}}_1\hat{\mathrm{o}}_2 - \hat{f}_1\hat{\mathrm{v}}_1\hat{\mathrm{v}}_2) \tag{2.69}$$
$$+ \hat{\mathrm{o}}_1\hat{f}_1\hat{\mathrm{o}}_1 - \hat{\mathrm{o}}_1\hat{f}_1\hat{\mathrm{o}}_1\hat{\mathrm{o}}_2 + \hat{\mathrm{v}}_1\hat{f}_1\hat{\mathrm{v}}_1\hat{\mathrm{v}}_2 \ , \tag{2.70}$$

where the linear operator $\hat{T}$ adds the transpose, e.g., $\hat{T}\hat{f}_1\hat{o}_1 = \hat{f}_1\hat{o}_1 + \hat{o}_1\hat{f}_1$. The terms in (2.70) involve only standard Fock matrix elements and evaluate to

$$\langle ij|F_{12}\hat{o}_1\hat{f}_1\hat{o}_1F_{12}|ij\rangle = \langle ijk|F_{12}F_{23}|lji\rangle f_k^l \tag{2.71}$$
$$\langle ij|F_{12}\hat{o}_1\hat{f}_1\hat{o}_1\hat{o}_2F_{12}|ij\rangle = \mathcal{F}_{lm}^{ij}f_k^l\mathcal{F}_{ij}^{km} \tag{2.72}$$
$$\langle ij|F_{12}\hat{v}_1\hat{f}_1\hat{v}_1\hat{v}_2F_{12}|ij\rangle = \mathcal{F}_{ab}^{ij}f_c^a\mathcal{F}_{ij}^{cb}\ . \tag{2.73}$$

The rest of the terms contain non-standard Fock matrix elements and they must be split up into the kinetic energy, nuclear attraction, exchange, and Coulomb operator contributions for exact evaluation. The terms that lead to at most two-electron integrals evaluate to

$$\langle ij|F_{12}\hat{t}_1F_{12}|ij\rangle = \frac{1}{2}\langle ij|(\nabla_1F_{12}\cdot\nabla_1F_{12})|ij\rangle - \frac{1}{2}\langle ij|F_{12}^2|(\Delta i)j\rangle \tag{2.74}$$
$$\langle ij|F_{12}\hat{v}_1F_{12}|ij\rangle = -\sum_A Z_A\langle ij|F_{12}^2r_{1A}^{-1}|ij\rangle \tag{2.75}$$
$$\langle ij|F_{12}\hat{t}_1(\hat{o}_1\hat{o}_2-\hat{v}_1\hat{v}_2)F_{12}|ij\rangle = -\frac{1}{2}\langle ij|F_{12}|(\Delta k)l\rangle\mathcal{F}_{ij}^{kl} + \frac{1}{2}\langle ij|F_{12}|(\Delta a)b\rangle\mathcal{F}_{ij}^{ab} \tag{2.76}$$
$$\langle ij|F_{12}\hat{v}_1(\hat{o}_1\hat{o}_2-\hat{v}_1\hat{v}_2)F_{12}|ij\rangle = -\sum_A\langle ij|F_{12}r_{1A}^{-1}|kl\rangle\mathcal{F}_{ij}^{kl} + \sum_A\langle ij|F_{12}r_{1A}^{-1}|ab\rangle\mathcal{F}_{ij}^{ab}\ . \tag{2.77}$$

Up to three-electron integrals result for the follwing terms:

$$\langle ij|F_{12}\hat{j}_1F_{12}|ij\rangle = \langle ijk|F_{12}^2r_{13}^{-1}|ijk\rangle \tag{2.78}$$
$$\langle ij|F_{12}\hat{k}_1F_{12}|ij\rangle = \langle ijk|F_{12}F_{23}r_{13}^{-1}|kji\rangle \tag{2.79}$$
$$\langle ij|F_{12}\hat{t}_1\hat{o}_1F_{12}|ij\rangle = -\frac{1}{2}\langle ijk|F_{12}F_{23}|(\Delta k)ji\rangle \tag{2.80}$$
$$\langle ij|F_{12}\hat{v}_1\hat{o}_1F_{12}|ij\rangle = -\sum_A Z_A\langle ijk|F_{12}F_{23}r_{1A}^{-1}|kji\rangle \tag{2.81}$$
$$\langle ij|F_{12}\hat{t}_1\hat{o}_2F_{12}|ij\rangle = -\frac{1}{2}\langle ijk|F_{12}\Delta_1F_{13}|ikj\rangle \tag{2.82}$$
$$= -\frac{1}{2}\langle ijk|F_{12}(\Delta_1F_{13})|ikj\rangle \tag{2.83}$$
$$-\langle ijk|F_{12}\nabla_1F_{13}\cdot|(\nabla i)kj\rangle \tag{2.84}$$
$$-\frac{1}{2}\langle ijk|F_{12}F_{13}|(\Delta i)kj\rangle \tag{2.85}$$
$$\langle ij|F_{12}\hat{v}_1\hat{o}_2F_{12}|ij\rangle = -\sum_A Z_A\langle ijk|F_{12}F_{13}r_{1A}^{-1}|ikj\rangle \tag{2.86}$$
$$\langle ij|F_{12}\hat{j}_1(\hat{o}_1\hat{o}_2-\hat{v}_1\hat{v}_2)F_{12}|ij\rangle = \langle ijm|F_{12}r_{13}^{-1}|klm\rangle\mathcal{F}_{ij}^{kl} + \langle ijk|F_{12}r_{13}^{-1}|abk\rangle\mathcal{F}_{ij}^{ab} \tag{2.87}$$
$$\langle ij|F_{12}\hat{k}_1(\hat{o}_1\hat{o}_2-\hat{v}_1\hat{v}_2)F_{12}|ij\rangle = \langle ijm|F_{12}r_{13}^{-1}|mlk\rangle\mathcal{F}_{ij}^{kl} + \langle ijk|F_{12}r_{13}^{-1}|kba\rangle\mathcal{F}_{ij}^{ab}\ . \tag{2.88}$$

The remaining four terms lead to four-electron integrals

$$\langle ij|F_{12}\hat{\jmath}_1\hat{\mathrm{o}}_1F_{12}|ij\rangle = \langle ijkl|F_{12}F_{23}r_{14}^{-1}|kjil\rangle \tag{2.89}$$

$$\langle ij|F_{12}\hat{k}_1\hat{\mathrm{o}}_1F_{12}|ij\rangle = \langle ijkl|F_{12}F_{23}r_{14}^{-1}|ljik\rangle \tag{2.90}$$

$$\langle ij|F_{12}\hat{\jmath}_1\hat{\mathrm{o}}_2F_{12}|ij\rangle = \langle ijkl|F_{12}F_{13}r_{14}^{-1}|ikjl\rangle \tag{2.91}$$

$$\langle ij|F_{12}\hat{k}_1\hat{\mathrm{o}}_2F_{12}|ij\rangle = \langle ijkl|F_{12}F_{34}r_{14}^{-1}|lkji\rangle. \tag{2.92}$$

Many of these integrals also arise in the Hylleraas-CI method and their calculation within this context has been described in Refs. [49–52] in terms of a linear ansatz $F_{12} = r_{12}$. Their calculation for general correlation factors is also discussed in Refs. [43, 53–61]. Due to their high cost, the three- and four-electron integrals are typically approximated in some form or another, see Section 2.5.

2.2.4 Explicitly Correlated Coupled Cluster Theory

The application of the R12/F12 approach to coupled cluster theory (CC-F12) [62–65] is achieved by including excitations into explicitly correlated configurations. The wave function ansatz is given by

$$\Phi_{\mathrm{CC-F12}} = \exp(\hat{T}^{(n)}_{\mathrm{CC-F12}})\Phi_{\mathrm{HF}} \ , \qquad \hat{T}^{(n)}_{\mathrm{CC-F12}} = \hat{R} + \hat{T}_1 + \hat{T}_2 + \ldots + \hat{T}_n \ , \tag{2.93}$$

with the excitation operators $\hat{T}_1$, $\hat{T}_2$, etc., defined as in regular CC theory (see eq. (2.22)) and the explicitly correlated excitation operator $\hat{R}$ defined as

$$\hat{R} = \sum_{\substack{i<j \\ x<y}} c_{xy}^{ij}\hat{\gamma}_{ij}^{xy} \ , \tag{2.94}$$

where the c_{xy}^{ij} are the geminal amplitudes and the $\hat{\gamma}_{ij}^{xy}$ describe two-electron excitations into orthogonalized geminal configurations:

$$\hat{\gamma}_{ij}^{xy}|ij\rangle = \hat{Q}_{12}F_{12}|(\hat{S}_{xy}xy)\rangle \tag{2.95}$$

One proceeds as in regular CC by formulating the similarity transformed Schrödinger equation

$$\exp(-\hat{T}^{(n)}_{\mathrm{CC-F12}})\hat{H}\exp(\hat{T}^{(n)}_{\mathrm{CC-F12}})|\Phi_{\mathrm{HF}}\rangle = E_{\mathrm{CC-F12}}|\Phi_{\mathrm{HF}}\rangle \ , \tag{2.96}$$

and projecting onto conventional singly, doubly, etc., excited determinants, as well as onto the linear geminal functions produced by $\hat{\gamma}_{ij}^{xy}|\Phi_{\mathrm{HF}}\rangle$, leading to equations for determining the excitation and geminal amplitudes. In practice, the most common choice is $n = 2$ so that the $\hat{T}$ operator is restricted to single and double excitations (CCSD-F12). In CCSD-F12 theory, in addition to the V, X, C, and B intermediates needed in MP2-F12 theory, additional intermediates result that are of similar complexity to the B intermediate. They can be factorized using the methods in Section 2.3 and screened using the bounds developed

in **Publication II**. However, there exist various very good approximations to full CCSD-F12 that only require the intermediates of MP2-F12 theory [66–71]. As in MP2-F12 theory, the use of Ten-no's fixed amplitudes leads to great savings and is standard practice [6], and the approximate CC-F12 methods show the same scaling with the system size as the base CC method. In contrast to MP2-F12, the cost of the CC-F12 calculations is not significantly higher than the base coupled cluster calculation, making the F12 correction in this context even more attractive. Due to the large overlap between the F12 corrections of MP2 and CC theories, the methods aimed at reduction of the scaling of the F12 correction presented in this thesis are also directly applicable to CC-F12 methods.

2.3 Approximations for Reduced Scaling

The overarching goal of quantum chemistry, to accurately characterize and simulate chemical systems of practical interest, requires the development of computational methods that exhibit low scaling with the size of the system M. M is used here as an abstract measure that generally corresponds to the number of atoms in a system, but not always to the number of basis functions used, since the latter may increase without adding more atoms by simply choosing a larger basis set. However, the scaling with the size of the basis set is also an important property that can be reduced by the integral factorizations introduced in this section, but not through integral screening, which relies on increasing spatial distances between basis functions.

The calculation of the four-electron terms that arise in MP2-F12 theory scales formally as $\mathcal{O}(M^8)$ in the AO basis, which is formidable considering that even the formally quartic scaling of the uncorrelated HF method (Sec. 2.1.1) can become intractable for moderately sized molecules if integral screening or other approximations are not used. The root problem of steeply scaling methods is that the size of systems that can be treated increases unacceptably slowly with increases in either compute time, memory size, or processor number and speed, so that improvements in computational resources alone cannot be relied upon to extend the reach of quantum chemistry. To overcome the resulting computational wall, the scaling of most methods can be reduced by using accurate approximations and additionally avoiding the calculation of negligible contributions to molecular energies through integral and density based screening. In many cases, the asymptotic scaling can be reduced to linear using these methods.

2.3.1 Resolution-of-the-Identity Approximation

The identity operator $\hat{1}$ on the space of one-electron functions can be resolved exactly, with respect to an electron n, by the CBS projector

$$\hat{\alpha}_n \, g(\ldots, \boldsymbol{r}_n, \ldots) = \sum_{\alpha} \alpha(\boldsymbol{r}_n) \Big(\int \alpha(\boldsymbol{r}_n) g(\ldots, \boldsymbol{r}_n, \ldots) d\boldsymbol{r}_n \Big) \, . \tag{2.97}$$

In the resolution-of-the-identity approximation (RI) [72–76], $\hat{\alpha}_n$ is approximated as the

projector onto the subspace spanned by a finite, orthonormal set of functions. The most straightforward way to construct a suitable projector is to orthogonalize a large auxiliary basis set $\{P\}$ to obtain an orthogonal set $\{P_o\}$, e.g., through Löwdin orthogonalization [13]:

$$P_o(\boldsymbol{r}) = [\boldsymbol{S}^{-\frac{1}{2}}]_Q^P Q(\boldsymbol{r}) \ , \tag{2.98}$$

where $\boldsymbol{S}$ is the overlap matrix in the auxiliary basis set.

This leads to the approximation

$$\hat{\alpha}_n \approx \sum_{P_o} |P_o)(P_o| = \sum_{PQ} |P)(PQ)^{-1}(Q| \ , \tag{2.99}$$

where the shorthand notation $(PQ)^{-1} = [\boldsymbol{S}^{-1}]_Q^P$ is used.

For a multiplicative two-electron operator

$$m : \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow [0, \infty], \ (\boldsymbol{r}_1, \boldsymbol{r}_2) \mapsto m(\boldsymbol{r}_1, \boldsymbol{r}_2) \ , \tag{2.100}$$

the corresponding two-electron integral

$$(\mu\nu|m|\lambda\sigma) = \int\int \chi_\mu(\boldsymbol{r}_1)\chi_\nu(\boldsymbol{r}_1)m(\boldsymbol{r}_1, \boldsymbol{r}_2)\chi_\lambda(\boldsymbol{r}_2)\chi_\sigma(\boldsymbol{r}_2)d\boldsymbol{r}_1 d\boldsymbol{r}_2 \ , \tag{2.101}$$

can be approximated through insertion of (2.99):

$$(\mu\nu|m|\lambda\sigma) = (\mu\nu\hat{\alpha}_1|m|\hat{\alpha}_2\lambda\sigma) \approx \sum_{PQRS} (\mu\nu P)(PQ)^{-1}(Q|m|R)(RS)^{-1}(S\lambda\sigma) \ . \tag{2.102}$$

Given a second two-electron operator n defined analogously to m and known as the RI metric, one can, starting from $\{P\}$, define a new set of functions $\{P_n\}$ as

$$P_n(\boldsymbol{r}) = \int P(\boldsymbol{r}')n(\boldsymbol{r}', \boldsymbol{r})d\boldsymbol{r}', \tag{2.103}$$

which, for certain n, can also be used as an auxiliary basis set for performing RI [77]. Using this set (2.102) can be generalized to

$$(\mu\nu|m|\lambda\sigma) \approx \sum_{PQRS} (\mu\nu|n|P)(P|n|Q)^{-1}(Q|m|R)(R|n|S)^{-1}(S|n|\lambda\sigma) \ , \tag{2.104}$$

which has the potential to be more accurate than (2.102). Here again, the shorthand notation $(P|n|Q)^{-1}$ is used to refer to elements of the inverse of the matrix with elements $(P|n|Q)$.

Through the closely related method of *density fitting* (DF) [78], one can show that, if m can be used as the RI metric, then the most accurate approximation of the type (2.104) is given by setting $n = m$, which leads to the most widely used RI expression

$$(\mu\nu|m|\lambda\sigma) \approx \sum_{PQ} (\mu\nu|m|P)(P|m|Q)^{-1}(Q|m|\lambda\sigma) \ . \tag{2.105}$$

The form (2.105) is available for all of the two-electron operators that appear in this work and is the preferred expression from an accuracy standpoint when factorizing two-electron integrals. However, for slowly decaying operators such as the Coulomb operator r_{12}^{-1}, it has been shown that the overall efficiency of the RI factorization can be increased by using more quickly decaying RI metrics, see, e.g., Refs. [79, 80]. This increases the sparsity in the three-center integral tensors of (2.104), which can be used to increase the speed of numerical implementations at the cost of slight reductions in accuracy.

In the following, the acronym RI is used to denote the insertion of a completeness relation in the form of (2.99), i.e., with respect to the overlap metric $\delta(|\boldsymbol{r}_1 - \boldsymbol{r}_2|)$, and the basis used to represent the RI will be assumed to be orthogonal. The acronym DF is used to denote the factorization of a two-electron integral given by the formula (2.105), with the usual choice of the RI metric being the one that is optimal in the sense of accuracy, i.e., equation (2.105).

2.3.2 Real Space Numerical Quadrature

An alternative to RI for factorizing molecular integrals is provided by three-dimensional numerical quadrature (QD). QD has a long history of use in density functional theory (DFT) methods, where analytical integration is usually not feasible. This has led to the development of efficient molecular grids [81–86], which consist of a set of three-dimensional points $\boldsymbol{r}_g$ with corresponding weights w_g, such that the integration over a function $f : \mathbb{R}^3 \to \mathbb{R}$ (such as the electronic density, or spatial orbitals) can be approximated as

$$\int f(\boldsymbol{r}) d\boldsymbol{r} \approx \sum_g w_g f(\boldsymbol{r}_g) \ , \tag{2.106}$$

where the accuracy of the approximation depends on the number of grid points N_g and their positions and weights. In general, the approximation can be made as accurate as desired by adding more grid points at the price of increased computational demand.

To factorize the molecular integrals that occur in the wave function based methods of Section 2.1, one can, e.g., set

$$f(\boldsymbol{r}_1) = \chi_\mu(\boldsymbol{r}_1)\chi_\lambda(\boldsymbol{r}_1) \int \chi_\nu(\boldsymbol{r}_2)\chi_\sigma(\boldsymbol{r}_2) r_{12}^{-1} d\boldsymbol{r}_2, \tag{2.107}$$

in (2.106), which leads to the approximation

$$(\mu\lambda|\nu\sigma) \approx w_g \chi_\mu^g \chi_\lambda^g \mathcal{G}_{\nu\sigma}^g \ , \tag{2.108}$$

where $\chi_\mu^g = \chi_\mu(\boldsymbol{r}_g)$ and

$$\mathcal{G}_{\nu\sigma}^g = \int \frac{\chi_\nu(\boldsymbol{r})\chi_\sigma(\boldsymbol{r})}{|\boldsymbol{r} - \boldsymbol{r}_g|} d\boldsymbol{r} \ . \tag{2.109}$$

The one-electron potential integrals $\mathcal{G}_{\nu\sigma}^g$ can be computed analytically and they are already required for the calculation of the nuclear electronic attraction term of the fixed nucleus

Hamiltonian, where instead of the potential at the grid points one needs the potential at the atomic centers. This type of factorization can be used to greatly reduce the scaling of molecular energy formulas and is used extensively in Sections 2.4 and 2.5.

2.3.3 Integral Screening

The methods presented in Sections 2.1 and 2.2 expand the wave function using products of molecular orbitals, which are themselves optimized linear combinations of fixed atomic orbitals. This leads to the need to calculate different types of interactions between a very large number of comparatively simple charge distributions. The prototypical interaction is the two-electron repulsion, which can be broken down into the direct E^D and exchange E^X energies:

$$E^D = (ii|jj) = P_{\mu\nu}J_{\mu\nu}, \qquad E^X = (ij|ij) = P_{\mu\nu}K_{\mu\nu} \ , \tag{2.110}$$

where

$$J_{\mu\nu} = P_{\lambda\sigma}(\mu\nu|\lambda\sigma), \qquad K_{\mu\nu} = P_{\lambda\sigma}(\mu\lambda|\nu\sigma) \ , \tag{2.111}$$

are the Coulomb and exchange matrices, respectively, and $P_{\mu\nu}$ is the one-particle density matrix.

In non-explicitly correlated theories, the electron repulsion integrals (ERIs) $(\mu\nu|\lambda\sigma)$ are the most computationally intensive integrals to calculate and also appear in the greatest number with a formal scaling of $\mathcal{O}(N_\mu^4)$ total integrals. Consequently, significant time and research has been devoted to improving the speed of their calculation [87–101]. In addition to the complexity of the ERIs, their large number means that prohibitively large amounts of computer memory would be required to store their values for reuse in the iterative HF scheme. This has led to the prevalence of direct SCF methods [102, 103], in which the ERIs are recalculated on-the-fly in each SCF iteration. In order to reduce the computational demands of these algorithms, estimates and bounds are used to avoid the exact calculation of integrals that are approximately zero, i.e., smaller than some given threshold ϑ. Initially, non-rigorous estimates were employed [102, 104], but they were quickly replaced by the very efficient and rigorous Schwarz bound [105]:

$$|(\mu\nu|\lambda\sigma)| \leqslant (\mu\nu|\mu\nu)^{1/2}(\lambda\sigma|\lambda\sigma)^{1/2} = Q_{\mu\nu}Q_{\lambda\sigma} \ . \tag{2.112}$$

This bound is a consequence of the fact that the ERI defines an inner product on the vector space of one-electron charge distributions [12, 72].

The usefulness of (2.112) derives from the local nature of Gaussian basis functions, which vanish at infinity and are in the Schwartz space of rapidly decreasing functions [106], so that as system sizes increase, only a linearly scaling number of significantly overlapping basis function pairs results, and products of non-overlapping basis functions vanish, as do integrals over them. This overlap decay is captured within the diagonal integrals $(\mu\nu|\mu\nu)$, which can be used to determine which basis function pairs never contribute, based on the size of $(\mu\nu|\mu\nu)^{1/2}$ alone, and which basis function quartets are negligible, based on the size of $(\mu\nu|\mu\nu)^{1/2}(\lambda\sigma|\lambda\sigma)^{1/2}$. The integrals $(\mu\nu|\mu\nu)$ depend on only two basis functions and can therefore be easily precomputed and stored for use in screening algorithms.

The overlap decay captured by the Schwarz bound (2.112) reduces the number of asymptotically non-negligible integrals to $\mathcal{O}(N_\mu^2)$. Due to the decay of the Coulomb operator r_{12}^{-1} at large distances, the number of significant integrals is actually asymptotically linear. This operator mediated distance decay is not captured by the Schwarz bound, but because r_{12}^{-1} decays so slowly with increasing distances, linear scaling would only be seen for systems significantly larger than can be treated today and the Schwarz bound is usually adequate in this case. However, in some cases, the operator decay can become very important. The multipole expansion of r_{12}^{-1} can be used to show that for charge distributions with vanishing low-order multipole moments, the operator decay increases. This is an important aspect in linear-scaling formulations of correlated theories such as MP2, where transformation of the charge distributions leads to vanishing monopoles [107]. In addition, single atomic orbitals have vanishing multipoles up to the order of their total angular momentum, leading to faster distance decay for the three-center integrals over r_{12}^{-1} that occur when using density fitting [108]. The first bound to successfully address distance decay was developed by Lambrecht, Doser, and Ochsenfeld [109, 110]. This bound is based directly on the multipole expansion of r_{12}^{-1}, and introduced the concept of charge distribution extents based on the convergence of the multipole expansion. Later, it was shown that a simpler and more effective approach could be obtained through a tight estimate [107, 111], which is not a strict upper bound but retains the extent-based form of the rigorous multipole-based bound. This so-called QQR estimate is a product of the Schwarz bound factors $Q_{\mu\nu}$, which capture the overlap decay, and a distance factor $R'^{-(1+m)}$, where m is the number of vanishing leading multipoles in the *bra* and *ket*, and R' is the distance between the two spheres defined by the charge distribution centers and extents, which are computed in the same way as in the continuous fast multipole method (CFMM) [112, 113]. The QQR estimate was shown to work very well and was used later in a more general way to estimate integrals over the operator $\mathrm{erfc}(0.11 r_{12})$ [114]. Here, the distance factor is simply replaced by the function $\mathrm{erfc}(0.11 R')$, while the Schwarz bound factors and CFMM extents are retained. In this context, the QQR estimate still performs very well, however, this changes for other operators as outlined below.

In **Publication I**, it is shown that the generalized QQR estimate becomes increasingly inaccurate as the rate of operator decay increases, leading to unacceptable errors for the very short-ranged operators encountered in explicitly correlated theories. In this same work, a new Schwarz-type bound is introduced that rigorously and effectively captures the distance decay of such operators without the use of distribution centers or extents. While the ability to effectively apply screening methods to explicitly correlated theories was a breakthrough, some drawbacks of the bound were quickly apparent. For one, it can only be applied effectively to two-electron four-center integrals and is therefore not useful in combination with the RI, DF, and QD introduced above, or for explicitly correlated methods with exact calculation of three- and four-electron integrals. In addition, the bound can fail to correctly discard integrals that are negligible due to roughly equal contributions of overlap and distance decay, and instead performs best when either overlap or distance decay individually lead to negligibility. The latter problem was addressed by introducing a non-rigorous integral estimate based on the bound in the same publication, but the estimate

has similar downsides for very-short ranged operators as the QQR estimate. Both problems are resolved in **Publication II**, where very general bounds for molecular integrals involving any number of electrons and all relevant combinations of integral operators are developed. The ability to treat many-electron integrals makes these bounds especially attractive for use in highly accurate explicitly correlated approaches such as Hylleraas-CI, where exact evaluation of many-electron integrals is more common.

The integral partition bounds (IPB's) developed in **Publication II** are based on the intuition gained from the earlier extent-based bounds and estimates, but rely on previously unused mathematical techniques to ensure formulas that are rigorous and tight for any of the two-electron operators used in quantum chemistry, and combinations thereof. While previous extents were determined based on the convergence of the multipole expansion of r_{12}^{-1}, IPB extents are generated in a much more direct fashion by explicitly partitioning the multi-electron integration space. The fundamental ideas behind the IPB's are detailed in the following.

The goal of the bounding procedure is to take advantage of the fast radial decay of the charge distributions Ω_i used to describe one-electron states. This radial decay remains for the corresponding absolute charge distributions $\overline{\Omega}_i = |\Omega_i|$ and it is therefore sufficient to develop bounds for absolute charge distributions.

Given a set of N absolute charge distributions $\{\overline{\Omega}_1, \ldots, \overline{\Omega}_N\}$, one can define distribution centers $\{\boldsymbol{C}_1, \ldots, \boldsymbol{C}_N\}$. The distribution centers can be chosen arbitrarily, but their positions have practical consequences for the sizes of the extents and the tightness of the resulting bounds (see **Publication II** and Chapter 3). The next step is to partition the integration spaces around each distribution into core regions – which are balls, centered at the distribution centers, with variable radii $\{R_1, \ldots, R_N\}$ – and tail regions, which are the spaces outside of these balls. A two-dimensional depiction of the partitioning of the integration space for an absolute charge distribution $\overline{\Omega}_i$ is shown in Fig. 2.2.

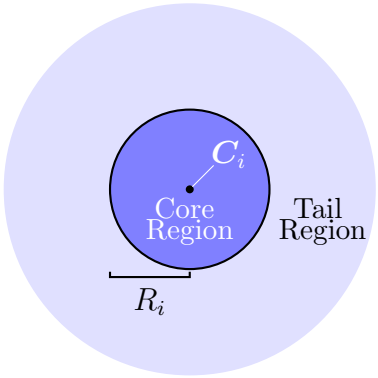


Figure 2.2: Depiction of the spherical partitioning of an absolute charge distribution $\overline{\Omega}_i$ into core and tail regions.

A depiction of an absolute two-electron integral within the IPB framework is shown in Fig. 2.3a. Because integrals are linear maps with regards to the integration space, one can represent a multi-electron integral exactly as the sum of a single integral over core regions

only, the principal integral, and further auxiliary integrals that have the core region removed for one electron. This exact partitioning into the principal and auxiliary integrals is shown for an absolute two-electron integral in Fig. 2.3b. For practical reasons, it is beneficial to use, instead of the exact partitioning, an upper bound partitioning that results in auxiliary integrals that are congruent in the sense that the have the same general form. This is obtained by adding any missing tail regions to the distributions of the auxiliary integrals of the exact partitioning. For the general n-electron integral, the resulting partitioning is the sum of the principal integral and n tail integrals describing the interaction of the tail region of one distribution with the other $(n-1)$ absolute charge distributions involved. A depiction of this upper bound partitioning for a two-electron integral is shown in Fig. 2.3c. In practice, the overestimation that occurs through the use of the upper bound partitioning scheme is insignificant, since the distribution radii are optimized such that the added tail regions become negligible.

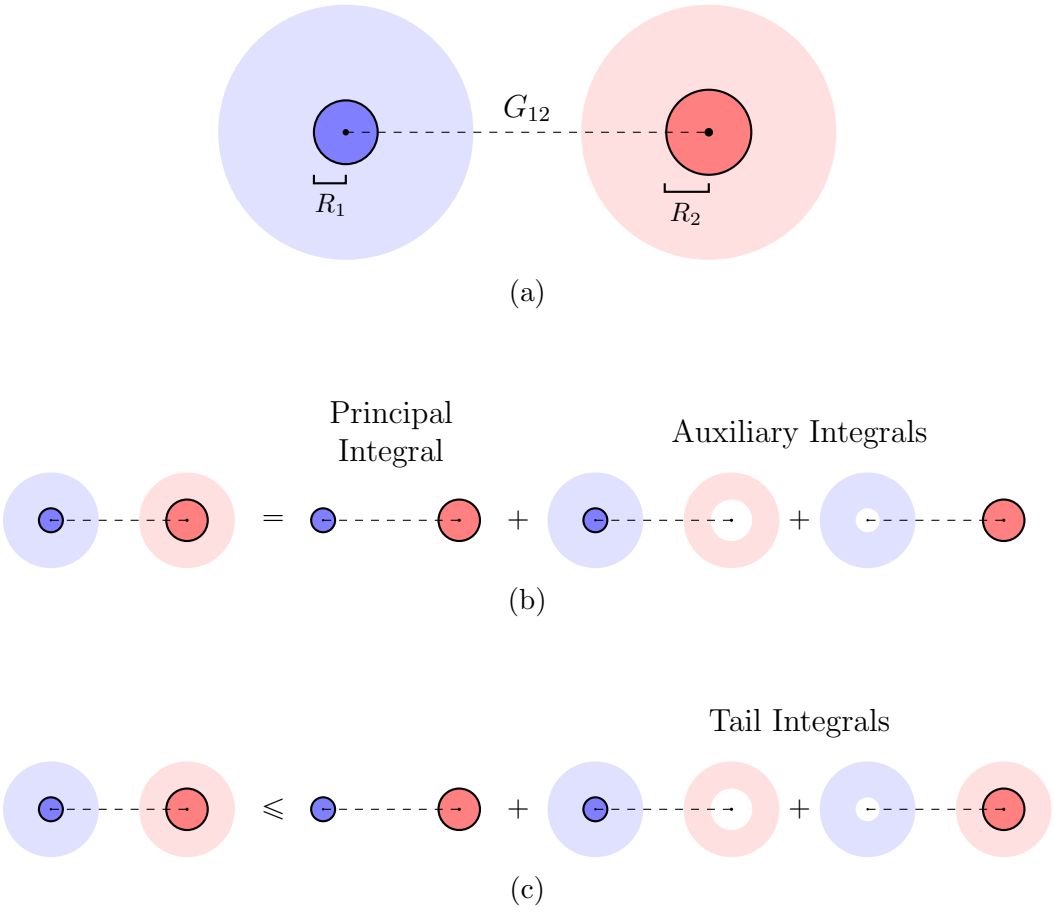


Figure 2.3: (a) Depiction of an absolute two-electron integral involving two absolute charge distributions interacting through the operator G_{12}. (b) Exact partitioning of absolute two-electron integral into principal integral and auxiliary integrals. (c) Upper bound partitioning of absolute two-electron integral in principal integral and tail integrals.

A large advantage of the upper bound partitioning is its generality. The partitioning of an absolute one-electron potential integral, which describes the interaction of a charge distribution with a point charge, is shown in Fig. 2.4a. In this particular case, the partitioning is exact. The partitioning for an absolute three-electron integral is shown in Fig. 2.4b.

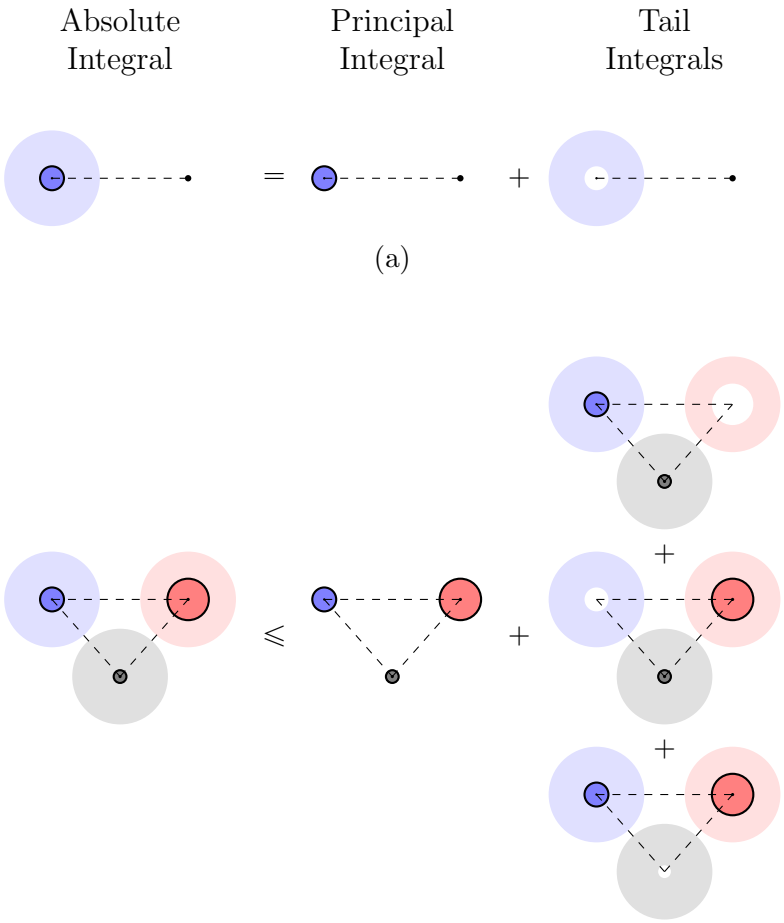


Figure 2.4: Depiction of the IPB partitioning procedure for absolute (a) one-electron potential and (b) three-electron interaction integrals.

In the general n-electron case, the tail integrals provide a natural starting point for the determination of rigorous extents, since they individually depend on the radius of only one distribution and can be made negligible by increasing this radius, because the charge distributions vanish at infinity. Given an integral threshold ϑ, if each tail integral in the partitioning is smaller than $\vartheta/(n+1)$, then the the entire integral is insignificant if the principle integral is also smaller than $\vartheta/(n+1)$. Using the symbol $T^i_{k_1,\cdots k_{(n-1)}}(R_i)$ to denote the tail integral that describes the interaction of the tail region of $\overline{\Omega}_i$ with the charge distributions $\overline{\Omega}_{k_1},\ldots,\overline{\Omega}_{k_{(n-1)}}$, the condition for determining distribution extents, i.e., the

R_i that ensure that tail integrals are negligible, is

$$\max_{k_1,\ldots,k_{(n-1)}} T^i_{k_1\cdots k_{(n-1)}}(R_i) \leqslant \vartheta/(n+1) \; . \tag{2.113}$$

This condition leads to rigorous extent equations for each charge distribution, independent of all others. The solution of these equations is no trivial task from a mathematical perspective and is a large part of the work in **Publication II**, where efficient numerical procedures are developed for all integral types of interest involving AO based charge distributions. Once the extents have been determined, only the principal integrals over bounded core regions must be taken into account for screening purposes. The bounded nature of the integration spaces within the principal integrals makes it possible to estimate them from above by simple, separable formulas that take both overlap and distance dependence into account.

2.3.4 Density Screening

For many classes of molecules, the electronic structure itself is a source of sparsity that can be used to reduce the time-complexity of computational algorithms. While the occupied molecular orbitals obtained from canonical Hartree-Fock theory are typically spread out throughout the molecule, the one-particle density matrix $\boldsymbol{P}$, defined through the expansion coefficients of the MOs in the AOs, $C_{\mu p}$, as

$$P_{\mu\nu} = C_{\mu i} C_{\nu i} \tag{2.114}$$

is often a sparse matrix, where spatially distant atomic orbital pairs correspond to vanishing density matrix elements. Taking advantage of this sparsity can lead to early onset asymptotic linear scaling of expressions such as the exchange matrix.

In **Publication II**, a variety of screening schemes which take density matrix sparsity into account are given for the different types of integrals that result in explicitly correlated and other methods. These schemes can easily be adapted to treat the three-center integrals and semi-numerical quantities that result from the use of RI, DF, and QD.

Screening the density matrix directly requires the formulation of energy expressions either partially or completely in the AO basis. While this has the advantage of increased sparsity, it can lead to inefficiencies for midsize and large basis sets, since the number of occupied orbitals is then much smaller than the number of AOs. This problem can be addressed by using localized MOs, which result from unitary transformations of canonical MOs. When combined with MO based integral bounds, this allows one to retain the low scaling nature of AO methods, while also reducing the dependence of the expressions on the size of the basis set. The most efficient localization scheme for large systems is obtained through Cholesky decomposition of the density matrix [115], which can be performed in a linear scaling, non-iterative fashion [116, 117]. The result is a lower-triangular coefficient matrix $\boldsymbol{L}$ such that

$$\boldsymbol{P} = \boldsymbol{L}\boldsymbol{L}^t, \tag{2.115}$$

where $\boldsymbol{L}$ retains the sparsity of the density matrix and defines an orthonormal set of MOs denoted Cholesky MOs.

In order to take advantage of such localized MOs, it is necessary to find upper bounds for molecular integrals containing them. In **Publication II**, the IPB's are developed for atomic orbitals, where the equations that arise can be largely calculated analytically. When dealing with MO-type orbitals, i.e., expansions of AOs, a numerical approach is required to ensure that the calculated extents and overlap factors lead to tight bounds. The formulas necessary in these cases are given in Chapter 3.

2.4 Semi-Numerical Hartree-Fock Exchange

The use of numerical quadrature is the basis for the semi-numerical calculation of the exchange matrix [118–127]

$$K_{\mu\nu} = P_{\lambda\sigma}(\mu\lambda|\nu\sigma). \tag{2.116}$$

The calculation of the exchange matrix scales formally as $\mathcal{O}(M^4)$, and, while the asymptotic scaling with the size of the system can be reduced to linear using integral and density based screening methods [128–130], the increases in computational demands when increasing the size of the basis set remain high. QD can be used to approximate the exchange matrix using the following formula:

$$\tilde{K}_{\mu\nu} = P_{\lambda\sigma} w_g \chi^g_\mu \chi^g_\lambda \mathcal{G}^g_{\nu\sigma} \ , \tag{2.117}$$

This expression can be calculated in three steps,

$$F^g_\sigma = \chi^g_\lambda P_{\lambda\sigma} \tag{2.118}$$

$$G^g_\nu = \mathcal{G}^g_{\nu\sigma} F^g_\sigma \tag{2.119}$$

$$\tilde{K}_{\mu\nu} = w_g \chi^g_\mu G^g_\nu \ , \tag{2.120}$$

each of which scales formally as $\mathcal{O}(N_g N_\mu^2)$. Because $\tilde{K}_{\mu\nu}$ is not exactly symmetric due to the asymmetry of the QD factorization, a symmetric approximation of $K_{\mu\nu}$ is obtained through the use of:

$$K_{\mu\nu} \approx \frac{\tilde{K}_{\mu\nu} + \tilde{K}_{\nu\mu}}{2} \ . \tag{2.121}$$

The improved formal scaling ($\mathcal{O}(M^3)$) of this method, combined with only quadratic scaling in the size of the basis set, leads to much improved computational efficiency for large basis sets and/or molecules compared to exchange matrix calculation with analytic integral evaluation. In addition, screening based on the integrals $\mathcal{G}^g_{\nu\sigma}$ and density matrix elements leads to a very efficient, asymptotically linear scaling method. However, bounds on the values of $\mathcal{G}^g_{\nu\sigma}$ were initially unavailable. For this reason, heuristic estimates were used, which, although effective in most cases, have drawbacks in reliability and flexibility [121, 127]. To remedy this problem, in **Publication IV**, the general and rigorous bounds developed in **Publication II** are applied to the screening of the $\mathcal{G}^g_{\nu\sigma}$ integrals, leading to a

rigorous screening procedure. In addition, the improved simplicity of the bounds allows for an efficient implementation of the screening routines on graphic processing units (GPUs), leading to further large gains in performance.

2.5 Reduced Scaling MP2-F12

RI Integral Factorization

A big reason for the success of Kutzelnigg and Klopper's original R12 method [7, 8] was the use of RI to break down the many-electron integrals into products of two-electron integrals. For example, inserting equation (2.99) into the three-electron integrals contained in the V_{ij}^{ij} intermediate gives the following approximation:

$$\langle ijk|\frac{F_{12}}{r_{23}}|kji\rangle = \langle ijk|F_{12}\hat{\alpha}_2\frac{1}{r_{23}}|kji\rangle \approx \langle ij|F_{12}|kP_o\rangle\langle kP_o|r_{12}^{-1}|ij\rangle = \mathcal{F}_{kP_o}^{ij}\mathcal{G}_{ij}^{kP_o} \ . \tag{2.122}$$

This is equivalent to first representing the $\hat{Q}_{12}$ projector as

$$\hat{Q}_{12} = 1 - \hat{\mathrm{v}}_1\hat{\mathrm{v}}_2 + \hat{\mathrm{o}}_1\hat{\mathrm{o}}_2 - \hat{\mathrm{o}}_1\hat{\alpha}_2 - \hat{\alpha}_1\hat{\mathrm{o}}_2 \ , \tag{2.123}$$

and then resolving $\hat{\alpha}_n$ approximately using the finite set $\{P_o\}$.

In the initial work on R12 methods, the finite set used to approximate the CBS was simply chosen as the MO basis itself [8, 62]. Later, efficiency was improved by employing a larger auxiliary basis set to construct the RI [131], allowing for better accuracy without the need to increase the size of the HF orbital basis set. Today, the most popular RI scheme is Valeev's *complementary auxiliary basis set* (CABS) method [14], in which the RI is performed using the union $\{p'\}$ of the HF orbitals $\{p\}$ and the CABS basis set $\{p''\}$. The corresponding projectors are denoted $\hat{\mathrm{p}}'_n$, $\hat{\mathrm{p}}_n$, and $\hat{\mathrm{p}}''_n$, respectively. Using $\hat{\alpha}_n \approx \hat{\mathrm{p}}'_n$ in (2.123) leads to the approximation

$$\hat{Q}_{12} \approx 1 - \hat{\mathrm{p}}_1\hat{\mathrm{p}}_2 - \hat{\mathrm{o}}_1\hat{\mathrm{p}}''_2 - \hat{\mathrm{p}}''_1\hat{\mathrm{o}}_2 \ , \tag{2.124}$$

and further use of RI in the form of $1 = \hat{\mathrm{p}}'_1\hat{\mathrm{p}}'_2$ gives

$$\hat{Q}_{12} \approx \hat{\mathrm{p}}''_1\hat{\mathrm{p}}''_2 + \hat{\mathrm{v}}_1\hat{\mathrm{p}}''_2 + \hat{\mathrm{p}}''_1\hat{\mathrm{v}}_2 \ . \tag{2.125}$$

In the CABS method, instead of using (2.123) and resolving the complete one-electron space with an auxiliary basis set, either (2.124) or (2.125) is employed, and one approximately resolves the projector $\hat{\alpha}'_n = \hat{\alpha}_n - \hat{\mathrm{p}}_n$ directly using the CABS basis set. The version (2.124) is used when the unity operator leads to tractable terms, while (2.125), which cannot be approximated as accurately due to the double RI insertion required for the first term, is used to avoid the unity operator. The construction of the CABS basis set is detailed in Ref. [14]. The utility of the CABS method is that more compact formulas are obtained and errors due to inexact cancellations that could arise in the last three terms of equation

(2.123) are avoided. Such errors would occur when using an approximation to $\hat{\alpha}_n$ that does not sufficiently span the HF orbital space.

For the $\mathcal{V}_{ij}^{ij}$ and $\mathcal{X}_{mj}^{ij} f_i^m$ terms, equation (2.124) is used to obtain the working equations

$$\mathcal{V}_{ij}^{ij} = \mathcal{FG}_{ij}^{ij} - \mathcal{F}_{kp''}^{ij}\mathcal{G}_{ij}^{kp''} - \mathcal{F}_{p''k}^{ij}\mathcal{G}_{ij}^{p''k} - \mathcal{F}_{pq}^{ij}\mathcal{G}_{ij}^{pq} \tag{2.126}$$

$$\mathcal{X}_{mj}^{ij} f_i^m = \left(\mathcal{FF}_{mj}^{ij} - \mathcal{F}_{kp''}^{ij}\mathcal{F}_{mj}^{kp''} - \mathcal{F}_{p''k}^{ij}\mathcal{F}_{mj}^{p''k} - \mathcal{F}_{pq}^{ij}\mathcal{F}_{mj}^{pq}\right) f_i^m \ . \tag{2.127}$$

The treatment of the $\mathcal{B}_{ij}^{ij}$ terms is complicated by the fact that the factorization of the many-electron terms requires the use of $\hat{Q}_{12}$ in the form (2.125), and care must be taken to ensure that the error of the required double RI insertion remains manageable.

Using eq. (2.64) and the expansion (2.67), $\mathcal{B}_{ij}^{ij}$ can be written exactly as

$$\mathcal{B}_{ij}^{ij} = \mathcal{A}_{ij}^{ij} + \mathcal{A}_{ji}^{ji} - \mathcal{Z}_{ij}^{ij} - \mathcal{Z}_{ji}^{ji} - \mathcal{F}_{ab}^{ij}\mathcal{C}_{ij}^{ab} - \mathcal{F}_{ab}^{ji}\mathcal{C}_{ji}^{ab}, \tag{2.128}$$

where

$$\mathcal{A}_{ij}^{ij} = \langle ij|F_{12}\hat{f}_1\hat{Q}_{12}F_{12}|ij\rangle \tag{2.129}$$

$$\mathcal{Z}_{ij}^{ij} = \langle ij|F_{12}\hat{\mathrm{o}}_1\hat{f}_1\hat{Q}_{12}F_{12}|ij\rangle \tag{2.130}$$

$$\mathcal{C}_{ij}^{ab} = \langle ab|\hat{f}_1(1-\hat{\mathrm{v}}_1)F_{12}|ij\rangle \ , \tag{2.131}$$

Defining $\hat{P}_{12} = 1 - \hat{Q}_{12}$, one can divide $\mathcal{A}_{ij}^{ij}$ into two terms:

$$\mathcal{A}_{ij}^{ij} = \mathcal{M}_{ij}^{ij} - \mathcal{N}_{ij}^{ij} = \langle ij|F_{12}\hat{f}_1F_{12}|ij\rangle - \langle ij|F_{12}\hat{f}_1\hat{P}_{12}F_{12}|ij\rangle \ . \tag{2.132}$$

In the so-called approximation C [132, 133], $\mathcal{N}_{ij}^{ij}$, $\mathcal{Z}_{ij}^{ij}$, and $\mathcal{C}_{ij}^{ij}$ are treated exclusively through RI. For $\mathcal{N}_{ij}^{ij}$, one uses

$$\hat{P}_{12} = -\hat{\mathrm{p}}_1\hat{\mathrm{p}}_2 - \hat{\mathrm{o}}_1\hat{\mathrm{p}}_2'' - \hat{\mathrm{p}}_1''\hat{\mathrm{o}}_2 \ , \tag{2.133}$$

to obtain

$$\langle ij|F_{12}\hat{f}_1\hat{P}_{12}F_{12}|ij\rangle \approx \langle ij|F_{12}\hat{f}_1|pq\rangle\mathcal{F}_{ij}^{pq} + \langle ij|F_{12}\hat{f}_1|kp''\rangle\mathcal{F}_{ij}^{kp''} + \langle ij|F_{12}\hat{f}_1|p''k\rangle\mathcal{F}_{ij}^{p''k} \ . \tag{2.134}$$

An additional RI insertion of the form $\hat{\alpha}_1 = \hat{\mathrm{p}}_1'$ leads to

$$\langle ij|F_{12}\hat{\alpha}_1\hat{f}_1|pq\rangle\mathcal{F}_{ij}^{pq} \approx \mathcal{F}_{r'q}^{ij} f_p^{r'}\mathcal{F}_{ij}^{pq} \tag{2.135}$$

$$\langle ij|F_{12}\hat{\alpha}_1\hat{f}_1|kp''\rangle\mathcal{F}_{ij}^{kp''} \approx \mathcal{F}_{r'p''}^{ij} f_k^{r'}\mathcal{F}_{ij}^{kp''} \tag{2.136}$$

$$\langle ij|F_{12}\hat{\alpha}_1\hat{f}_1|p''k\rangle\mathcal{F}_{ij}^{p''k} \approx \mathcal{F}_{r'k}^{ij} f_{p''}^{r'}\mathcal{F}_{ij}^{p''k} \ . \tag{2.137}$$

The $\mathcal{Z}_{ij}^{ij}$ and $\mathcal{C}_{ij}^{ab}$ terms are typically small and vanish in the approximation of the GBC and EBC, respectively. $\mathcal{Z}_{ij}^{ij}$ is treated using equation (2.125) for $\hat{Q}_{12}$:

$$\langle ij|F_{12}\hat{\mathrm{o}}_1 f_1\hat{Q}_{12}F_{12}|ij\rangle \approx \mathcal{F}_{kq''}^{ij} f_{p''}^{k}\mathcal{F}_{ij}^{p''q''} + \mathcal{F}_{ka}^{ij} f_{p''}^{k}\mathcal{F}_{ij}^{p''a} \ . \tag{2.138}$$

For $\mathcal{C}_{ij}^{ab}$, one uses $1 - \hat{\mathrm{v}}_1 \approx \hat{\mathrm{o}}_1 + \hat{\mathrm{p}}_1''$ to obtain

$$\mathcal{C}_{ij}^{ab} \approx \langle ab|f_1(\hat{\mathrm{o}}_1 + \hat{\mathrm{p}}_1'')F_{12}|ij\rangle = f_{p''}^{a}\mathcal{F}_{ij}^{p''b} \ . \tag{2.139}$$

In both (2.138) and (2.139) the identity $f_a^i = 0$ is used.

The remaining terms $\mathcal{M}_{ij}^{ij} = \langle ij|F_{12}\hat{f}_1 F_{12}|ij\rangle$ could be factorized as

$$\langle ij|F_{12}\hat{\alpha}_1\hat{f}_1\hat{\alpha}_1\hat{\alpha}_2 F_{12}|ij\rangle \approx \mathcal{F}_{p'q'}^{ij} f_{r'}^{p'}\mathcal{F}_{ij}^{r'q'} \ , \tag{2.140}$$

but the error this would introduce has been shown to converge too slowly with the size of the RI basis set to be useful. Instead, one can use the fact that the operators $\hat{v}_1$ and $\hat{j}_1$ commute with F_{12} to write

$$F_{12}\hat{f}_1 F_{12} = \frac{1}{2}\big[[F_{12},\hat{t}_1],F_{12}\big] - F_{12}\hat{k}_1 F_{12} + \frac{1}{2}\big((\hat{f}_1+\hat{k}_1)F_{12}^2 + F_{12}^2(\hat{f}_1+\hat{k}_1)\big) \ , \tag{2.141}$$

where it is noted that $\hat{f}_1 + \hat{k}_1 = \hat{t}_1 + \hat{v}_1 + \hat{j}_1$.

Using the product rules for the Laplace Δ_1 and nabla ∇_1 operators one can show that

$$\big[[F_{12},\hat{t}_1],F_{12}\big] = (\nabla_1 F_{12}\cdot\nabla_1 F_{12}) \ , \tag{2.142}$$

which leads to two-electron integrals that can be readily calculated for all relevant forms of F_{12}.

The last two terms in (2.141) can be treated using a single RI insertion:

$$\langle ij|(\hat{f}_1+\hat{k}_1)\hat{\alpha}_1 F_{12}^2|ij\rangle \approx (f+k)_{p'}^{i}\mathcal{FF}_{ij}^{p'j} \tag{2.143}$$

$$\langle ij|F_{12}^2\hat{\alpha}_1(\hat{f}_1+\hat{k}_1)|ij\rangle \approx \mathcal{FF}_{p'j}^{ij}(f+k)_{i}^{p'} \ . \tag{2.144}$$

The remaining exchange operator term $\langle ij|F_{12}\hat{k}_1 F_{12}|ij\rangle$ is typically treated as in (2.140), because the errors that result for the exchange operator are acceptable [7, 8]:

$$\langle ij|F_{12}\hat{\alpha}_1\hat{k}_1\hat{\alpha}_1\hat{\alpha}_2 F_{12}|ij\rangle \approx \mathcal{F}_{p'q'}^{ij} k_{r'}^{p'}\mathcal{F}_{ij}^{r'q'} \ , \tag{2.145}$$

It is noted that these terms partially cancel with the exchange operator parts of the terms given in equations (2.135)-(2.137), (2.138), and (2.139).

The formulas given here are working equations suitable for a computer implementation. The main ingredients necessary are the determination of the CABS orbitals, the calculation of the mixed MO-CABS and CABS-CABS matrix elements of the kinetic energy, nuclear attraction, exchange and Coulomb operators, and the calculation of the two electron integrals over the operators r_{12}^{-1}, F_{12}, $F_{12}r_{12}^{-1}$, F_{12}^2, and $(\nabla_1 F_{12}\cdot\nabla_1 F_{12})$.

Density Fitting

The most costly terms that arise in the RI-factorized F12 correction energy exhibit the same $\mathcal{O}(M^5)$ scaling with the system size as the MP2 energy. They are essentially of the

same form as the numerator in the MP2 energy expression, which is made up of the direct and exchange terms,

$$D_{ijab}^{\mathrm{MP2}} = \mathcal{G}_{ab}^{ij}\mathcal{G}_{ij}^{ab}, \qquad X_{ijab}^{\mathrm{MP2}} = \mathcal{G}_{ab}^{ij}\mathcal{G}_{ji}^{ab} \ , \tag{2.146}$$

respectively.

The formal fifth order scaling in these expressions comes from the AO to MO transformation of the four-center integrals – the most expensive step scales formally as $\mathcal{O}(N_i N_\mu^4)$ – and another significant bottleneck is the $\mathcal{O}(N_i^2 N_a^2)$ scaling for the memory required to store the transformed integrals. These bottlenecks can be reduced by using the density fitted expressions for the energies

$$D_{ijab}^{\mathrm{MP2}} \approx \mathcal{G}_{ia}^{P}\tilde{\mathcal{G}}_{P}^{Q}\mathcal{G}_{jb}^{Q}\ \mathcal{G}_{ia}^{R}\tilde{\mathcal{G}}_{R}^{S}\mathcal{G}_{jb}^{S}, \qquad X_{ijab}^{\mathrm{MP2}} \approx \mathcal{G}_{ia}^{P}\tilde{\mathcal{G}}_{P}^{Q}\mathcal{G}_{jb}^{Q}\ \mathcal{G}_{ja}^{R}\tilde{\mathcal{G}}_{R}^{S}\mathcal{G}_{ib}^{S} \ , \tag{2.147}$$

where

$$\mathcal{G}_{pq}^{P} = (pq|r_{12}^{-1}|P) \tag{2.148}$$

$$\tilde{\mathcal{G}}_{Q}^{P} = [\mathcal{G}^{-1}]_{Q}^{P} \tag{2.149}$$

$$\mathcal{G}_{Q}^{P} = (P|r_{12}^{-1}|Q) \ . \tag{2.150}$$

This reduces the memory costs to $\mathcal{O}(N_P N_i N_a)$. If the Laplace transform is used to factorize the denominator in the MP2 energy expression [134–137], the direct energy can be calculated in $\mathcal{O}(N_P^2 N_i N_a)$ steps, i.e., fourth order scaling with the system size, while the scaling of the exchange energy remains fifth order ($\mathcal{O}(N_P N_i^2 N_a^2)$), but with a reduced prefactor compared to no use of DF.

This type of density fitting is also used extensively in the calculation of the F12 correction, where, e.g., (2.145) becomes in direct and exchange forms, respectively,

$$\mathcal{F}_{r'q'}^{ij} k_{p'}^{r'} \mathcal{F}_{ij}^{p'q'} \approx \mathcal{F}_{ir'}^{P}\tilde{\mathcal{F}}_{P}^{Q}\mathcal{F}_{jq'}^{Q} k_{p'}^{r'} \mathcal{F}_{ir'}^{R}\tilde{\mathcal{F}}_{R}^{S}\mathcal{F}_{jp'}^{S} \tag{2.151}$$

$$\mathcal{F}_{r'q'}^{ij} k_{p'}^{r'} \mathcal{F}_{ji}^{p'q'} \approx \mathcal{F}_{ir'}^{P}\tilde{\mathcal{F}}_{P}^{Q}\mathcal{F}_{jq'}^{Q} k_{p'}^{r'} \mathcal{F}_{jr'}^{R}\tilde{\mathcal{F}}_{R}^{S}\mathcal{F}_{ip'}^{S} \ . \tag{2.152}$$

Analogous definitions to (2.148)-(2.150) are used here. Again, the memory costs are reduced to $\mathcal{O}(M^3)$, and the calculation of the direct and exchange energies scale as $\mathcal{O}(M^4)$ and $\mathcal{O}(M^5)$, respectively. Due to the large number of terms and the need to use large auxiliary basis sets for both the F12 specific RIs and density fitting of the four-center two-electron integrals, the F12 correction costs considerably more than the calculation of the MP2 energy. However, the short ranged nature of the correlation factor means that considerable gains in efficiency are possible through the screening of negligible integrals.

Numerical Quadrature For MP2-Type Terms

In this section, possibilities are explored for the use of QD, in combination with RI and DF, for reducing the formal scaling of the MP2-F12 method to $\mathcal{O}(M^4)$.

For an MP2-type term such as $\mathcal{F}_{ab}^{ij} f_c^a \mathcal{F}_{ij}^{cb}$ (eq. (2.73)), density fitting already leads to low prefactor $\mathcal{O}(M^4)$ scaling in the case of the direct terms. For exchange-type terms, a combination of DF and QD leads to the decomposition

$$\mathcal{F}_{ab}^{ij} f_c^a \mathcal{F}_{ji}^{cb} \approx \mathcal{F}_{ia}^{P} \tilde{\mathcal{F}}_{P}^{Q} \mathcal{F}_{jb}^{Q} f_c^a w_g \phi_j^g \phi_c^g \mathcal{F}_{ib}^{g} \ , \tag{2.153}$$

where

$$\phi_j^g = \phi_j(\boldsymbol{r}_g) \tag{2.154}$$
$$\mathcal{F}_{ib}^{g} = \langle i | F_{1g} | b \rangle \ , \tag{2.155}$$

and $F_{1g} = F(|\boldsymbol{r}_1 - \boldsymbol{r}_g|)$. The calculation of the integrals $\mathcal{F}_{ib}^{g}$ scales formally as $\mathcal{O}(M^4)$ and is potentially asymptotically linearly scaling when using localized orbitals. The steps necessary for the calculation of the final energy, with their respective scaling behaviors, are

$$\mathcal{F}_{ia}^{\tilde{Q}} = \mathcal{F}_{ia}^{P} \tilde{\mathcal{F}}_{P}^{Q}, \qquad N_P^2 N_i N_a \tag{2.156}$$
$$\phi_a^g = f_c^a \phi_c^g, \qquad N_g N_a^2 \tag{2.157}$$
$$\check{\mathcal{F}}_i^{g\tilde{Q}} = \mathcal{F}_{ia}^{\tilde{Q}} \phi_a^g, \qquad N_g N_P N_i N_a \tag{2.158}$$
$$\mathring{\mathcal{F}}_b^{gQ} = \mathcal{F}_{jb}^{Q} \phi_j^g, \qquad N_g N_P N_i N_a \tag{2.159}$$
$$\bar{\mathcal{F}}_{ib}^{g} = \check{\mathcal{F}}_i^{g\tilde{Q}} \mathring{\mathcal{F}}_b^{gQ}, \qquad N_g N_P N_i N_a \tag{2.160}$$
$$w_g \bar{\mathcal{F}}_{ib}^{g} \mathcal{F}_{ib}^{g}, \qquad N_g N_i N_a \tag{2.161}$$

Therefore, the formal scaling is $\mathcal{O}(M^4)$ with the potential for only needing asymptotically linear computational time using the bounds in **Publication II**.

Despite the potential computational savings afforded by expression (2.153), the memory bottleneck for storing the three-center DF integrals remains. An advantage of the QD approach is that the grid points are completely independent and do not need to be transformed so that the three-center QD integrals can be computed and used in a batch-wise manner. This reduced memory footprint is beneficial for the use of high-performance computing architectures such as GPUs, which can offset the increased computational prefactor that results due to the fact that N_g is typically much larger than the size of an auxiliary basis set. To fully leverage these advantages a method that strictly uses QD is needed.

The following formulas alleviate the memory bottleneck through two QD integrations. The resulting expressions scale formally as $\mathcal{O}(M^4)$ but are expensive due to a quadratic dependence on the size of the grid and will therefore be most useful for large calculations on high-performance computers. The direct-type and exchange-type expressions are factorized according to

$$\mathcal{F}_{ab}^{ij} f_c^a \mathcal{F}_{ij}^{cb} \approx \mathcal{F}_{ia}^{g} w_g \phi_j^g \phi_b^g l_d^a l_c^d w_h \phi_i^h \phi_c^h \mathcal{F}_{jb}^{h}, \qquad \mathcal{F}_{ab}^{ij} f_c^a \mathcal{F}_{ji}^{cb} \approx \mathcal{F}_{ia}^{g} w_g \phi_j^g \phi_b^g l_d^a l_c^d w_h \phi_j^h \phi_c^h \mathcal{F}_{ib}^{h} \ , \tag{2.162}$$

where a factorization of the Fock matrix $f_c^a = l_d^a l_c^d$ is used for symmetry purposes. The

steps to calculate the energies, with their respective scaling behaviors, are then

$$\phi_{\tilde{a}}^{h} = l_b^a \phi_b^h, \qquad N_g N_a^2 \tag{2.163}$$

$$\mathcal{F}_i^{gh} = \mathcal{F}_{ia}^{g} \phi_{\tilde{a}}^{h}, \qquad N_g^2 N_i N_a \tag{2.164}$$

$$\mathcal{FX}^{gh} = \mathcal{F}_i^{gh} \phi_i^h, \qquad N_g^2 N_i \tag{2.165}$$

$$\mathcal{FF}^{gh} = \mathcal{F}_i^{gh} \mathcal{F}_i^{gh}, \qquad N_g^2 N_i \tag{2.166}$$

$$\mathcal{XX}^{gh} = \phi_i^g \phi_i^h, \qquad N_g^2 N_i \tag{2.167}$$

$$w_g w_h \mathcal{FX}^{gh} \mathcal{FX}^{gh}, \qquad N_g^2 \quad \text{direct} \tag{2.168}$$

$$w_g w_h \mathcal{FF}^{gh} \mathcal{XX}^{gh}, \qquad N_g^2 \quad \text{exchange} \tag{2.169}$$

Numerical Quadrature For Three- and Four-Electron Integrals

As shown above, all the contributions to the MP2-F12 energy can be put into the form of (2.73) through RI factorization, so the QD formulas given so far are enough to formulate an $\mathcal{O}(M^4)$ scaling MP2-F12 theory. However, some terms, particularly three-electron integrals, can be treated much more efficiently by using QD without first bringing them into the form of equation (2.73) through RI. In Refs. [47] and [138], the use of QD for simplifying the three- and four-electron integrals in F12 theory yields very promising results. In this work, QD is used in a similar fashion, with some differences in the way the B intermediate is treated, and the scaling of the resulting formulas is discussed.

The three-electron integrals of the V and X intermediates can be effectively factorized with a single QD integration. One obtains for example

$$\langle ijk|\frac{F_{12}}{r_{23}}|kji\rangle \approx w_g \phi_j^g \phi_j^g \mathcal{F}_{ik}^g \mathcal{G}_{ik}^g \tag{2.170}$$

$$\langle ijk|\frac{F_{12}}{r_{23}}|kij\rangle \approx w_g \phi_j^g \phi_i^g \mathcal{F}_{ik}^g \mathcal{G}_{jk}^g \ . \tag{2.171}$$

These expressions scale very favorably as $\mathcal{O}(N_g N_i^2)$, i.e., $\mathcal{O}(M^3)$, with the actual bottleneck coming from the preceding AO to MO transformation of the one-electron three-center integrals. For the exchange term (2.171), this $\mathcal{O}(M^4)$ transformation could be avoided through direct calculation in the AO basis (see eqs. (2.118) and (2.119)). These expressions are therefore potentially much more efficient than those that result from the usual RI insertions.

For the B intermediate a combination of RI and QD is the most effective strategy. Here again, the identity

$$\hat{Q}_{12}\hat{f}_1\hat{Q}_{12} = \hat{f}_1 - \hat{f}_1\hat{\mathrm{o}}_2 \tag{2.172}$$

$$+ \hat{T}(-\hat{f}_1\hat{\mathrm{o}}_1 + \hat{f}_1\hat{\mathrm{o}}_1\hat{\mathrm{o}}_2 - \hat{f}_1\hat{\mathrm{v}}_1\hat{\mathrm{v}}_2) \tag{2.173}$$

$$+ \hat{\mathrm{o}}_1\hat{f}_1\hat{\mathrm{o}}_1 - \hat{\mathrm{o}}_1\hat{f}_1\hat{\mathrm{o}}_1\hat{\mathrm{o}}_2 + \hat{\mathrm{v}}_1\hat{\mathrm{v}}_2\hat{f}_1\hat{\mathrm{v}}_1 \ , \tag{2.174}$$

is used.

The first term in (2.172) can be treated through commutator relations as in equation (2.141), where the only difficult term that arises is the integral over $F_{12}\hat{k}_1 F_{12}$. Instead of using a triple RI insertion here, a double RI insertion leads to

$$\langle ij|F_{12}\hat{\alpha}_1\hat{k}_1\hat{\alpha}_1 F_{12}|ij\rangle \approx \langle ijp'|F_{12}F_{23}|q'ji\rangle k^{q'}_{p'} \ , \tag{2.175}$$

and QD can be used to get the following formulas

$$\langle ijp'|F_{12}F_{23}|q'ji\rangle k^{q'}_{p'} \rightarrow \quad w_g\phi^g_j\phi^g_j\mathcal{F}^g_{iq'}\mathcal{F}^g_{ip'}k^{q'}_{p'}, \quad w_g\phi^g_i\phi^g_j\mathcal{F}^g_{iq'}\mathcal{F}^g_{jp'}k^{q'}_{p'} \ , \tag{2.176}$$

where here, and in the following, the QD factorization notation

$$\text{direct integral} \rightarrow \quad \text{QD form (direct)}, \quad \text{QD form (exchange)} \ ,$$

is used. These terms scale formally as $\mathcal{O}(N_g N^2_{p'} N_i)$ and $\mathcal{O}(N_g N^2_{p'})$, respectively. The scaling of the exchange-type term decreases quite significantly from $\mathcal{O}(M^5)$ for the RI+DF expression to $\mathcal{O}(M^3)$. This is very promising, especially considering that this is the most expensive term in the computation of the MP2-F12 correction using RI+DF.

The second term in (2.172) can be treated in a few ways. One way is to procede as in eqs. (2.134)-(2.137), which imply a double RI insertion, and treat the resulting MP2-type terms with QD, DF, or mixed DF-QD. To avoid a double RI insertion, it is necessary to treat the different operators in $\hat{f}_1$ separately. The three terms that arise from $\hat{t}_1$ (eqs. (2.83)-(2.85)) can be treated directly with QD:

$$\begin{aligned}
\langle ijk|F_{12}(\Delta_1 F_{13})|ikj\rangle \rightarrow \quad & w_g\phi^g_i\phi^g_i\mathcal{F}^g_{jk}(\Delta\mathcal{F})^g_{jk}, && w_g\phi^g_i\phi^g_j\mathcal{F}^g_{jk}(\Delta\mathcal{F})^g_{ik} \\
-\langle ijk|F_{12}\nabla_1 F_{13}\cdot|(\nabla i)kj\rangle \rightarrow \quad & w_g\phi^g_i\big((\nabla\phi)^g_i\cdot(\nabla\mathcal{F})^g_{jk}\big)\mathcal{F}^g_{jk}, && w_g\phi^g_i\big((\nabla\phi)^g_j\cdot(\nabla\mathcal{F})^g_{ik}\big)\mathcal{F}^g_{jk} \\
\langle ijk|F_{12}F_{13}|(\Delta i)kj\rangle \rightarrow \quad & w_g\phi^g_i(\Delta\phi)^g_i\mathcal{F}^g_{jk}\mathcal{F}^g_{jk}, && w_g\phi^g_i(\Delta\phi)^g_j\mathcal{F}^g_{jk}\mathcal{F}^g_{ik} \ ,
\end{aligned}$$

where

$$(\Delta\mathcal{F})^g_{pq} = \langle p|\Delta_1 F_{1g}|q\rangle, \qquad (\nabla\mathcal{F})^g_{pq} = \langle p|\nabla_1 F_{1g}|q\rangle, \tag{2.177}$$

$$(\Delta\phi)^g_i = (\Delta\phi_i)(\boldsymbol{r}_g), \qquad (\nabla\phi)^g_i = (\nabla\phi_i)(\boldsymbol{r}_g) \ , \tag{2.178}$$

can be calculated analytically, and $\Delta_1 F_{13} = \Delta_3 F_{13}$ and $\nabla_1 F_{13} = -\nabla_3 F_{13}$ have been used.

The operator $\hat{j}_1\hat{o}_2$ leads to the four-electron integral of (2.91), which can be treated with a single QD factorization:

$$\langle ijkl|F_{12}F_{13}r^{-1}_{14}|ikjl\rangle \rightarrow \quad w_g\phi^g_i\phi^g_i\mathcal{F}^g_{jk}\mathcal{F}^g_{jk}\mathcal{G}^g_{ll}, \quad w_g\phi^g_i\phi^g_j\mathcal{F}^g_{jk}\mathcal{F}^g_{ik}\mathcal{G}^g_{ll} \ . \tag{2.179}$$

A single RI insertion before QD factorization is required for the operator $\hat{v}_1\hat{o}_2$ in order to avoid the singular behavior of $|\boldsymbol{r}_g - \boldsymbol{r}_A|^{-1}$:

$$\langle ijk|F_{12}F_{13}\hat{\alpha}_1 r^{-1}_{1A}|ikj\rangle \approx \langle ijk|F_{12}F_{13}|p'kj\rangle v^{p'}_i \rightarrow \quad w_g\phi^g_i\phi^g_{p'}\mathcal{F}^g_{jk}\mathcal{F}^g_{jk}v^{p'}_i, \quad w_g\phi^g_i\phi^g_{p'}\mathcal{F}^g_{jk}\mathcal{F}^g_{ik}v^{p'}_j \ .$$

These expressions for the integrals over $F_{12}(\hat{t}_1 + \hat{v}_1 + \hat{j}_1)\hat{o}_2 F_{12}$ can all be calculated at $\mathcal{O}(M^3)$ cost. The remaining term over $\hat{k}_1\hat{o}_2$, however, can only be reduced to $\mathcal{O}(M^4)$. Here, a single RI insertion and then QD factorization leads to the formulas

$$\langle ij|F_{12}\hat{\alpha}_1\hat{k}_1\hat{o}_2F_{12}|ij\rangle \rightarrow \qquad \mathcal{F}^{ij}_{p'l}w_g\phi^g_k\phi^g_i\mathcal{F}^g_{lj}\mathcal{G}^g_{p'k}, \qquad \mathcal{F}^{ij}_{p'l}w_g\phi^g_k\phi^g_j\mathcal{F}^g_{li}\mathcal{G}^g_{p'k}\ , \tag{2.180}$$

which require a further factorization of $\mathcal{F}^{ij}_{p'l}$ through DF or QD.

The rest of the terms from (2.173) and (2.174) are approximated in the following way

$$\hat{T}(-\hat{f}_1\hat{o}_1 + \hat{f}_1\hat{o}_1\hat{o}_2 - \hat{f}_1\hat{v}_1\hat{v}_2) + \hat{o}_1\hat{f}_1\hat{o}_1 - \hat{o}_1\hat{f}_1\hat{o}_1\hat{o}_2 + \hat{v}_1\hat{f}_1\hat{v}_1\hat{v}_2 \tag{2.181}$$

$$\approx \hat{T}(-\hat{\alpha}_1\hat{f}_1\hat{o}_1 + \hat{\alpha}_1\hat{f}_1\hat{o}_1\hat{o}_2 - \hat{\alpha}_1\hat{f}_1\hat{v}_1\hat{v}_2) + \hat{o}_1\hat{f}_1\hat{o}_1 - \hat{o}_1\hat{f}_1\hat{o}_1\hat{o}_2 + \hat{v}_1\hat{f}_1\hat{v}_1\hat{v}_2 \tag{2.182}$$

$$= \hat{T}(-\hat{\alpha}'_1\hat{f}_1\hat{o}_1 + \hat{\alpha}'_1\hat{f}_1\hat{o}_1\hat{o}_2 - \hat{\alpha}'_1\hat{f}_1\hat{v}_1\hat{v}_2) - \hat{o}_1\hat{f}_1\hat{o}_1 + \hat{o}_1\hat{f}_1\hat{o}_1\hat{o}_2 - \hat{v}_1\hat{f}_1\hat{v}_1\hat{v}_2\ , \tag{2.183}$$

where $\hat{v}_1 f_1 \hat{o}_1 = \hat{o}_1 f_1 \hat{v}_1 = 0$ is used and at most a single RI insertion is required. The three-electron integrals that result are treated with QD and the two-electron integrals with DF or QD. This leads to $\mathcal{O}(M^3)$ scaling for the terms $\hat{T}\hat{\alpha}'_1\hat{f}_1\hat{o}_1$ and $\hat{o}_1\hat{f}_1\hat{o}_1$, and otherwise $\mathcal{O}(M^4)$ scaling.

These formulas show that numerical integration can play a large part in reducing the computational demands of explicitly correlated MP2-F12 theory by both reducing the scaling of the most expensive terms and reducing the dependence on the RI basis set. Based on the work in **Publication II**, tight integral bounds are available for all of the three- and four-electron integrals and their factorizations with either RI, DF, or QD. This allows for screening algorithms to further reduce the asymptotic scaling of all the F12 expressions to potentially $\mathcal{O}(M)$ or $\mathcal{O}(M^2)$, depending on the electronic structure.

2.6 Electronic Resonances and Non-Hermitian Quantum Chemistry

Electronic resonances are metastable electronic states characterized by a finite lifetime. These states are embedded in the continuous spectrum of the many-body Hamiltonian and are unstable with respect to ionization or detachment. Because they are part of the continuum, resonances cannot be described as a single state within Hermitian quantum mechanics, but are characterized by an increased density of states within the continuous spectrum [139]. For this reason, the methods presented in Sections 2.1 and 2.2 cannot be applied directly. A number of alternative methods have been employed to characterize resonance states, including time-dependent treatments [140–142], stabilization techniques [143–145], the Stieltjes-Tchebycheff approach [146], methods based on the core-valence separation approximation [147, 148], or methods based on finding the poles of the scattering matrix [149].

An elegant route to the characterization of resonance states is provided by *non-Hermitian* approaches, in which analytic continuation of the Hamiltonian to the complex plane is employed. The advantage of these methods is that resonance states are described by a single,

square integrable wave function. Through analytic continuation, the continuous spectrum of the Hamiltonian is rotated into the complex plane, and, additionally, discrete complex eigenvalues of the resulting non-Hermitian Hamiltonian appear [139]. These eigenvalues are the so-called Siegert energies, whose real and imaginary parts give the positions $\mathcal{E}$ and widths Γ of resonance states, respectively:

$$E = \mathcal{E} - i\Gamma/2 \tag{2.184}$$

The square integrability of the resonance eigenfunctions opens the door for the use of methods analogous to those developed for bound states.

One way to achieve this is to augment the Hamiltonian with a *complex absorbing potential* (CAP) [150–152] of the form $i\eta W(\boldsymbol{r})$. The CAP is constructed to absorb the diverging tail of the resonance wave function and the resulting Hamiltonian

$$\hat{H}_\eta = \hat{H} - i\eta W(\boldsymbol{r}) \ , \tag{2.185}$$

is non-Hermitian and its spectrum is purely discrete for finite values of η [150].

Alternatively, complex-scaling (CS) methods [10, 139] can be employed, in which all coordinates within the Hamiltonian are scaled by a complex number $e^{i\theta}$, which can be expressed as a similarity transformation of the form [10, 153]:

$$\hat{H}^\theta = e^{i\theta r(\partial/\partial r)} \hat{H} e^{-i\theta r(\partial/\partial r)} \ . \tag{2.186}$$

Because the transformation operator $e^{i\theta r(\partial/\partial r)}$ used here is unbounded (in contrast to, e.g., the operator $\exp(-\hat{T}^{(n)})$ used in coupled cluster theory, see Section 2.1.3), the spectrum of $\hat{H}^\theta$, $\sigma(\hat{H}^\theta)$, is changed compared to that of $\hat{H}$ [154]. These changes are illustrated in Fig. 2.5, which shows the rotation by an angle of 2θ of segments of the continuous spectrum around the ionization and detachment energies (thresholds) into the lower half of the complex plane. Notice that the bound-state energies and thresholds of $\hat{H}$ are invariant under complex scaling.

The mathematical basis for such a transformation is given by the Balslev-Combes theorem [156–158] and its subsequent extensions for resonances in external fields [159–161]. In the original CS formulation (2.186), the similarity transformation is not compatible with the Born-Oppenheimer approximation [162, 163]. This is not a problem for atomic systems [155], but a different approach is required for molecules. In a variant called exterior complex scaling [162], only the molecular coordinates outside a given radius are complex scaled. This variant is compatible with the Born-Oppenheimer approximation, but its direct application leads to practical difficulties in the context of electronic structure theory using Gaussian basis functions.

An alternative and related complex-variable technique that is readily applicable to molecular systems is the *method of complex basis functions*, originally proposed by McCurdy and Rescigno [163]. In this method, the asymptotic effects of CS are described through a complex-scaling of the exponents of the most diffuse basis functions. The Gaussian basis functions used in such calculations take the form

$$\chi_\mu(\boldsymbol{r}, \boldsymbol{A}) = N_\mu S_\mu(\boldsymbol{r}_{\boldsymbol{A}}) \exp[-\alpha_\mu e^{-2i\theta_\mu} \boldsymbol{r}_{\boldsymbol{A}}^2]. \tag{2.187}$$

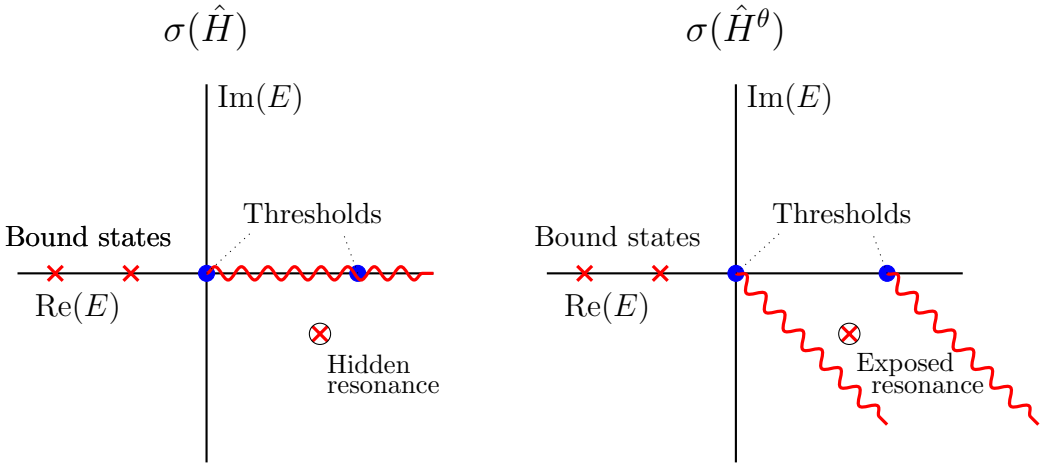


Figure 2.5: The transformation of the spectrum of the Hamiltonian through complex scaling as described by the Balslev-Combes theorem. Adapted from Ref. [155] with permission. Copyright 2013 AIP.

where N_μ is a normalization constant, $\boldsymbol{A}$ is an atomic center, $\boldsymbol{r_A} = \boldsymbol{r} - \boldsymbol{A}$, and S_μ is a real polynomial in the components of $\boldsymbol{r_A}$ that depends on the angular quantum numbers of χ_μ. Each basis function is either strictly real, $\theta_\mu = 0$, or scaled using a global complex scaling factor $\theta_\mu = \theta \in \mathbb{R}$. The mathematical justification for these methods is that, in the limit of a complete basis set, the scaling $r \to re^{i\theta}$ of the coordinates within the Hamiltonian is mathematically equivalent to the scaling $r \to re^{-i\theta}$ of the coordinates of the basis functions [139]. This approach had long been restricted to small systems and small basis sets [164–170] due to the need to calculate non-standard two-electron integrals over complex basis functions, but renewed interest led to efficient implementations at the non-Hermitian static exchange (SE) and Hartree-Fock (HF) levels of theory [171, 172]. Extensions to include electron correlation at the second-order Møller-Plesset (MP2) and coupled cluster singles and doubles (CCSD) levels of theory [173] soon followed, allowing for accurate descriptions of the resonances of small molecules. Additionally, applications to atomic and molecular Stark resonances have been detailed in Refs. [174] and [175] for coupled cluster wave functions and, very recently, the acceleration of the complex basis function HF and MP2 methods through the use of density fitting has been accomplished in Refs. [176].

The scaling used within the method of complex basis functions leads to new types of molecular integrals not present in bound state calculations and the efficient screening of these integrals is the subject of **Publication III**. In Ref. [172], it was identified that one barrier to reaching the efficiency of bound state calculations is the lack of a simple upper bound on the value of the electron repulsion integrals (ERIs) over complex basis functions. In bound state calculations, one has the Schwarz bound for screening ERIs over real basis functions [72, 105]:

$$|(\mu\nu|\lambda\sigma)| \leqslant (\mu\nu|\mu\nu)^{1/2}(\lambda\sigma|\lambda\sigma)^{1/2}. \tag{2.188}$$

In complex basis function methods, the bound (2.188) cannot be used directly, because the ERIs are complex-symmetric and not positive-semidefinite. In **Publication III**, it is shown that the proper Schwarz bound in the case of non-Hermitian complex basis function methods is given by

$$|(\mu\nu|\lambda\sigma)| \leqslant (\mu^*\nu^*|\mu\nu)^{1/2}(\lambda^*\sigma^*|\lambda\sigma)^{1/2}. \tag{2.189}$$

This bound makes the efficient screening methods of bound-state quantum chemistry available to the non-Hermitian, complex basis function description of molecular resonances. The resulting improvements in efficiency will be especially useful for the description of resonance states of large systems, where integral screening becomes vital. Two examples of areas where the resonances of large systems are important are the study of DNA damage processes caused by ionizing radiation [177–180] and in plasmonic catalysis [181, 182].

Chapter 3

Integral Partition Bounds for Local Molecular Orbitals

In **Publication II**, the integral partition bounds (IPBs) are introduced. The basic structure of these bounds is largely independent of the type of charge distributions involved, provided they tend to zero at large distances and are absolutely integrable. Given a set of such distributions $\{\Omega_1, \ldots, \Omega_N\}$, with $N \geqslant 2$ one can choose corresponding centers $\{\boldsymbol{C}_1, \ldots, \boldsymbol{C}_N\}$. These are points in $\mathbb{R}^3$ which can be chosen arbitrarily, but should be close to the centers of absolute charge to maximize effectiveness of the bounds. The IPB formalism then requires the calculation of (or the calculation of a tight upper bound for) the following overlap factor integrals:

$$\mathcal{S}_i(R) = \int\limits_{\bar{B}_i(R)} |\Omega_i(\boldsymbol{r})| d\boldsymbol{r} = \int \Theta_R(|\boldsymbol{r} - \boldsymbol{C}_i|) |\Omega_i(\boldsymbol{r})| d\boldsymbol{r} \tag{3.1}$$

$$\mathcal{V}_i(R) = \max_{\boldsymbol{r}' \in \mathbb{R}^3} \int\limits_{\bar{B}_i(R)} \frac{|\Omega_i(\boldsymbol{r})|}{|\boldsymbol{r} - \boldsymbol{r}'|} d\boldsymbol{r} = \max_{\boldsymbol{r}' \in \mathbb{R}^3} \int \Theta_R(|\boldsymbol{r} - \boldsymbol{C}_i|) \frac{|\Omega_i(\boldsymbol{r})|}{|\boldsymbol{r} - \boldsymbol{r}'|} d\boldsymbol{r} \ . \tag{3.2}$$

Here, $\bar{B}_i(R)$ is the complement (in $\mathbb{R}^3$) of a ball centered at $\boldsymbol{C}_i$ of radius R, and Θ_R is the shifted Heaviside step function

$$\Theta_R(x) = \begin{cases} 0 & x < R \\ 1 & x \geqslant R \end{cases} . \tag{3.3}$$

The closed form calculation of upper bounds for these functions is given in **Publication II** in the case of distributions made up of contracted Gaussians (AOs), or contracted Gaussian products (AO pairs). In this context, it is beneficial to treat the special cases $\mathcal{S}_i = \mathcal{S}_i(0)$ and $\mathcal{V}_i = \mathcal{V}_i(0)$ separately, as these integrals are particularly important and tighter bounds are available than in the general case. In most situations, $\mathcal{V}_i(R)$ is actually only needed in the special case of $R = 0$.

In the context of distributions made up of local MO (LMO) products some special considerations are required. In principle, an LMO, ϕ_p, is simply a contraction in the AO basis:

$$\phi_p = \sum_\mu L_{\mu p}\chi_\mu \ , \tag{3.4}$$

where the matrix $\boldsymbol{L}$ is determined, e.g., by an incomplete Cholesky factorization of the density matrix $\boldsymbol{P}$. Thus, after choosing centers $\boldsymbol{C}_{pq}$ for each MO product distribution $\phi_p\phi_q$, one could calculate

$$\mathcal{S}_{pq}(R) = \int\limits_{\bar{B}_{pq}(R)} |\phi_p(\boldsymbol{r})\phi_q(\boldsymbol{r})|d\boldsymbol{r} \leqslant \sum_{\mu\nu} |L_{\mu p}L_{\nu q}| \int\limits_{\bar{B}_{pq}(R)} |\chi_\mu(\boldsymbol{r})\chi_\nu(\boldsymbol{r})|d\boldsymbol{r} \ , \tag{3.5}$$

and similarly,

$$\mathcal{V}_{pq}(R) \leqslant \sum_{\mu\nu} |L_{\mu p}L_{\nu q}| \max_{\boldsymbol{r}'\in\mathbb{R}^3} \int\limits_{\bar{B}_{pq}(R)} \frac{|\chi_\mu(\boldsymbol{r})\chi_\nu(\boldsymbol{r})|}{|\boldsymbol{r}-\boldsymbol{r}'|}d\boldsymbol{r} \ , \tag{3.6}$$

which can be calculated from the AO formulas derived in **Publication II**. The problem that arises here is that, in contrast to the fixed contracted Gaussian coefficients, taking the absolute values over the LMO coefficients will generally lead to large overestimates.

3.1 Overlap Factors for LMOs

For the IPB overlap factors over LMOs, one can use numerical quadrature to avoid the need to take absolute values of the LMO coefficients. We note that this procedure should also lead to tighter estimates for many of the AO function pairs ($\boldsymbol{L} = \boldsymbol{1}$) as well, since some of the upper bounds needed for an analytical approach are no longer necessary. The numerical procedure consists of the following steps:

Step 1: Determining MO centers

MO centers $\boldsymbol{C}_{pq} = (C^x_{pq}, C^y_{pq}, C^z_{pq})$ are determined as the centers of absolute charge:

$$C^i_{pq} = \frac{1}{\mathcal{S}_{pq}(0)} \int r^i |\phi_p(\boldsymbol{r})\phi_q(\boldsymbol{r})|d\boldsymbol{r}, \quad i = x, y, z \tag{3.7}$$

where the integrals involved are calculated numerically using QD:

$$\mathcal{S}_{pq}(0) = \int |\phi_p(\boldsymbol{r})\phi_q(\boldsymbol{r})|d\boldsymbol{r} \approx w_g|\phi^g_p\phi^g_q| = w_g F_{pg}F_{qg}, \tag{3.8}$$

where,

$$F_{pg} = |\sum_\mu L_{\mu p}\chi^g_\mu| \ . \tag{3.9}$$

If $\mathcal{S}_{pq}(0)$ is negligible based on the integral threshold, then the LMO pair pq can be neglected entirely and the remaining procedure is only needed for significant LMO pairs.

The remaining integral in (3.7) can be calculated similarly:

$$\int r^i |\phi_p(\boldsymbol{r})\phi_q(\boldsymbol{r})| d\boldsymbol{r} \approx w_g r_g^i F_{pg} F_{qg}, \qquad i = x, y, z \ . \tag{3.10}$$

Step 2: Determining Preliminary Extents

Using the AO product centers $\boldsymbol{C}_{\mu\nu}$ and maximal extents $E_{\mu\nu}$ from **Publication II**, one can determine rough overestimates $\mathfrak{E}_{pq}$ for rigorous extents using:

$$\mathfrak{E}_{pq} = \max_{\substack{\mu\nu \\ |L_{\mu p} L_{\nu q}| \mathcal{S}_{\mu\nu} > \vartheta}} \left\{ |\boldsymbol{C}_{pq} - \boldsymbol{C}_{\mu\nu}| + E_{\mu\nu} \right\} , \tag{3.11}$$

where ϑ is the integral screening threshold and $\mu\nu$ runs over AO pairs that contribute significantly to $\phi_p\phi_q$ (determined by the criterion under the max in (3.11)).

If extents are needed based on (3.2), equation (3.11) is used with the corresponding AO extents and $\mathcal{V}_{\mu\nu}$ instead of $\mathcal{S}_{\mu\nu}$ in the significance criteria.

Step 3: Calculating $\mathcal{S}_{pq}(R)$

For the final calculation of $\mathcal{S}_{pq}(R)$, an atomic grid centered at $\boldsymbol{C}_{pq}$ is used, which can be broken down into a spherical (e.g., Lebedev) grid $\mathfrak{S}$ combined with a chosen radial grid $\mathfrak{R}$, which should integrate on the interval $[0, \max_{pq} \mathfrak{E}_{pq}]$. The total atomic grid is then $\mathfrak{G}^A = \mathfrak{R} \otimes \mathfrak{S}$, and it is noted that the number of grid points for such a grid is constant with respect to the size of the system. The integral can then be approximated as

$$\mathcal{S}_{pq}(R) \approx \sum_{\substack{r \in \mathfrak{R} \\ R < r < \mathfrak{E}_{pq}}} \sum_{g \in r \otimes \mathfrak{S}} w_g F_{pg}^{(pq)} F_{qg}^{(pq)}, \tag{3.12}$$

where

$$F_{pg}^{(pq)} = |\sum_{\mu} L_{\mu p} \chi_{\mu g}^{(pq)}|, \qquad \chi_{\mu g}^{(pq)} = \chi_\mu(\boldsymbol{r}_g + \boldsymbol{C}_{pq}) \ . \tag{3.13}$$

Because both $\mathcal{S}_{pq}(R)$ and its numerical representation are decreasing in R, one can easily use (3.12) to find, to a desired numerical precision, the smallest R, i.e., the smallest extent, such that

$$\mathcal{S}_{pq}(R) < T \ , \tag{3.14}$$

where T is a ϑ-dependent threshold that is given by the general IPB procedure. The easiest way is to simply increase R incrementally until (3.14) is fulfilled, but more sophisticated optimization procedures can also be used.

Step 4: Upper bounds for $\mathcal{V}_{pq}(R)$

If one proceeds for $\mathcal{V}_{pq}(R)$ in the same manner as for $\mathcal{S}_{pq}(R)$, the determination of the maximum would necessitate a numerical global optimization procedure over a complicated function in three dimensions. To avoid this we use the following upper bound for the charge distribution:

$$|\phi_p(\boldsymbol{r})\phi_q(\boldsymbol{r})| \leqslant \tilde{\Omega}_{pq}(|\boldsymbol{r} - \boldsymbol{C}_{pq}|) \ , \tag{3.15}$$

where

$$\tilde{\Omega}_{pq}(r) = \max_{\substack{\boldsymbol{x}\in\mathbb{R}^3 \\ |\boldsymbol{x}-\boldsymbol{C}_{pq}|=r}} |\phi_p(\boldsymbol{x})\phi_q(\boldsymbol{x})| \ . \tag{3.16}$$

Here, the value of the distribution $|\phi_p\phi_q|$ at a point $\boldsymbol{r}$ is bound by the maximum of the distribution on the ($\boldsymbol{r}$-containing) sphere centered at $\boldsymbol{C}_{pq}$ with radius $|\boldsymbol{r} - \boldsymbol{C}_{pq}|$. This leads to the upper bound

$$\mathcal{V}_{pq}(R) = \max_{\boldsymbol{r}'\in\mathbb{R}^3} \int \Theta_R(|\boldsymbol{r} - \boldsymbol{C}_{pq}|) \frac{|\phi_p(\boldsymbol{r})\phi_q(\boldsymbol{r})|}{|\boldsymbol{r} - \boldsymbol{r}'|} d\boldsymbol{r} \tag{3.17}$$

$$\leqslant \max_{\boldsymbol{r}'\in\mathbb{R}^3} \int \Theta_R(|\boldsymbol{r} - \boldsymbol{C}_{pq}|) \frac{\tilde{\Omega}_{pq}(|\boldsymbol{r} - \boldsymbol{C}_{pq}|)}{|\boldsymbol{r} - \boldsymbol{r}'|} d\boldsymbol{r} \ . \tag{3.18}$$

Per construction, the function $\boldsymbol{r} \to \tilde{\Omega}_{pq}(|\boldsymbol{r} - \boldsymbol{C}_{pq}|)$ is non-negative and spherically symmetric around the point $\boldsymbol{C}_{pq}$. This is true for $\Theta_R(|\boldsymbol{r} - \boldsymbol{C}_{pq}|)$ as well, and thus also for the product $\Theta_R(|\boldsymbol{r} - \boldsymbol{C}_{pq}|)\tilde{\Omega}_{pq}(|\boldsymbol{r} - \boldsymbol{C}_{pq}|)$. In Section 3.2, it is proven that the maximum potential of a non-negative, spherically symmetric function is always obtained at the spherical center. This allows for the determination of the remaining maximum directly:

$$\max_{\boldsymbol{r}'\in\mathbb{R}^3} \int \Theta_R(|\boldsymbol{r} - \boldsymbol{C}_{pq}|) \frac{\tilde{\Omega}_{pq}(|\boldsymbol{r} - \boldsymbol{C}_{pq}|)}{|\boldsymbol{r} - \boldsymbol{r}'|} d\boldsymbol{r} = \int \Theta_R(|\boldsymbol{r} - \boldsymbol{C}_{pq}|) \frac{\tilde{\Omega}_{pq}(|\boldsymbol{r} - \boldsymbol{C}_{pq}|)}{|\boldsymbol{r} - \boldsymbol{C}_{pq}|} d\boldsymbol{r} \tag{3.19}$$

$$= \int \Theta_R(|\boldsymbol{r}|) \frac{\tilde{\Omega}_{pq}(|\boldsymbol{r}|)}{|\boldsymbol{r}|} d\boldsymbol{r} \ . \tag{3.20}$$

The remaining integral can be calculated on the radial grid:

$$\mathcal{V}_{pq}(R) \leqslant \int \Theta_R(|\boldsymbol{r}|) \frac{\tilde{\Omega}_{pq}(|\boldsymbol{r}|)}{|\boldsymbol{r}|} d\boldsymbol{r} \tag{3.21}$$

$$= 4\pi \int_{r=R}^{\infty} r\, \tilde{\Omega}_{pq}(r) dr \tag{3.22}$$

$$\approx 4\pi \sum_{\substack{r\in\mathfrak{R} \\ R<r<\mathfrak{E}_{pq}}} w_r\, r\, \tilde{\Omega}_{pq}(r) \ , \tag{3.23}$$

The function $\tilde{\Omega}_{pq}$ can be approximated numerically at each grid point $r \in \mathfrak{R}$ as the maximum

$$\tilde{\Omega}_{pq}(r) \approx \max_{g \in r \otimes \mathfrak{S}} |\phi_p(\boldsymbol{r}_g + \boldsymbol{C}_{pq})\phi_q(\boldsymbol{r}_g + \boldsymbol{C}_{pq})| \tag{3.24}$$

$$= \max_{g \in r \otimes \mathfrak{S}} F_{pg}^{(pq)} F_{qg}^{(pq)} \ . \tag{3.25}$$

Again, both $\mathcal{V}_{pq}(R)$ and its numerical bound are decreasing in R and solving

$$\mathcal{V}_{pq}(R) < T \ , \tag{3.26}$$

is straightforward.

3.2 Maximal Coulomb Potential of Non-Negative Radial Distributions

Here, it is proven that for any integrable, non-negative radial charge distribution φ, spherically symmetric around a point $\boldsymbol{C} \in \mathbb{R}^3$, whose Coulomb potential

$$\Phi[\varphi](\boldsymbol{x}) = \int d\boldsymbol{r} \frac{\varphi(\boldsymbol{r})}{|\boldsymbol{r} - \boldsymbol{x}|} \ , \tag{3.27}$$

is finite for any $\boldsymbol{x} \in \mathbb{R}^3$, it holds that

$$\max_{\boldsymbol{x} \in \mathbb{R}^3} \Phi[\varphi](\boldsymbol{x}) = \Phi[\varphi](\boldsymbol{C}) \ , \tag{3.28}$$

i.e., the Coulomb potential of φ takes its maximum at $\boldsymbol{x} = \boldsymbol{C}$. This can be seen as follows: Translating to the center of spherical symmetry, $\boldsymbol{r} \to \boldsymbol{r} + \boldsymbol{C}$, we have

$$\Phi[\varphi](\boldsymbol{x}) = \int d\boldsymbol{r} \frac{\varphi(\boldsymbol{r} + \boldsymbol{C})}{|\boldsymbol{r} - (\boldsymbol{x} - \boldsymbol{C})|} \ . \tag{3.29}$$

Because the numerator of the integrand is now spherically symmetric with respect to the origin, Newton's shell theorem [183, 184] can be applied to obtain

$$\Phi[\varphi](\boldsymbol{x}) = \frac{1}{|\boldsymbol{x} - \boldsymbol{C}|} \int\limits_{|\boldsymbol{r}| \leqslant |\boldsymbol{x} - \boldsymbol{C}|} d\boldsymbol{r} \varphi(\boldsymbol{r} + \boldsymbol{C}) + \int\limits_{|\boldsymbol{r}| > |\boldsymbol{x} - \boldsymbol{C}|} d\boldsymbol{r} \frac{\varphi(\boldsymbol{r} + \boldsymbol{C})}{|\boldsymbol{r}|} \ . \tag{3.30}$$

Using $\{|\boldsymbol{r}| > |\boldsymbol{x} - \boldsymbol{C}|\} = \mathbb{R}^3 \backslash \{|\boldsymbol{r}| \leqslant |\boldsymbol{x} - \boldsymbol{C}|\}$ for the integration space of the second term, one obtains

$$\Phi[\varphi](\boldsymbol{x}) = \int d\boldsymbol{r} \frac{\varphi(\boldsymbol{r} + \boldsymbol{C})}{|\boldsymbol{r}|} + \int\limits_{|\boldsymbol{r}| \leqslant |\boldsymbol{x} - \boldsymbol{C}|} d\boldsymbol{r} \Big(\frac{1}{|\boldsymbol{x} - \boldsymbol{C}|} - \frac{1}{|\boldsymbol{r}|} \Big) \varphi(\boldsymbol{r} + \boldsymbol{C}) \tag{3.31}$$

$$= \Phi[\varphi](\boldsymbol{C}) + \int\limits_{|\boldsymbol{r}| \leqslant |\boldsymbol{x} - \boldsymbol{C}|} d\boldsymbol{r} \Big(\frac{1}{|\boldsymbol{x} - \boldsymbol{C}|} - \frac{1}{|\boldsymbol{r}|} \Big) \varphi(\boldsymbol{r} + \boldsymbol{C}) \tag{3.32}$$

The identity (3.32) allows for the calculation of the limit

$$\lim_{\boldsymbol{x}\to\boldsymbol{C}} \int\limits_{|\boldsymbol{r}|\leqslant|\boldsymbol{x}-\boldsymbol{C}|} d\boldsymbol{r}\Big(\frac{1}{|\boldsymbol{x}-\boldsymbol{C}|}-\frac{1}{|\boldsymbol{r}|}\Big)\varphi(\boldsymbol{r}+\boldsymbol{C}) = \lim_{\boldsymbol{x}\to\boldsymbol{C}}\Big(\Phi[\varphi](\boldsymbol{x})-\Phi[\varphi](\boldsymbol{C})\Big) = 0\ . \tag{3.33}$$

From $\varphi \geqslant 0$, and because $\frac{1}{|\boldsymbol{x}-\boldsymbol{C}|}-\frac{1}{|\boldsymbol{r}|} \leqslant 0$ follows from $|\boldsymbol{r}| \leqslant |\boldsymbol{x}-\boldsymbol{C}|$, we conclude that

$$\int\limits_{|\boldsymbol{r}|\leqslant|\boldsymbol{x}-\boldsymbol{C}|} d\boldsymbol{r}\Big(\frac{1}{|\boldsymbol{x}-\boldsymbol{C}|}-\frac{1}{|\boldsymbol{r}|}\Big)\varphi(\boldsymbol{r}+\boldsymbol{C}) \leqslant 0\ , \tag{3.34}$$

so that, together with (3.33), one obtains

$$\max_{\boldsymbol{x}\in\mathbb{R}^3} \int\limits_{|\boldsymbol{r}|\leqslant|\boldsymbol{x}-\boldsymbol{C}|} d\boldsymbol{r}\Big(\frac{1}{|\boldsymbol{x}-\boldsymbol{C}|}-\frac{1}{|\boldsymbol{r}|}\Big)\varphi(\boldsymbol{r}+\boldsymbol{C}) = 0 \tag{3.35}$$

Taking the maximum over $\Phi[\varphi](\boldsymbol{x})$ thus gives

$$\max_{\boldsymbol{x}\in\mathbb{R}^3}\Phi[\varphi](\boldsymbol{x}) = \Phi[\varphi](\boldsymbol{C}) + \max_{\boldsymbol{x}\in\mathbb{R}^3} \int\limits_{|\boldsymbol{r}|\leqslant|\boldsymbol{x}-\boldsymbol{C}|} d\boldsymbol{r}\Big(\frac{1}{|\boldsymbol{x}-\boldsymbol{C}|}-\frac{1}{|\boldsymbol{r}|}\Big)\varphi(\boldsymbol{r}+\boldsymbol{C}) \tag{3.36}$$

$$= \Phi[\varphi](\boldsymbol{C})\ , \tag{3.37}$$

which concludes the proof.

Chapter 4

Publications

4.1 Publication I

Distance-including rigorous upper bounds and tight estimates for two-electron integrals over long- and short-range operators

T. H. Thompson and C. Ochsenfeld,
J. Chem. Phys. **147**, 144101 (2017).

Abstract:

We introduce both rigorous and non-rigorous distance-dependent integral estimates for four-center two-electron integrals derived from a distance-including Schwarz-type inequality. The estimates are even easier to implement than our so far most efficient distance-dependent estimates [S. A. Maurer *et al.*, J. Chem. Phys. **136**, 144107 (2012)] and, in addition, do not require well-separated charge distributions. They are also applicable to a wide range of two-electron operators such as those found in explicitly correlated theories and in short-range hybrid density functionals. For two such operators with exponential distance decay [$e^{-r_{12}}$ and $\mathrm{erfc}(0.11 \cdot r_{12})/r_{12}$], the rigorous bound is shown to be much tighter than the standard Schwarz estimate with virtually no error penalty. The non-rigorous estimate gives results very close to an exact screening for these operators and for the long-range $1/r_{12}$ operator, with errors that are completely controllable through the integral screening threshold. In addition, we present an alternative form of our non-rigorous bound that is particularly well- suited for improving the PreLinK method [J. Kussmann and C. Ochsenfeld, J. Chem. Phys. **138**, 134114 (2013)] in the context of short-range exchange calculations.

The following article is reproduced in agreement with its publisher (AIP Publishing LLC) and can be found online at: https://doi.org/10.1063/1.4994190

THE JOURNAL OF CHEMICAL PHYSICS **147**, 144101 (2017)

Distance-including rigorous upper bounds and tight estimates for two-electron integrals over long- and short-range operators

Travis H. Thompson and Christian Ochsenfeld
Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), Butenandtstr. 7, D-81377 Munich, Germany and Center for Integrated Protein Science Munich (CIPSM) at the Department of Chemistry, University of Munich (LMU), Butenandtstr. 5-13, D-81377 Munich, Germany

(Received 4 July 2017; accepted 5 September 2017; published online 9 October 2017)

We introduce both rigorous and non-rigorous distance-dependent integral estimates for four-center two-electron integrals derived from a distance-including Schwarz-type inequality. The estimates are even easier to implement than our so far most efficient distance-dependent estimates [S. A. Maurer *et al.*, J. Chem. Phys. **136**, 144107 (2012)] and, in addition, do not require well-separated charge-distributions. They are also applicable to a wide range of two-electron operators such as those found in explicitly correlated theories and in short-range hybrid density functionals. For two such operators with exponential distance decay [$e^{-r_{12}}$ and $\mathrm{erfc}(0.11 \cdot r_{12})/r_{12}$], the rigorous bound is shown to be much tighter than the standard Schwarz estimate with virtually no error penalty. The non-rigorous estimate gives results very close to an exact screening for these operators and for the long-range $1/r_{12}$ operator, with errors that are completely controllable through the integral screening threshold. In addition, we present an alternative form of our non-rigorous bound that is particularly well-suited for improving the PreLinK method [J. Kussmann and C. Ochsenfeld, J. Chem. Phys. **138**, 134114 (2013)] in the context of short-range exchange calculations. *Published by AIP Publishing.*
https://doi.org/10.1063/1.4994190

I. INTRODUCTION

Reducing the computational effort of quantum-chemical calculations requires taking advantage of the sparse nature of the interactions involved. This can be done efficiently through low-cost estimates that allow one to skip the exact calculation of numerically vanishing contributions. Traditionally, the development of reliable estimates of the four-center two-electron integrals over the Coulomb operator $1/r_{12}$ has played an important role in decreasing the storage and computational requirements. The use of the Schwarz estimate (QQ),

$$|(\mu\nu|\lambda\sigma)| \leq (\mu\nu|\mu\nu)^{1/2}(\lambda\sigma|\lambda\sigma)^{1/2} = Q_{\mu\nu}Q_{\lambda\sigma}, \quad (1)$$

where

$$(\mu\nu|\lambda\sigma) = \int\int \frac{\chi_\mu(\mathbf{r}_1)\chi_\nu(\mathbf{r}_1)\chi_\lambda(\mathbf{r}_2)\chi_\sigma(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2, \quad (2)$$

and $\{\chi_\mu\}$ is a finite set of sufficiently local orbitals (usually Gaussians), represents the first breakthrough in integral screening techniques.[1,2] It accurately captures the—in the case of Gaussian basis functions—exponential decay of two-electron integrals with increasing distances between the centers of χ_μ and χ_ν, and χ_λ and χ_σ (overlap dependence), while also being simple and very efficient to calculate.

More recently,[3] estimates have been developed that not only capture the overlap decay but also incorporate the decay of (2) arising from the behaviour of the Coulomb operator as the distance between the local charge distributions $\chi_\mu\chi_\nu$ and $\chi_\lambda\chi_\sigma$ becomes large (distance dependence). In this regard, Maurer *et al.* showed that a good approximation of the integrals can be even more useful than a strict upper bound and introduced the QQR estimate[3] given by

$$|(\mu\nu|\lambda\sigma)| \approx \begin{cases} Q_{\mu\nu}Q_{\lambda\sigma}(\tilde{R}^{\lambda\sigma}_{\mu\nu})^{-1}, & \tilde{R}^{\lambda\sigma}_{\mu\nu} > 1 \\ Q_{\mu\nu}Q_{\lambda\sigma}, & \tilde{R}^{\lambda\sigma}_{\mu\nu} \leq 1 \end{cases}, \quad (3)$$

where

$$\tilde{R}^{\lambda\sigma}_{\mu\nu} = R^{\lambda\sigma}_{\mu\nu} - \mathrm{ext}_{\mu\nu} - \mathrm{ext}_{\lambda\sigma}. \quad (4)$$

Here, $R^{\lambda\sigma}_{\mu\nu}$ is the distance between the centers of the two charge distributions $\chi_\mu\chi_\nu$ and $\chi_\lambda\chi_\sigma$, while $\mathrm{ext}_{\mu\nu}$ and $\mathrm{ext}_{\lambda\sigma}$ are their respective numerical extents (precise definitions in terms of entire shells are given in Ref. 3).

Further advances include the development of tighter estimates for three-center Coulomb integrals[4] and the extension of QQR to the operator $\mathrm{erfc}(\omega r_{12})/r_{12}$[5] used in some density functional theory (DFT) methods employing short-range exchange.

In the following, we introduce integral estimates that accurately describe the distance dependence and are applicable to a wide range of multiplicative, distance-dependent operators, including those that arise in some modern quantum-chemical methods, such as explicitly correlated F12[6,7] and short-range hybrid DFT methods.[8] The expressions are as simple and easily computable as the Schwarz estimates and are used to form both rigorous and even tighter non-rigorous estimates. In addition, they employ a distance-including term that is simpler and more efficient to calculate than that used in QQR type estimates. We test the estimates for integrals over the Coulomb operator $1/r_{12}$, which we call "long-range" due to their slow inverse distance decay, and over the operators $e^{-r_{12}}$ and $\mathrm{erfc}(0.11 \cdot r_{12})/r_{12}$, which we call "short-range" due to

their exponential distance decay. The operator $e^{-\gamma r_{12}}$ is important in explicitly correlated F12 theory with a true Slater-type geminal[9] and optimal values of γ depend on the basis set but are typically close to 1.0.[10] The operator $\mathrm{erfc}(0.11 \cdot r_{12})/r_{12}$ is used in, e.g., the HSE06,[11] HSE-3c,[12] and N12-SX[13] density functionals.

II. THEORY

The QQ estimate (1) remains applicable when the Coulomb operator $1/r_{12}$ is replaced by any positive-definite integral kernel, and in the following, we use the integral notation $(\mu\nu|\lambda\sigma)$ as in (2), but with the Coulomb operator replaced by some general function G of the inter-electronic distance. For proof of this statement and of the positive-definiteness of various important operators used in quantum chemistry, including those in this work, see Appendix B. Thus, from now on, the factor $Q_{\mu\nu}$ always depends implicitly on the operator G. In addition, we consider the generalized QQR estimate of the form

$$|(\mu\nu|\lambda\sigma)| \approx \begin{cases} Q_{\mu\nu}Q_{\lambda\sigma}G(\tilde{R}^{\lambda\sigma}_{\mu\nu}), & 0 < G(\tilde{R}^{\lambda\sigma}_{\mu\nu}) < 1 \\ Q_{\mu\nu}Q_{\lambda\sigma}, & \text{otherwise} \end{cases}, \quad (5)$$

with $\tilde{R}^{\lambda\sigma}_{\mu\nu}$ as in Eq. (4).

A. A rigorous upper bound for short-range operators

The QQ estimate is the Schwarz inequality on the space of one-electron charge distributions defined as $\Omega_{\mu\nu}(r_1) = \chi_\mu(r_1)\chi_\nu(r_1)$. It is applicable due to the fact that the two-electron integral over some distance-dependent operator G is an inner product on this space, as long as G is positive definite. We show in Appendix A that the integral is also an inner product on the space of real continuous two-electron functions, if G is strictly positive for $r_{12} > 0$. The two-electron orbital products $\tilde{\Omega}_{\mu\nu}(r_1, r_2) = \chi_\mu(r_1)\chi_\nu(r_2)$ are a subset of this space and the corresponding Schwarz inequality leads, due to symmetry, to the following two upper bounds:

$$|(\mu\nu|\lambda\sigma)| \le (\mu\mu|\lambda\lambda)^{1/2}(\nu\nu|\sigma\sigma)^{1/2} = M_{\mu\lambda}M_{\nu\sigma}, \quad (6)$$

$$|(\mu\nu|\lambda\sigma)| \le (\mu\mu|\sigma\sigma)^{1/2}(\nu\nu|\lambda\lambda)^{1/2} = M_{\mu\sigma}M_{\nu\lambda}. \quad (7)$$

In contrast to the original QQ estimates, the right-hand sides of (6) and (7) inherently contain distance dependence, while lacking any overlap dependence. Equality holds in (6) and (7) for integrals of type $(\mu\mu|\nu\nu)$, i.e., for "perfect" overlap in forming the charge distributions. We combine the original QQ estimates (1) and the Schwarz-type inequalities (6) and (7) into a rigorous distance-including upper bound, the combined Schwarz bound (CSB),

$$|(\mu\nu|\lambda\sigma)| \le \min\left(Q_{\mu\nu}Q_{\lambda\sigma}, M_{\mu\lambda}M_{\nu\sigma}, M_{\mu\sigma}M_{\nu\lambda}\right). \quad (8)$$

For the long-range Coulomb operator, overlap dependence is much more important than distance dependence and we find that the CSB estimate is no more useful than the QQ estimate for currently tractable systems. However, for operators with much stronger distance-decay, such as $e^{-r_{12}}$ and $\mathrm{erfc}(0.11 \cdot r_{12})/r_{12}$, we find that that the rigorous CSB provides a much tighter bound and correctly captures the linear scaling increase in the number of significant two-electron integrals already for medium-sized systems. This is in contrast to the QQ estimates, which scale quadratically with system size regardless of the operator, due to the missing distance-dependence.

Lastly, we note that the CSB estimates are exact for both $(\mu\nu|\mu\nu)$-type and $(\mu\mu|\nu\nu)$-type integrals. This is because the QQ estimates are exact for the former integral types, while Eqs. (6) and (7) are exact for the latter, and all three are upper bounds.

B. Tight estimates for short- and long-range operators

The descriptions provided by the QQ bounds and those in Eqs. (6) and (7) can be seen as being complementary. While the QQ estimates describe the overlap dependence at zero distance, the new bounds provide a description of the distance dependence at full overlap. This makes it plausible that they can be combined in a way that captures both properties simultaneously and gives a good estimate for most, if not all, of the two-electron integrals. In order to preserve the exactness of the estimates for $(\mu\nu|\mu\nu)$-type integrals, we developed the following normalized description of the distance decay:

$$\tilde{M}_{\mu\lambda} = \frac{M_{\mu\lambda}}{Q^{1/2}_{\mu\mu}Q^{1/2}_{\lambda\lambda}}. \quad (9)$$

We note that $0 < \tilde{M}_{\mu\lambda} \le 1$ and that $\tilde{M}_{\mu\lambda} = 1$ if $\chi_\mu = \chi_\lambda$. This follows from the QQ estimate of $M_{\mu\lambda}$,

$$M_{\mu\lambda} \le Q^{1/2}_{\mu\mu}Q^{1/2}_{\lambda\lambda}, \quad (10)$$

where equality holds if $\chi_\mu = \chi_\lambda$. The denominator in (9) depends neither on the overlap nor on the distance but ensures that the factor $\tilde{M}_{\mu\lambda}$ gives unity for $\chi_\mu = \chi_\lambda$ so that an estimate derived by multiplying the QQ bound with this factor will reduce to the QQ bound for $(\mu\nu|\mu\nu)$-type integrals, which is exact in this case. On the other hand, the distance decay of the factor $M_{\mu\lambda}$ is included. We use $\tilde{M}_{\mu\lambda}$ to formulate our non-rigorous combined Schwarz approximations (CSAs),

$$|(\mu\nu|\lambda\sigma)| \approx Q_{\mu\nu}Q_{\lambda\sigma}\mathcal{M}_{\mu\nu\lambda\sigma}, \quad (11)$$

with three different approximations [CSA1, CSA2, and CSA max (CSAM)] given by the choice of $\mathcal{M}_{\mu\nu\lambda\sigma}$,

$$\mathcal{M}_{\mu\nu\lambda\sigma} = \begin{cases} \tilde{M}_{\mu\lambda}\tilde{M}_{\nu\sigma}, & \text{CSA1} \\ \tilde{M}_{\mu\sigma}\tilde{M}_{\nu\lambda}, & \text{CSA2} \\ \max\left(\tilde{M}_{\mu\lambda}\tilde{M}_{\nu\sigma}, \tilde{M}_{\mu\sigma}\tilde{M}_{\nu\lambda}\right), & \text{CSAM} \end{cases}. \quad (12)$$

We note that in addition to $0 < \tilde{M}_{\mu\lambda} \le 1$, one also has

$$0 < \frac{Q_{\mu\nu}}{Q^{1/2}_{\mu\mu}Q^{1/2}_{\nu\nu}} \le 1, \quad (13)$$

with equality on the right-hand side for $\chi_\mu = \chi_\nu$, which follows by estimating $Q_{\mu\nu}$ with Eq. (6) [or (7)]. This ensures that the CSA estimates are also exact for integrals of type $(\mu\mu|\nu\nu)$ because they then reduce to the bound (6), which is exact in this case.

We note that CSA1 and CSA2 offer more flexibility due the separated form of $\mathcal{M}_{\mu\nu\lambda\sigma}$ but are slightly less accurate than CSAM (see Sec. III A). We have found, as one might expect, that CSA1 and CSA2 perform nearly identically and we only present results for CSA1.

The combined Schwarz approximations exhibit many excellent qualities. For one, they require only diagonal elements of the integrals themselves. Furthermore, no well-separatedness condition is required for their application, in contrast to QQR, and the error incurred in their use is mediated through only one screening parameter, the integral threshold. In addition, their use is very inexpensive because the distance factors $\tilde{M}_{\mu\nu}$ can be easily precalculated and stored, and only simple multiplication is needed in the screening step. Lastly, they lead, in terms of integral shell counts, to a nearly exact screening, i.e., the neglect of all integral shells with exact norms below the integral threshold, for all of the operators we have tested.

C. Incorporation into Coulomb and exchange matrix screening algorithms

Modern methods allow for the linear-scaling calculation of both the Coulomb (**J**) and exchange (**K**) matrices needed in Hartree-Fock (HF) theory and hybrid DFT (see, e.g., Ref. 14). **J** and **K** are defined in terms of the density matrix **P** and the two-electron integrals as

$$J_{\mu\nu} = \sum_{\lambda\sigma} P_{\lambda\sigma} (\mu\nu|\lambda\sigma), \tag{14}$$

$$K_{\mu\nu} = \sum_{\lambda\sigma} P_{\lambda\sigma} (\mu\lambda|\nu\sigma), \tag{15}$$

and traditionally, individual contributions are screened, i.e., contributions estimated below a chosen screening threshold are neglected, via the QQ bound,

$$|P_{\lambda\sigma} (\mu\nu|\lambda\sigma)| \leq |P_{\lambda\sigma}| Q_{\mu\nu} Q_{\lambda\sigma}, \tag{16}$$

$$|P_{\lambda\sigma} (\mu\lambda|\nu\sigma)| \leq |P_{\lambda\sigma}| Q_{\mu\lambda} Q_{\nu\sigma}. \tag{17}$$

In the case of the Coulomb matrix, the continuous fast multipole method (CFMM)[15] can be used to partition **J** into near-field and far-field contributions based on interaction distance. The near-field part inherently contains only a linear scaling number of interactions due to the distance cutoff and it is calculated using exact integrals, while the formally quadratic scaling far-field contribution is handled with arbitrary precision using a highly efficient multipole expansion based algorithm in a linear scaling fashion. Typically, the near-field contribution dominates Coulomb matrix calculation times and integral screening via Eq. (16) is used to reduce the prefactor.

The exchange matrix can be calculated in a linear scaling fashion for systems with non-vanishing HOMO-LUMO gaps, such that the density matrix **P** is sparse.[16–18] This is because the two charge distributions present in exchange terms are coupled through the density matrix as seen in (17) so that the number of significant contributions to **K** scales linearly in this case. Alternatively, very recent work[19] explores the use of fast multipole methods for calculating the exchange matrix.

While a naïve screening using Eq. (17) will scale quadratically with the system size, the LinK method[17,18] introduces a prescreening/preordering step preceding the Schwarz screening that reduces the complexity to linear for sparse **P** matrices.

We implement our estimates within the Coulomb near-field screening and LinK screening algorithms by simply adding a final screening step in which integrals are neglected when the estimates,

$$|P_{\lambda\sigma} (\mu\nu|\lambda\sigma)| \approx |P_{\lambda\sigma}| A_{\mu\nu\lambda\sigma}, \tag{18}$$

$$|P_{\lambda\sigma} (\mu\lambda|\nu\sigma)| \approx |P_{\lambda\sigma}| A_{\mu\lambda\nu\sigma}, \tag{19}$$

are less than some fixed thresholds ϑ_j and ϑ_k, for Coulomb and exchange contributions, respectively. Here, $A_{\mu\nu\lambda\sigma}$ represents one of the approximations to the two-electron integrals described above. Although one could tailor the prescreening step to the particular approximation used, we have found that this has little impact on the efficiency of calculating the exact exchange matrix, i.e., for two-electron integrals over the $1/r_{12}$ operator, to which this work is restricted. This is because the slow distance-decay allows for only small improvements in the coarse prescreening step. In addition, the large amount of non-negligible integrals means that the screening step makes up a small fraction of the calculation time in this case.

Lastly, we note that while the LinK prescreening algorithm is very efficient on central processing units (CPUs), it is not suited for use on graphical processing units (GPUs) due to the complex control flow required. However, the PreLinK prescreening method[20] allows for linear-scaling through a coarse preselection of significant **K** matrix elements before integral calculation begins using only efficient sparse matrix multiplications. Defining $|\mathbf{P}|$ as the matrix of absolute density matrix elements and **Q** as the matrix formed from the factors $Q_{\mu\nu}$, the approximated matrix of absolute exchange elements $|\mathbf{K}|'$ is given by the simple expression,

$$|\mathbf{K}|' = \mathbf{Q}|\mathbf{P}|\mathbf{Q}. \tag{20}$$

Because the CSB, QQR, and CSAM estimates do not separate into two-index quantities, they cannot be used straightforwardly within the PreLinK method because they do not lead to an efficient prescreening involving only matrix operations. On the other hand, the CSA1 estimates lead to a very simple expression for $|\mathbf{K}|'$,

$$|\mathbf{K}|' = \tilde{\mathbf{M}} \circ (\mathbf{Q}(\tilde{\mathbf{M}} \circ |\mathbf{P}|)\mathbf{Q}), \tag{21}$$

where $\circ$ symbolizes the Hadamard matrix product and $\tilde{\mathbf{M}} = (\tilde{M}_{\mu\nu})_{\mu\nu}$. Thus, only two Hadamard products and two matrix multiplications are required, which again can be performed in a linear-scaling fashion.

We have found that, in terms of speed vs. accuracy, Eq. (21) performs the same as (20) for the exact exchange calculations to which this work is restricted. This is not surprising considering the long-range nature of the $1/r_{12}$ operator and that the PreLinK method is a coarse prescreening. We note however that for short-range operators, the matrix $\tilde{\mathbf{M}}$ will become sparse for currently tractable systems and we have found Eq. (21) to perform significantly better than (20) in this case. Additionally, the sparsity of $\tilde{\mathbf{M}}$ means that for both the LinK and PreLinK methods, linear-scaling will be achievable even for small band-gap systems with densely populated density matrices. We will present results for short-range exchange calculations in future work.

III. COMPUTATIONAL DETAILS AND RESULTS

We compare the performance of the rigorous CSB [Eq. (8)] estimate and the non-rigorous CSA1 and CSAM [Eqs.

(11) and (12)] estimates to that of the established QQ [Eq. (1)] and QQR [Eq. (5)] estimates. For the QQR estimates, we employ a fixed well-separatedness threshold of 10^{-1}, which has been shown to give sufficient accuracy.[3] As test systems, we use a subset of the screening benchmark test suite described in Ref. 3.

In Sec. III A, preliminary results based on statistics and integral only screenings are given for the long-range Coulomb operator $1/r_{12}$ and the short-range operators $e^{-r_{12}}$ and $\mathrm{erfc}(0.11 \cdot r_{12})/r_{12}$. The importance of these operators is given at the end of Sec. I. We give results for a range of thresholds and also compare estimate performance for a series of linear alkanes containing up to 160 carbon atoms.

In Sec. III B, we give results for the $1/r_{12}$ operator in the context of Hartree-Fock SCF calculations to demonstrate the applicability of our estimates in electronic structure theory.

Due to the way in which molecular integrals are calculated, it is advantageous to screen at the level of integral shell-quartets. We estimate the contribution of a shell-quartet by taking the maximum norm of our estimates with respect to entire shells, before combining these shellwise estimates in the screening step to a final estimate. If a shell-quartet is deemed non-significant, none of the corresponding integrals are calculated.

A. Statistics and estimate performance

As a first test, we evaluate the performance of the estimates through statistics of the ratio

$$F = I_{\text{estimate}}/I_{\text{exact}}, \tag{22}$$

where I_{estimate} and I_{exact} are the estimated and actual norms of an integral shell-quartet, respectively. The maximum ($F_{\max}$), minimum ($F_{\min}$), average (F_{av}), and standard deviation [$\sigma(F)$] of F are given for the molecule Amylose_{16} from the test suite and the basis sets cc-pVDZ and cc-pVTZ in Tables I and II, respectively. Due to the way the integrals over the $\mathrm{erfc}(0.11 \cdot r_{12})/r_{12}$ operator are calculated as the difference between the integrals over the $1/r_{12}$ and $\mathrm{erf}(0.11 \cdot r_{12})/r_{12}$ operators, numerical instability occurs for large distances between charge distributions. Such integrals are calculated as the small difference between large values. As we show below, this does not affect the real performance of our estimates when using a fixed integral threshold; however, it distorts the statistics of F. To avoid this distortion, we restrict our statistics to shell-quartets with exact norms larger than 10^{-12}.

For the cc-pVDZ basis set we see, in the case of the $1/r_{12}$ operator, only slightly improved statistics with the CSB versus the QQ estimate. This is expected due to the weak distance dependence for this operator and we suspect that most of the improvement comes from $(\mu\mu|\nu\nu)$-type integrals for which CSB is exact while the QQ bounds are strict overestimates (assuming $\chi_\mu \neq \chi_\nu$). The non-rigorous QQR, CSAM, and CSA1 bounds perform similarly, with significantly improved averages, standard deviations, and maximal overestimates compared to the two rigorous bounds. The two CSA estimates have slightly improved standard deviations at a cost of slightly larger errors indicated by lower $F_{\min}$ values.

In the case of the $e^{-r_{12}}$ operator, the CSB estimate shows large improvements to the rather poorly performing QQ bound.

TABLE I. Statistics of the ratio $F = I_{\text{estimate}}/I_{\text{exact}}$ for the molecule Amylose_{16}, the cc-pVDZ basis set, and the estimates QQ, CSB, QQR, CSAM, and CSA1. Only shell-quartets with exact norms above 10^{-12} are included in the statistics.

cc-pVDZ					
Operator	Estimates	F_{av}	$\sigma(F)$	$F_{\min}$	$F_{\max}$
$1/r_{12}$	QQ	70.38	1.5×10^{3}	1.000	4.3×10^{6}
	CSB	62.06	954.00	1.000	2.3×10^{6}
	QQR	2.27	24.29	0.376	4.6×10^{4}
	CSAM	2.32	17.01	0.161	1.9×10^{4}
	CSA1	2.30	17.01	0.083	1.9×10^{4}
$e^{-r_{12}}$	QQ	1.4×10^{6}	2.0×10^{7}	1.000	7.3×10^{9}
	CSB	1.5×10^{3}	6.0×10^{3}	1.000	5.1×10^{5}
	QQR	12.76	47.45	0.179	1.6×10^{4}
	CSAM	2.39	6.20	0.002	1.3×10^{3}
	CSA1	1.88	4.97	2.4×10^{-4}	911.86
$\frac{\mathrm{erfc}(0.11 \cdot r_{12})}{r_{12}}$	QQ	2.0×10^{6}	4.0×10^{7}	1.000	3.9×10^{10}
	CSB	2.0×10^{3}	7.1×10^{3}	1.000	5.5×10^{5}
	QQR	5.14	5.49	0.448	741.05
	CSAM	1.48	1.30	0.042	132.70
	CSA1	1.34	1.22	0.014	125.33

Again, the non-rigorous estimates outperform the rigorous bounds greatly but come with the cost of underestimates which are more severe than in the case of the Coulomb operator. The CSA estimates are significantly better on average than the QQR estimates but also show much larger underestimates.

The results for the operator $\mathrm{erfc}(0.11 \cdot r_{12})/r_{12}$ are similar to those for $e^{-r_{12}}$. The non-rigorous estimates however show improved values in each category with the same general trends between them.

The results for the cc-pVTZ basis are very similar to those for the cc-pVDZ basis. The trends stay the same, but all

TABLE II. Statistics of the ratio $F = I_{\text{estimate}}/I_{\text{exact}}$ for the molecule Amylose_{16}, the cc-pVTZ basis set, and the estimates QQ, CSB, QQR, CSAM, and CSA1. Only shell-quartets with exact values above 10^{-12} are included in the statistics.

cc-pVTZ					
Operator	Estimates	F_{av}	$\sigma(F)$	$F_{\min}$	$F_{\max}$
$1/r_{12}$	QQ	101.54	2.6×10^{3}	1.000	9.0×10^{6}
	CSB	90.08	2.1×10^{3}	1.000	9.0×10^{6}
	QQR	3.05	43.28	0.337	9.9×10^{5}
	CSAM	3.15	30.06	0.145	3.4×10^{4}
	CSA1	3.12	30.06	0.069	3.4×10^{4}
$e^{-r_{12}}$	QQ	1.5×10^{6}	1.9×10^{7}	1.000	7.4×10^{9}
	CSB	1.8×10^{3}	8.1×10^{3}	1.000	1.2×10^{7}
	QQR	11.05	98.69	0.119	1.0×10^{6}
	CSAM	3.88	19.17	0.001	3.4×10^{5}
	CSA1	2.97	14.74	1.1×10^{-4}	2.2×10^{5}
$\frac{\mathrm{erfc}(0.11 \cdot r_{12})}{r_{12}}$	QQ	2.0×10^{6}	3.5×10^{7}	1.000	4.0×10^{10}
	CSB	2.3×10^{3}	8.6×10^{3}	1.000	6.5×10^{6}
	QQR	4.74	20.67	0.306	7.8×10^{5}
	CSAM	1.97	3.84	0.033	1.2×10^{5}
	CSA1	1.77	3.63	0.010	1.2×10^{5}

estimates perform slightly worse than with the smaller basis set.

We also present results regarding the efficiency and accuracy of direct integral screenings and compare to an *a posteriori* exact screening. In these tests, integrals are discarded when they are estimated to be smaller than a fixed threshold ϑ. In the case of the exact screening, they are discarded if they truly are smaller than the threshold. As a measure of the efficiency, we use significant shell-quartet totals, and as a measure of accuracy, the entrywise 1-norm (W) of the difference between the exact two-electron integral tensor $\{(\mu\nu|\lambda\sigma)\}$ and the "screened" tensor $\{(\mu\nu|\lambda\sigma)'\}$ with integrals from non-significant shell-quartets set to zero,

$$W = \sum_{\mu\nu\lambda\sigma} |\,(\mu\nu|\lambda\sigma) - (\mu\nu|\lambda\sigma)'\,|. \tag{23}$$

We note that W is just the unsigned sum of all discarded integral values. Although W will be much larger than the actual screening error due to error cancellation, it offers a good metric for comparison with the well-tested estimates QQ and QQR and with the error involved in an exact screening. We note that for non-rigorous methods, W will include contributions from integrals that are larger than the screening threshold; however, our accuracy metric treats such terms on equal footing with correctly discarded integrals, making a comparison with rigorous estimates possible. In almost all cases, the contribution from accurately screened integrals is larger or of the same order of magnitude as that coming from those discarded erroneously.

We tested our estimates using various screening thresholds. In Fig. 1, the ratio of the number of significant shell-quartets N_{SQ} of the most conservative ($\vartheta = 14$) QQ screening to those of each method/threshold (speedup) and the W values are plotted against a sequence of increasing screening thresholds ϑ for all three operators tested. The numbers are calculated using sums of the N_{SQ} and W data from the molecules Amylose_{32}, Polyethyne_{128}, and $(\text{S}_8)_{20}$ from the test suite.

For the $1/r_{12}$ operator [Figs. 1(a) and 1(d)], we find no improvement in efficiency for the rigorous CSB estimate and no change in accuracy compared to the QQ estimate. In fact, virtually identical sets of integrals are discarded in both cases. On the other hand, the non-rigorous estimates, which perform nearly identically, show significant improvements in efficiency close to that achieved with an exact screening. This increased speed is accompanied by an increase in error of more than an order of magnitude compared to the rigorous bounds. However the error is virtually identical to that of an exact screening and is controllable through the screening threshold. The error increases linearly with increasing ϑ.

In Fig. 1(b), we see that the speedups for the $\text{erfc}(0.11 \cdot r_{12})/r_{12}$ operator are much higher than for the $1/r_{12}$ operator for all estimates except QQ, for which they are more or less identical (notice the different scales for the speedup). Remarkably, the rigorous CSB estimate shows a speedup of ca. 2.5 already for $\vartheta = 10^{-14}$ with virtually no error increase. This can only be matched by the QQ estimate by increasing ϑ by six orders of magnitude, which is accompanied by an increase in error of roughly six orders of magnitude. The speedup gap between the two rigorous estimates increases with increasing ϑ, while the errors incurred [see Fig. 1(e)] remain virtually identical. Therefore, it seems that there can be no justification for the use of QQ instead of CSB for this operator. Looking at the non-rigorous estimates, we see speedups rivaling the exact screening from the CSA estimates with the QQR estimate discarding considerably less. The gap in speed between the non-rigorous and rigorous estimates has widened significantly in comparison to the $1/r_{12}$ operator. In contrast to the speedups, the errors of the screenings decrease across all estimates compared to the $1/r_{12}$ operator, especially for the QQ, QQR, and CSB estimates. This is notable considering the fact that many more integrals are discarded for this operator. One sees as well that the speedup of the CSA estimates compared to QQR comes at the price of a slight decrease in accuracy.

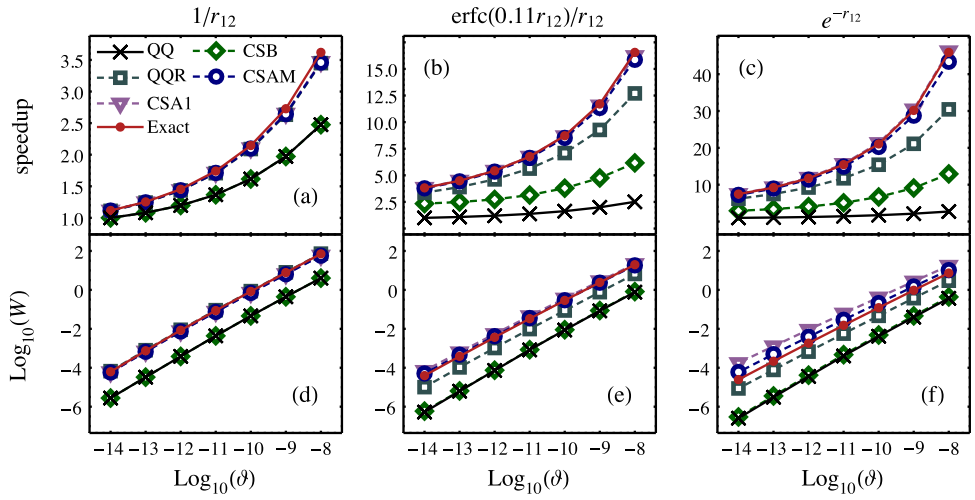


FIG. 1. [(a)-(c)] Speedup (relative to QQ estimates with screening threshold $\vartheta = 10^{-14}$) and [(d)-(f)] decadic logarithm of the error measure W [see Eq. (23)] for the molecular systems Amylose_{32}, Polyethyne_{128}, and $(\text{S}_8)_{20}$ for a series of screening thresholds ϑ. The number of significant shell-quartets and the W values are summed over the three molecules. The rigorous screenings QQ and CSB are compared to the non-rigorous QQR, CSAM, and CSA1 screenings as well as an *a posteriori* exact screening for the operators $1/r_{12}$ [(a) and (d)], $\text{erfc}(0.11 \cdot r_{12})/r_{12}$ [(b) and (e)], and $e^{-r_{12}}$ [(c) and (f)]. Note that while the range of speedup of the upper plots varies for the different operators, the range of the lower error plots is fixed.

The results for the $e^{-r_{12}}$ operator [Figs. 1(c) and 1(f)] are similar to that for $\mathrm{erfc}(0.11 \cdot r_{12})/r_{12}$, except that the speedups are much larger for all but the QQ estimates. Again, we find that CSB significantly outperforms the QQ estimates in efficiency with virtually no error penalty. We see the same trends between the non-rigorous estimates with a wider speed gap between the CSA and QQR estimates and between CSA1 and CSAM. The accuracy of the exact screening and rigorous estimates again increases, for QQR and CSAM it stays about the same, while the error of CSA1 increases slightly in comparison to the values for the $\mathrm{erfc}(0.11 \cdot r_{12})/r_{12}$ operator.

We also evaluated the performance of our estimates with increasing system size. In Fig. 2, the number of significant integral shell-quartets N_{SQ} and the W values are given for a set of linear alkanes ranging from methane to $C_{160}H_{322}$. Here, a fixed screening threshold of $\vartheta = 10^{-10}$ is used.

For the $1/r_{12}$ operator [Figs. 2(a) and 2(d)], we observe a quadratic increase in the number of significant shell-quartets with increasing system size for all estimates and for the exact screening, as expected. The non-rigorous estimates are nearly exact in terms of both significant shell-quartets and error.

For the $\mathrm{erfc}(0.11 \cdot r_{12})/r_{12}$ and $e^{-r_{12}}$ operators, we observe that all estimates except for QQ, including the rigorous CSB, correctly predict a linear scaling number of significant shell-quartets. While the CSB has an increased prefactor compared to the non-rigorous estimates, it shows almost no error increase compared to the QQ bound. Remarkably, for the largest alkane ($C_{160}H_{322}$), the QQ bound predicts 5.14 times as many significant shell-quartets as the CSB bound in the case of the $\mathrm{erfc}(0.11 \cdot r_{12})/r_{12}$ operator. For the $e^{-r_{12}}$ operator, this number increases to 8.09. The non-rigorous bounds are much closer to the exact screening in terms of significant integral shells. The CSA estimates are barely distinguishable from the exact screening, while the QQR estimate discards slightly less integrals. In terms of the accuracy of the non-rigorous estimates, CSA1 performs the worst in both cases, followed by CSAM and then QQR. We also note that the errors for all estimates and operators increase linearly with increasing system size.

In Sec. III B, we focus on the use of our newly introduced CSA estimates in Hartree-Fock SCF calculations, in which only the $1/r_{12}$ operator is needed. We will publish further results on the effectiveness of these estimates for the $\mathrm{erfc}(0.11 \cdot r_{12})/r_{12}$, $e^{-r_{12}}$, and other operators when integrated into the corresponding chemical theories in future work. For now, we point out that in these preliminary results, the errors of all estimates for the short-range operators $\mathrm{erfc}(0.11 \cdot r_{12})/r_{12}$ and $e^{-r_{12}}$ are lower than those seen for the QQR, CSAM, and CSA1 estimates for the $1/r_{12}$ operator. This is encouraging because, as shown in Ref. 3 for the QQR estimate and as we show in the following, these estimates can be well utilized in the efficient calculation of Coulomb and exchange matrices in SCF calculations of large molecules, while incurring minimal error. Thus we expect that our new estimates, CSB, CSAM, and CSA1, will perform very well for these and other operators when used in the respective theories.

B. Exact exchange SCF calculations

In the following, we present Hartree-Fock SCF calculations and compare the performance of the QQ bound with the QQR and our CSAM estimates. All calculations were run on 12 core CPU servers (2x Intel Xeon CPU E5-2620 @ 2.00 GHz). We use one screening threshold for exchange matrix integrals (ϑ_k) and a separate threshold (ϑ_j) for near-field Coulomb matrix integrals, which are calculated separately. Far-field Coulomb-type contributions are calculated using the CFMM[15] method for sufficiently large systems. As an SCF convergence criterion, the error calculated in the direct inversion of the iterative subspace method (DIIS error) is required to be below 10^{-7}, unless otherwise noted. The superposition of atomic densities is used as an initial guess.

We do not present results for the CSB bound because it performs identically to the QQ bound for the systems we tested. We expect that, for the $1/r_{12}$ operator, CSB will only perform

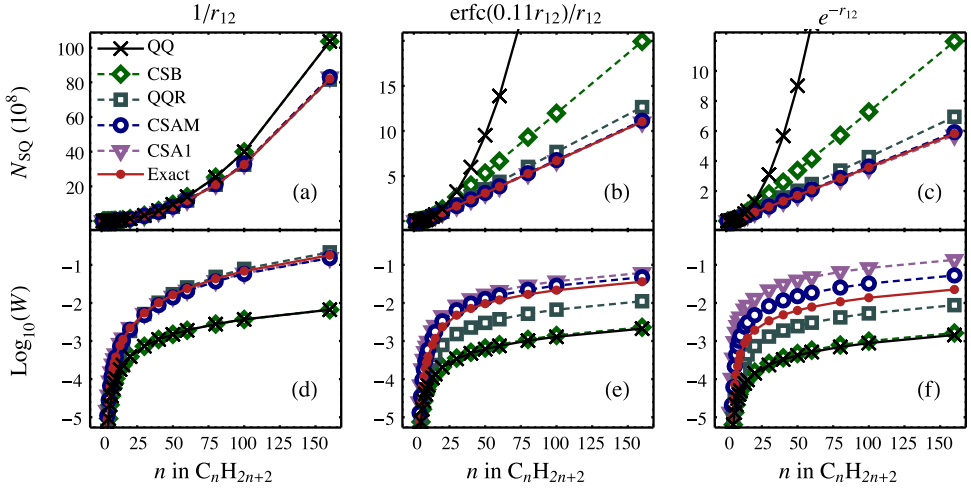


FIG. 2. [(a)-(c)] Number of significant shell-quartets (N_{SQ}) and [(d)-(f)] decadic logarithm of the error measure W [see Eq. (23)] for a series of linear alkanes of increasing length, five integral estimates, and the operators $1/r_{12}$ [(a) and (d)], $\mathrm{erfc}(0.11 \cdot r_{12})/r_{12}$ [(b) and (e)], and $e^{-r_{12}}$ [(c) and (f)]. The screenings were performed with a threshold of $\vartheta = 10^{-10}$ in all cases. Note that while the range of N_{SQ} in the upper plots varies for the different operators, the range of the lower error plots is fixed. In (b) and (c), some QQ points are omitted for large n for clarity. For the largest alkane ($n = 160$), the ratios of significant QQ to CSB shell-quartets are 5.14 and 8.09 for the $\mathrm{erfc}(0.11 \cdot r_{12})/r_{12}$ (b) and $e^{-r_{12}}$ (c) operators, respectively.

better than the QQ bound for unrealistically large systems. This is because of the slow distance-decay of the corresponding two-electron integrals.

1. Threshold dependence

As a first test of our CSA estimates for the Coulomb operator in the context of HF SCF calculations, we compare their performance for various screening thresholds. We vary the exchange screening threshold ϑ_k, while keeping the Coulomb matrix threshold ϑ_j fixed at $\vartheta_j = 10^{-12}$. In Fig. 3, we plot the decadic logarithm of the error vs. the computational time required in the SCF procedure, given as a percentage of the reference time. Here the molecule DNA_2 and the basis set cc-pVTZ were used. The error is calculated as the absolute difference between the converged energies of the reference calculation and the calculation performed with the respective estimates and thresholds. The exchange screening threshold varies from 10^{-9} (top left) to 10^{-12} (bottom right). The calculation with the QQ bound and $\vartheta_k = 10^{-12}$ is used as the reference and is shown as the dotted vertical line at the right of the plot. It has zero error by definition. Here, we see that the non-rigorous distance-including estimates perform much better than the QQ bound. They offer either improved speed with little loss of accuracy for a fixed threshold or vastly improved accuracy with the same calculation speed if used with a more conservative threshold. The exception is the QQR estimate with $\vartheta_k = 10^{-9}$, where four additional iterations are needed for SCF convergence, leading to no speedup compared to the QQ bound calculation with the same threshold. (The CSAM calculation with $\vartheta_k = 10^{-9}$ requires one additional SCF iteration, which can be attributed to small numerical differences.) CSAM gives very similar results to QQR with the exception of the just mentioned better convergence for $\vartheta_k = 10^{-9}$, causing it to be much faster with slightly less accuracy in this case. With $\vartheta_k = 10^{-10}$, CSAM is only slightly faster than QQR with a slight loss of accuracy, while for ϑ_k equal to 10^{-11} or 10^{-12}, it is slightly faster than QQR with no loss in accuracy.

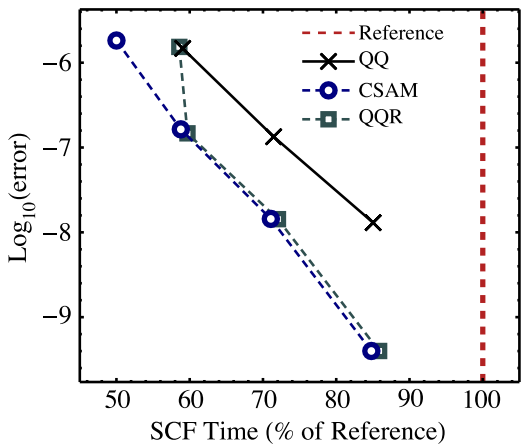


FIG. 3. Threshold dependence of the SCF calculations for the system DNA_2 and the basis set cc-pVTZ. Errors and SCF calculation times are given for the estimates QQ, QQR, and CSAM and exchange matrix integral thresholds ϑ_k varying from 10^{-9} (top left) to 10^{-12} (bottom right), with the combination QQ/10^{-12} used as the reference. The Coulomb matrix integrals are screened with a constant threshold $\vartheta_j = 10^{-12}$. The error is given as the unsigned difference between the converged energies of the reference and the respective estimate/threshold combinations.

2. System and basis set dependence

In the following, we give results for a wide range of chemical systems from the screening test suite and three different basis sets. For each estimate and system tested, we give errors and speedups with respect to reference calculations.

Unless otherwise noted, the calculations are performed with Coulomb and exchange integral thresholds of

TABLE III. Errors (E) and speedups (SU) with respect to reference calculations for HF SCF calculations and the cc-pVDZ basis set. The calculations are performed with $\vartheta_j = \vartheta_k = 10^{-10}$ unless noted otherwise. References are calculated with the QQ estimate and $\vartheta_j = \vartheta_k = 10^{-12}$.

cc-pVDZ basis set						
	QQ		QQR		CSAM	
System	E (μH)	SU	E (μH)	SU	E (μH)	SU
$Amylose_2$	−0.02	1.28	−0.02	1.30	−0.05	1.43
$Amylose_4$	−0.04	1.34	−0.06	1.48	−0.12	1.60
$Amylose_8$	−0.09	1.50	−0.13	1.76	−0.26	1.84
$Amylose_{16}$	−0.19	1.53	−0.27	1.88	−0.54	1.97
$Amylose_{32}$	−0.39	1.57	−0.57	1.88	−1.16	1.98
$Amylose_{48}$	−0.60	1.55	−0.86	1.88	−1.74	1.99
$Amylose_{64}$	−0.80	1.56	−1.14	1.88	−2.32	1.97
Angiotensin	−0.10	1.50	−0.12	1.72	−0.19	1.85
Angiotensin dep.	−0.10	1.50	−0.12	1.74	−0.18	1.87
Angiotensin zw.	−0.10	1.49	−0.13	1.73	−0.18	1.83
CNT_{20}[a]	−0.02	1.18	−0.01	1.17	−0.06	1.24
CNT_{40}[a]	−0.07	1.28	−0.13	1.33	−0.24	1.41
CNT_{80}[a]	−0.22	1.36	−0.31	1.48	−0.44	1.55
$CNT(6,3)_8$[a]	−4.18	1.76	−6.70	2.74	−9.35	2.78
$Diamond_{102}$	−0.20	1.35	−0.37	1.45	−1.70	1.55
DNA_1	−0.03	1.29	−0.03	1.39	−0.06	1.48
DNA_2	−0.13	1.49	−0.15	1.71	−0.23	1.82
DNA_4	−0.33	1.62	−0.41	2.03	−0.56	2.15
DNA_8	−0.71	1.79	−0.92	2.46	−1.25	2.57
DNA_{16}	−1.49	1.92	−1.88	2.74	−2.62	2.81
$Graphite_{24}$	−0.02	1.23	−0.04	1.28	−0.05	1.34
$Graphite_{54}$	−0.15	1.32	−0.25	1.42	−0.34	1.49
$Graphite_{96}$	−0.35	1.37	−0.68	1.54	−0.91	1.61
$(H_2O)_{68}$	−0.08	1.74	−0.05	2.19	0.03	2.32
$(H_2O)_{142}$	−0.23	1.88	−0.14	2.59	0.13	2.72
$(H_2O)_{285}$	−0.54	2.10	−0.27	3.04	0.39	3.19
$(H_2O)_{569}$	−1.17	2.30	−0.55	3.55	0.98	3.70
$(LiF)_{32}$	−0.02	1.13	−0.10	1.18	−0.13	1.20
$(LiF)_{72}$	0.09	1.19	−0.65	1.25	−0.60	1.29[b]
Phtalocyanine c.	−0.05	1.25	−0.09	1.34	−0.16	1.40
$Polyethyne_{64}$	−0.02	1.34	−0.02	1.71	−0.07	1.73
$Polyethyne_{128}$	−0.04	1.44	−0.05	1.91	−0.15	1.92
$Polyyne_{64}$	0.04	1.14	−0.54	1.30[c]	−0.05	1.31
$Polyyne_{1024}$[d]	0.67	1.20	−0.01[e]	1.28[c,e]	0.74[e]	1.31[e]
$(S_8)_5$	−0.10	1.37	−0.10	1.53	−0.09	1.66
$(S_8)_{20}$	−0.54	1.66	−0.45	2.19	−0.29	2.36
Triphenylmethyl	−0.01	1.23	−0.02	1.28	−0.05	1.34

[a]Convergence criterion relaxed to DIIS error below 10^{-6}.
[b]One less SCF iteration than reference calculation required.
[c]One more SCF iteration than reference calculation required.
[d]Convergence criterion relaxed to DIIS error below 10^{-5}.
[e]$\vartheta_j = 10^{-11}$, $\vartheta_k = 10^{-12}$.

$\vartheta_j = \vartheta_k = 10^{-10}$, while the reference calculations are performed with the QQ bound and $\vartheta_j = \vartheta_k = 10^{-12}$.

In Table III, we give the results for the cc-pVDZ basis set for a wide range of chemical systems. The errors (E) are given as differences in converged energies compared to the reference in microhartree (μH), while the speedups (SU) are given as the ratios of the sum of Coulomb and exchange matrix computation times (using the converged density) to that of the reference. In general, the errors are roughly proportional to the system size for each estimate. For almost all systems, the speedups are the lowest with the QQ screening and highest using our CSAM. The only exception is for the CNT_{20} molecule for which QQR is slightly slower than QQ. In some cases (see footnotes of Table III), the calculations using the QQR estimate require an additional SCF iteration for convergence, while in one case [$(LiF)_{72}$] the calculation with CSAM requires one less iteration than the reference. Typically, the size of the errors increases when going from QQ to QQR and from QQR to CSAM. The H_2O and S_8 clusters are exceptions to this rule; here, the non-rigorous bounds give energies closer or just as close to the reference. Most importantly, we find small absolute errors for all of the estimates and systems tested. The largest errors occur for the sizeable $CNT(6,3)_8$ molecule, with absolute errors of 4.18, 6.70, and 9.35 μH with the QQ, QQR, and CSAM estimates, respectively. For all but the largest systems, errors of below 2 μH are observed.

In Table IV, we present analogous results for the cc-pVTZ basis set and smaller subset of the test suite. For this larger basis set, we see increased speedups and typically a larger gap between QQ and the non-rigorous estimates compared to the same systems with a smaller cc-pVDZ basis set. There is almost no speedup gap between QQR and CSAM for this basis set and CSAM is only marginally faster in most cases. Interestingly, the errors incurred using the QQ bound typically increase using the larger basis set, while those of the QQR estimate are not significantly different and those of the CSAM estimate typically decrease. This leads to errors that are very similar across screening methods. Often the CSAM gives less error than QQR, while still being slightly faster. The largest absolute errors are only 1.02 μH ($Graphite_{54}$, CSAM) and 1.01 μH ($Polyyne_{64}$, QQR).

TABLE IV. Errors (E) and speedups (SU) with respect to reference calculations for HF SCF calculations and the cc-pVTZ basis set. The calculations are performed with $\vartheta_j = \vartheta_k = 10^{-10}$. References are calculated with the QQ estimate and $\vartheta_j = \vartheta_k = 10^{-12}$.

cc-pVTZ basis set						
	QQ		QQR		CSAM	
System	E (μH)	SU	E (μH)	SU	E (μH)	SU
$Amylose_8$	−0.10	1.59	−0.11	2.02	−0.12	2.05
Angiotensin	−0.12	1.57	−0.14	1.99[a]	−0.12	2.01[a]
Angiotensin dep.	−0.13	1.59	−0.15	1.98	−0.13	2.01
Angiotensin zw.	−0.12	1.60	−0.15	2.01[a]	−0.12	2.03
CNT_{40}[b]	−0.10	1.33[a]	−0.16	1.44	−0.22	1.49[a]
$Diamond_{102}$	−0.31	1.43	−0.99	1.61[a]	−0.97	1.67
DNA_2	−0.14	1.54	−0.20	1.89	−0.17	1.92
$Graphite_{54}$	−0.22	1.34[a]	−0.38	1.51[c]	−1.02	1.54[c]
$(H_2O)_{68}$	−0.12	1.85	−0.22	2.53	−0.25	2.54
$(LiF)_{32}$	−0.02	1.18	0.02	1.27	0.02	1.27
$Polyyne_{64}$	−0.50	1.18	−1.01	1.42	−0.13	1.42
$(S_8)_{20}$	−0.46	1.72	−0.26	2.33	0.38	2.42
Triphenylmethyl	−0.02	1.29	−0.04	1.38[d]	−0.06	1.42

[a]One more SCF iteration than reference calculation required.
[b]Convergence criterion relaxed to the DIIS error below 10^{-6}.
[c]Two more SCF iterations than reference calculation required.
[d]One less SCF iteration than reference calculation required.

TABLE V. Errors (E) and speedups (SU) with respect to reference calculations for HF SCF calculations and the aug-cc-pVDZ basis set. The calculations are performed with $\vartheta_j = \vartheta_k = 10^{-12}$. References are calculated with the QQ estimate and $\vartheta_j = \vartheta_k = 10^{-14}$.

aug-cc-pVDZ basis set						
	QQ		QQR		CSAM	
System	E (nH)	SU	E (nH)	SU	E (nH)	SU
$Amylose_8$	−0.40	1.21	1.40	1.33	−0.20	1.34
Angiotensin	−1.50	1.18	−1.80	1.25	1.80	1.29
DNA_2	−0.50	1.19	−2.60	1.26	−0.80	1.27
$(H_2O)_{68}$	0.70	1.35	−6.80	1.53	−0.80	1.52
$(LiF)_{32}$	−0.10	1.06	0.70	1.08	1.00	1.09
Triphenylmethyl	−0.20	1.06	−0.30	1.06	−0.10	1.09

In Table V, we give results for a basis set augmented with diffuse basis functions, the aug-cc-pVDZ basis set. In this case, tighter thresholds are required to ensure SCF convergence regardless of the estimate used and thus we choose Coulomb and exchange integral thresholds of $\vartheta_j = \vartheta_k = 10^{-12}$, with reference calculations using the QQ bound and $\vartheta_j = \vartheta_k = 10^{-14}$. Due to the tighter thresholds, the speedups are less than seen for the previous two basis sets. On the other hand, the errors are reduced drastically and are now given in nanohartree (nH). The speedups of the QQR and CSAM estimates are largely similar, with CSAM marginally faster than QQR for all but one system [$(H_2O)_{68}$]. CSAM outperforms QQR in accuracy for all but two systems [angiotensin, $(LiF)_{32}$].

IV. CONCLUSION AND OUTLOOK

We have presented first results for our newly developed distance-including rigorous bound CSB and non-rigorous estimates CSA for three different operators [$1/r_{12}$, $\mathrm{erfc}(0.11 \cdot r_{12})/r_{12}$, and $e^{-r_{12}}$] based on direct integral screenings that are independent of any quantum-chemical theory. In addition we have implemented the CSA estimates in the context of Hartree-Fock SCF calculations and shown that they can be used to speed up integral calculation compared to both the distance-neutral QQ bound and distance-including QQR estimates. The errors incurred are, in almost every case, of the same order of magnitude as for the QQ and QQR estimates and are controllable through the chosen screening threshold.

We expect, based on the results in Sec. III A, that the CSB bound and CSA estimates will perform even better in comparison to the QQ and QQR estimates for operators with much stronger distance-decay than the $1/r_{12}$ Coulomb

operator. In fact, one could already consider CSB as the new default estimate for short-range operators due to its rigorous nature, ease of implementation, low cost, and vastly superior performance compared to the QQ bound. On the other hand, the CSAM estimate has several advantages over the QQR estimate that become more pronounced for short-range operators. One of these advantages is the near-exact nature of the estimate for the short-range operators that we have tested. Another is the fact that the screening itself can be performed more efficiently because the distance-including factor is calculated through two multiplications and one logical operation, whereas in QQR the distance between shell-pairs must be calculated and inserted into an operator function that is relatively expensive to compute. This extra cost is also more important for short-range operators because the ratio of screening time to integral calculation time will naturally go up due to the much larger numbers of insignificant integrals. Regarding the CSA1 estimate, it remains to be seen how it will perform when used in the PreLinK method for short-range exchange DFT calculations. We will compare it to the QQ bound and other distance-including PreLinK methods[5] in this context in future work.

ACKNOWLEDGMENTS

The authors thank Dr. Jörg Kussmann (LMU Munich) for valuable suggestions for improving the manuscript.

Financial support by the DFG funding initiatives Oc35/4-1 and SFB749 (C7) and the Excellence Cluster EXC114 (CIPSM) is acknowledged.

APPENDIX A: PROOF OF EQS. (6) AND (7)

We first prove a more general statement for any pair of two-electron functions. We denote with $\mathcal{E}_2$ the real vector space of all continuous real-valued functions of the positions of two-electrons. We consider the two-electron integral with multiplicative operator G as a map $\langle\cdot,\cdot\rangle$ from the product space $\mathcal{E}_2 \times \mathcal{E}_2$ onto the real numbers,

$$\langle e,f\rangle = \iint e(r_1,r_2)G(r_{12})f(r_1,r_2)dr_1dr_2\,, \tag{A1}$$

where e and f are functions in $\mathcal{E}_2$, and show that $\langle\cdot,\cdot\rangle$ defines an inner product on $\mathcal{E}_2$. We assume that $G(r_{12})$ is positive for $r_{12} > 0$ (positivity property). In order for $\langle\cdot,\cdot\rangle$ to be an inner product, it must satisfy the following requirements for any functions e, f, g in $\mathcal{E}_2$ and any real number λ:

1. $\langle e,f\rangle = \langle f,e\rangle$,
2. $\langle \lambda e,f\rangle = \lambda\langle e,f\rangle$,
3. $\langle e+g,f\rangle = \langle e,f\rangle + \langle g,f\rangle$,
4. $\langle e,e\rangle \geq 0$,
5. $\langle e,e\rangle = 0$ if and only if $e \equiv 0$.

The first requirement follows from the fact that G is multiplicative, while requirements two and three follow from the linearity of integrals. To show four and five, we note the fact that the value of the integral (A1) does not change when we restrict the integration to values of r_1 and r_2, such that $r_{12} \neq 0$. Therefore, we have, using the multiplicity and positivity property of G, that

$$\langle e,e\rangle = \iint e^2(r_1,r_2)G(r_{12})dr_1dr_2 \geq 0 \tag{A2}$$

because the integrand is always positive for $r_{12} \neq 0$. This is requirement four. If $e(r_1, r_2)$ does not vanish for some pair of positions (r_1, r_2), then the continuity of e and the positivity property of G dictate that $\langle e,e\rangle > 0$ must hold. On the other hand, if e vanishes for all possible (r_1, r_2), then $\langle e,e\rangle$ must be zero. These two facts together are requirement five and this ends the proof that $\langle\cdot,\cdot\rangle$ is an inner product.

Because $\langle\cdot,\cdot\rangle$ is an inner product, the Schwarz inequality for any two functions e, f in $\mathcal{E}_2$ holds,

$$\langle e,f\rangle \leq \langle e,e\rangle^{1/2}\langle f,f\rangle^{1/2}\,, \tag{A3}$$

with equality if and only if e and f are linearly dependent. Equations (6) and (7) are a direct consequence of (A3) when e and f are replaced by the corresponding products of Gaussian basis functions (and Mulliken integral notation is used).

APPENDIX B: A NOTE ON THE APPLICABILITY OF THE SCHWARZ INEQUALITY (QQ BOUND) TO INTEGRALS OVER MULTIPLICATIVE, DISTANCE BASED OPERATORS

Proofs for the Schwarz inequality for two-electron integrals over the Coulomb operator $1/r_{12}$ have appeared in various forms in the quantum chemical literature[1,21,22] and more general theorems can be found in mathematical physics textbooks, see, for example, Refs. 23 and 24. Although the inequality has been used, as in this work, for two-electron integrals over other operators,[25,26] this has been done without giving explicit proofs or references to proofs for these cases. We have also been unable to find explicit proofs elsewhere in the literature. Thus, in this section, we aim to clarify the applicability of the Schwarz inequality by giving a simple, sufficient condition for any multiplicative, distance based operator. Our condition is based on the proof for the Coulomb operator given in Ref. 22. Further, we show explicitly that this condition is met not only for the operators used in this work but also for some other important operators in quantum chemical theories.

Let G be some multiplicative operator depending on the distance between the positions of two electrons. In order to show that

$$|\,(\mu\nu|_G|\lambda\sigma)\,| \leq (\mu\nu|_G|\mu\nu)^{1/2}\,(\lambda\sigma|_G|\lambda\sigma)^{1/2}\,, \tag{B1}$$

where

$$(\mu\nu|_G|\lambda\sigma) = \iint \chi_\mu(\mathbf{r}_1)\chi_\nu(\mathbf{r}_1)G(\mathbf{r}_1-\mathbf{r}_2) \times \chi_\lambda(\mathbf{r}_2)\chi_\sigma(\mathbf{r}_2)d\mathbf{r}_1d\mathbf{r}_2, \tag{B2}$$

it suffices to show that the integral $(\mu\nu|_G|\lambda\sigma)$ defines an inner product on the space spanned by all the orbital products $\chi_\mu\chi_\nu$. For this, the same conditions given in Appendix A, with e, f, and g replaced by one-electron functions, must be fulfilled. Again, the first three conditions follow directly from the

multiplicativity of G and the linearity of the integral. However, conditions four and five turn out to be slightly more difficult to show, due to the fact that the integrand on the right-hand side of the equation,

$$(f|f) = \int\int f(\mathbf{r}_1)f(\mathbf{r}_2)G(\mathbf{r}_1 - \mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2, \tag{B3}$$

is not everywhere positive. Here f is any one-electron function for which (B3) converges. Assuming G has a well-defined three-dimensional Fourier transform $\mathcal{F}_G$, given by

$$\mathcal{F}_G(k) = \frac{1}{2\pi}\int G(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}d\mathbf{r}, \tag{B4}$$

we can represent G in terms of its inverse Fourier transform,

$$G(\mathbf{r}) = \int \mathcal{F}_G(k)e^{i\mathbf{k}\cdot\mathbf{r}}d\mathbf{k}. \tag{B5}$$

Here we write $\mathcal{F}_G$ as a function of $k = |\mathbf{k}|$, which is possible because G only depends on $r = |\mathbf{r}|$. This allows us to write (B3) as

$$(f|f) = \int \mathcal{F}_G(k)|\tilde{f}(\mathbf{k})|^2 d\mathbf{k}, \tag{B6}$$

where

$$\tilde{f}(\mathbf{k}) = \int f(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}d\mathbf{r}. \tag{B7}$$

Thus, when $\mathcal{F}_G$ has the positivity property defined in Appendix A, the same arguments given there apply in this case so that conditions four and five follow analogously, and the Schwarz inequality can be used with the operator G. The condition $\mathcal{F}_G(k) > 0$ for $k > 0$ is, according to Bochner's theorem,[27] equivalent to the statement that G is a positive-definite function.

It thus suffices to calculate the Fourier transform of G and check that it is positive for $k > 0$. This can be done, e.g., using the following formula of Grafakos and Teschl:[28]

TABLE VI. Fourier transforms for some important operators in quantum chemical theories. The parameters γ, ω, and α are positive real numbers. All transforms are strictly positive for $k > 0$.

Operator	Fourier transform
$1/r_{12}$	$\frac{1}{2\pi^2 k^2}$
$e^{-\gamma r_{12}}$	$\frac{\gamma}{\pi^2(k^2+\gamma^2)^2}$
$e^{-\gamma r_{12}}/r_{12}$	$\frac{1}{2\pi^2(k^2+\gamma^2)}$
$\mathrm{erf}(\omega r_{12})/r_{12}$	$\frac{1}{2\pi^2 k^2}e^{-\frac{k^2}{4\omega^2}}$
$\mathrm{erfc}(\omega r_{12})/r_{12}$	$\frac{1}{2\pi^2 k^2}(1 - e^{-\frac{k^2}{4\omega^2}})$
$e^{-\alpha r_{12}^2}$	$\frac{1}{8\pi^{3/2}\alpha^{3/2}}e^{-\frac{k^2}{4\alpha}}$

$$\mathcal{F}_G(k) = -\frac{1}{2\pi k}\frac{d}{dk}\tilde{\mathcal{F}}_G(k), \tag{B8}$$

where $\tilde{\mathcal{F}}_G$ is the *one-dimensional* Fourier transform given by

$$\tilde{\mathcal{F}}_G(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(r)e^{-ikr}dr. \tag{B9}$$

Using this equation, we have calculated the Fourier transforms given in Table VI for a variety of important operators in quantum chemical theories, including those used in this work. In each case, the transforms are easily seen to fulfill the necessary conditions and the Schwarz inequality holds for each of these operators.

[1] J. L. Whitten, J. Chem. Phys. **58**, 4496 (1973).
[2] M. Häser and R. Ahlrichs, J. Comput. Chem. **10**, 104 (1989).
[3] S. A. Maurer, D. S. Lambrecht, D. Flaig, and C. Ochsenfeld, J. Chem. Phys. **136**, 144107 (2012).
[4] D. S. Hollman, H. F. Schaefer, and E. F. Valeev, J. Chem. Phys. **142**, 154106 (2015).
[5] M. Beuerle, J. Kussmann, and C. Ochsenfeld, J. Chem. Phys. **146**, 144108 (2017).
[6] C. Hättig, W. Klopper, A. Köhn, and D. P. Tew, Chem. Rev. **112**, 4 (2012).
[7] L. Kong, F. A. Bischoff, and E. F. Valeev, Chem. Rev. **112**, 75 (2012).
[8] J. Heyd, G. E. Scuseria, and M. Ernzerhof, J. Chem. Phys. **118**, 8207 (2003).
[9] S. Ten-no, Chem. Phys. Lett. **398**, 56 (2004).
[10] K. A. Peterson, T. B. Adler, and H. J. Werner, J. Chem. Phys. **128**, 084102 (2008).
[11] A. V. Krukau, O. A. Vydrov, A. F. Izmaylov, and G. E. Scuseria, J. Chem. Phys. **125**, 224106 (2006).
[12] J. G. Brandenburg, E. Caldeweyher, and S. Grimme, Phys. Chem. Chem. Phys. **18**, 15519 (2016).
[13] R. Peverati and D. G. Truhlar, Phys. Chem. Chem. Phys. **14**, 16187 (2012).
[14] J. Kussmann, M. Beer, and C. Ochsenfeld, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **3**, 614 (2013).
[15] C. A. White, B. G. Johnson, P. M. Gill, and M. Head-Gordon, Chem. Phys. Lett. **230**, 8 (1994).
[16] E. Schwegler, M. Challacombe, and M. Head-Gordon, J. Chem. Phys. **106**, 9708 (1997).
[17] C. Ochsenfeld, C. A. White, and M. Head-Gordon, J. Chem. Phys. **109**, 1663 (1998).
[18] C. Ochsenfeld, Chem. Phys. Lett. **327**, 216 (2000).
[19] H.-A. Le and T. Shiozaki, e-print arXiv:1708.05353v1 (2017).
[20] J. Kussmann and C. Ochsenfeld, J. Chem. Phys. **138**, 134114 (2013).
[21] C. C. J. Roothaan, Rev. Mod. Phys. **23**, 69 (1951).
[22] T. Helgaker, P. Jorgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (Wiley, 2000), pp. 431–432.
[23] E. H. Lieb and M. Loss, *Analysis*, 2nd ed. (American Mathematical Society, 2001).
[24] E. H. Lieb and R. Seiringer, *The Stability of Matter in Quantum Mechanics* (Cambridge University Press, 2010).
[25] J. Heyd and G. E. Scuseria, J. Chem. Phys. **121**, 1187 (2004).
[26] T. B. Adler, H.-J. Werner, and F. R. Manby, J. Chem. Phys. **130**, 054106 (2009).
[27] S. Bochner, *Lectures on Fourier Integrals* (Princeton University Press, 1959).
[28] L. Grafakos and G. Teschl, J. Fourier Anal. Appl. **19**, 167 (2013).

4.2 Publication II

Integral partition bounds for fast and effective screening of general one-, two-, and many-electron integrals

T. H. Thompson and C. Ochsenfeld,
J. Chem. Phys. **150**, 044101 (2019).

Abstract:

We introduce tight upper bounds for a variety of integrals appearing in electronic structure theories. These include electronic interaction integrals involving any number of electrons and various integral kernels such as the ubiquitous electron repulsion integrals and the three- and four-electron integrals found in explicitly correlated methods. Our bounds are also applicable to the one-electron potential integrals that appear in great number in quantum mechanical (QM), mixed quantum and molecular mechanical (QM/MM), and semi-numerical methods. The bounds are based on a partitioning of the integration space into balls centered around electronic distributions and their complements. Such a partitioning leads directly to equations for rigorous extents, which we solve for shell pair distributions containing shells of Gaussian basis functions of arbitrary angular momentum. The extents are the first general rigorous formulation we are aware of, as previous definitions are based on the inverse distance operator $1/r_{12}$ and typically only rigorous for simple spherical Gaussians. We test our bounds for six different integral kernels found throughout quantum chemistry, including exponential, Gaussian, and complementary error function based forms. We compare to previously developed estimates on the basis of significant integral counts and their usage in both explicitly correlated second-order Møller-Plesset theory (MP2-F12) and density functional theory calculations employing screened Hartree-Fock exchange.

The following article is reproduced in agreement with its publisher (AIP Publishing LLC) and can be found online at: https://doi.org/10.1063/1.5048491

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

Integral partition bounds for fast and effective screening of general one-, two-, and many-electron integrals

Cite as: J. Chem. Phys. 150, 044101 (2019); doi: 10.1063/1.5048491
Submitted: 15 July 2018 • Accepted: 3 December 2018 •
Published Online: 22 January 2019

View Online Export Citation CrossMark

Travis H. Thompson and Christian Ochsenfeld

AFFILIATIONS

Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), Butenandtstr. 7, D-81377 Munich, Germany and Center for Integrated Protein Science Munich (CIPSM) at the Department of Chemistry, University of Munich (LMU), Butenandtstr. 5-13, D-81377 Munich, Germany

ABSTRACT

We introduce tight upper bounds for a variety of integrals appearing in electronic structure theories. These include electronic interaction integrals involving any number of electrons and various integral kernels such as the ubiquitous electron repulsion integrals and the three- and four-electron integrals found in explicitly correlated methods. Our bounds are also applicable to the one-electron potential integrals that appear in great number in quantum mechanical (QM), mixed quantum and molecular mechanical (QM/MM), and semi-numerical methods. The bounds are based on a partitioning of the integration space into balls centered around electronic distributions and their complements. Such a partitioning leads directly to equations for rigorous extents, which we solve for shell pair distributions containing shells of Gaussian basis functions of arbitrary angular momentum. The extents are the first general rigorous formulation we are aware of, as previous definitions are based on the inverse distance operator $1/r_{12}$ and typically only rigorous for simple spherical Gaussians. We test our bounds for six different integral kernels found throughout quantum chemistry, including exponential, Gaussian, and complementary error function based forms. We compare to previously developed estimates on the basis of significant integral counts and their usage in both explicitly correlated second-order Møller-Plesset theory (MP2-F12) and density functional theory calculations employing screened Hartree-Fock exchange.

Published under license by AIP Publishing. https://doi.org/10.1063/1.5048491

I. INTRODUCTION

The large numbers of integrals that arise in electronic structure theories pose a difficult challenge for accurate computations involving large systems. The most demanding cases are interaction integrals between charge distributions described by basis function pairs, with the prototypical example being the electronic repulsion integrals (ERIs)

$$(\mu\nu|\lambda\sigma)_{C_{12}} = \int d\mathbf{r}_1\, \Omega_{\mu\nu}(\mathbf{r}_1) \int d\mathbf{r}_2\, \Omega_{\lambda\sigma}(\mathbf{r}_2)\, C_{12}\,, \tag{1}$$

where $C_{12} = 1/r_{12}$ is the Coulomb operator, $r_{12} = |\mathbf{r}_1 - \mathbf{r}_2|$, and $\Omega_{\mu\nu} = \chi_\mu\chi_\nu$ is an electronic distribution that is a product of two basis functions in a basis set $\{\chi_\mu\}_{\mu\in\mathbb{B}}$, where $\mathbb{B}$ is a finite set of basis set indices with $|\mathbb{B}| = M$.

Modern theories often require additional forms of such two-electron integrals in which the multiplicative operator C_{12} is replaced by a different function of inter-electronic distances. Examples of such methods include variants of density functional theory (DFT), which require the calculation of screened Hartree-Fock exchange terms[1–4] and explicitly correlated methods employing various correlation factors,[5–8] where two-electron integrals result from the insertion of the resolution of the identity (RI) approximation into the required many-electron integrals. Table I gives a non-exhaustive list of some important interaction functions and where they are found.

In addition, explicitly correlated methods that do not employ RI require many-electron integrals. The general

Published under license by AIP Publishing

TABLE I. Some important operators in quantum chemical theories. The parameters γ, ω, and α are positive real numbers.

Operator	Theory
$e^{-\gamma r_{12}}$	Explicitly correlated theories[7]
$e^{-\gamma r_{12}}/r_{12}$	Explicitly correlated theories[7]
$\mathrm{erfc}(\omega r_{12})/r_{12}$	Range-separated DFT[1]
$e^{-\alpha r_{12}^2}$	Range-separated DFT[4] explicitly correlated theories[6,9–11]
$1/r_{12}$	Ubiquitous

formula for an N-electron integral involving only multiplicative, two-electron operators is given by

$$(k_1|\cdots|k_N)_{\Pi F_{ij}} = \int d\boldsymbol{r}_1\, \Omega_{k_1} \ldots \int d\boldsymbol{r}_N\, \Omega_{k_N} \prod_{\substack{j=1 \\ i<j}}^{N} F_{ij}. \tag{2}$$

Here, F_{ij} is the r_{ij}-dependent interaction function between electrons i and j, and Ω_{k_i} denotes the charge distribution of electron i, with k_i being an index in the set of all distribution indices $\mathbb{D}$. Usually, the set of distributions is made up of all basis function products so that $|\mathbb{D}| = |\mathbb{B} \times \mathbb{B}| = M^2$, and the number of integrals represented by (2) scales as M^{2N}, while the asymptotic scaling of the number of numerically significant integrals with increasing system size is typically not larger than quadratic and often linear, due to the local nature of the basis functions and the decay of the interaction functions with increasing distances between distributions.

In order to take advantage of this inherent sparsity, it is necessary to develop simple upper bounds or reliable estimates that take the decay properties of the integrals into account as best as possible.

The use of the Cauchy-Schwarz inequality,[12,13]

$$|(\mu\nu|\lambda\sigma)_{C_{12}}| \leq Q_{\mu\nu} Q_{\lambda\sigma}, \quad Q_{\mu\nu} = \sqrt{(\mu\nu|\mu\nu)_{C_{12}}}\,, \tag{3}$$

to screen ERI contributions is essential to efficient calculations in quantum chemistry. Inequality (3), which also holds when C_{12} is replaced with any positive definite function,[14] allows for the screening of integrals based on their exponential decay with increasing distances between the two basis functions that describe the electronic distributions (overlap dependence). However, it neglects the operator dependent decay due to increasing distances between the distributions themselves (distance dependence). We denote a bound that captures the asymptotic overlap and distance dependence as *scaling consistent*. In the case of the untransformed ERIs, distance dependence is negligible because of the slow decay of C_{12} and (3) can be considered scaling consistent in practice. However, for most other operators of interest and in the case of transformed integrals such as those occurring in electron correlation theories (see, e.g., Ref. 15), the distance decay is much faster and it becomes essential to take distance dependence into account. Often, it is sufficient to use approximate formulas that are not strict upper bounds, as long as the error is still controllable, as shown by Maurer *et al.*,[16] who introduced the first efficient estimate for the C_{12} operator denoted as QQR. A similar non-rigorous estimate with improved performance for three-center two-electron integrals over C_{12} was introduced by Hollman *et al.*[17] and non-rigorous estimates for the short-range $\mathrm{erfc}(\omega r_{12})/r_{12}$ operator[1,18,19] as well as for some operators involved in F12 theory[20] have also been formulated. Recently,[14] we introduced both non-rigorous estimates and the rigorous combined Schwarz bound (CSB) for screening four-center two-electron integrals over general operators in a way that includes both operator and distance dependence. The very simple CSB uses a distance-including Cauchy-Schwarz type inequality combined with inequality (3) and works well with quickly decaying operators because it is tightest in the limiting cases of strong overlap or strong distance decay. However, the generalization of both the CSB and the Cauchy-Schwarz inequality to many-electron integrals leads to complicated factors that are less practical for screening purposes.

Recently, Barca *et al.*[21–23] presented upper bounds for various operators and three- and four-electron integrals. Their work is based on bounding primitive basis functions by *spherical bounding Gaussians* (SBGs) to get scaling-consistent bounds at the cost of an integral over spherical Gaussians. While the tightest bounds achieved in this way are inseparable, a hierarchy of increasingly separable bounds is obtained using Hölders inequality at the expense of general scaling consistency.

In this work, we present upper bounds that are tight, separable, and scaling-consistent. They can be used with a large variety of integral kernels and are applicable in the many-electron case. These properties allow the formulation of very low cost, truly scaling-consistent screening algorithms with negligible computational and storage requirements.

Our bounds are also easily adapted for use in electronic potential integrals of the type

$$(\mathbf{R}|\mu\nu)_F = \int d\boldsymbol{r}\Omega_{\mu\nu}(\boldsymbol{r})F(|\boldsymbol{r}-\mathbf{R}|), \tag{4}$$

where F is a general distance function and $\mathbf{R}$ is some point in space. These integrals appear in great number in mixed quantum and molecular mechanical (QM/MM)[24–26] and semi-numerical methods.[27–29] Here as well, the bounds account for both the overlap dependence in $\Omega_{\mu\nu}$ and the decay with increasing distance between $\mathbf{R}$ and the center of $\Omega_{\mu\nu}$.

This paper is organized as follows: we present the concepts behind our new upper bounds in Sec. II A and give the rigorous extent equations that result in Sec. II B. The bounds that result are presented in Secs. II C and II D. In Sec. II E, we discuss how to take the effects of density coupling into account during screening, leading in some cases to asymptotic linear scaling even when the number of significant integrals is quadratic. In Sec. II F, we give scaling-consistent screening procedures for two-, three-, and four-electron integrals and discuss the costs involved.

In Sec. III, we quantify the performance of our bounds on the basis of significant integral counts for six different integral kernels and in the context of two quantum chemical theories: explicitly correlated MP2-F12 theory, and

Published under license by AIP Publishing

range-separated DFT, where we use the estimates to speed up the formation of the screened exchange matrix used, e.g., in the HSE06 DFT functional.[2] The results shown in this work are based on two-electron integrals; however, we are currently working on an implementation of the equations for the exact calculation of three- and four-electron integrals,[22,30,31] in order to test our bounds in these cases as well. Because our bounds generalize very straightforwardly to these cases and because the decay properties are the same, i.e., largely dependent on overlap and operator decay, we expect our very promising results for the two-electron case to carry over to the many-electron case. In addition, our one-electron integral bounds are currently being incorporated into our recently published screening scheme[29] for semi-numerical DFT calculations.

II. THEORY

Our upper bounds are based on the intuition that, due to the exponential decay of the typical Gaussian basis functions, it should be possible to restrict integration to spheres that contain all non-negligible contributions, while the sum of the contributions outside the spheres remains below a given screening threshold ϑ.

In this work, we are concerned with integral operators whose kernels are positive, monotonically decreasing functions of the distance between two electrons. We distinguish between two different classes of such operators: *bounded-type* operators, whose kernels are bounded in the limit of zero distance, and *Coulomb-type* operators, which are the product of a bounded-type operator and the Coulomb operator C_{12}. For any operator F_{ij}, we denote its bounded part as F^b_{ij} such that $F_{ij} = F^b_{ij}$ for bounded-type operators and $F_{ij} = F^b_{ij}C_{ij}$ for Coulomb-type operators. This distinction is made because of the simplifications that arise for bounded-type operators. In particular, the overlap factors needed for the three- and four-electron integrals of F12 theory are the same as for the two-electron case when the correlation factor used is a bounded-type operator, which is essentially always the case and is assumed in this work.

A. General ideas

1. Space partitioning

The basic procedure for developing our integral partition bounds (IPBs) is as follows: for an arbitrary charge distribution $\Omega_k, k \in \mathbb{D}$, we choose its center $\mathbf{C}_k$ and consider the ball centered at $\mathbf{C}_k$ with radius R, which we denote $\dot{\mathrm{R}}_k$. This ball is defined by the set equation

$$\dot{\mathrm{R}}_k = \{\boldsymbol{x} \in \mathbb{R}^3 : |\boldsymbol{x} - \mathbf{C}_k| \leq R\}, \tag{5}$$

while its complement, $\mathring{\mathrm{R}}_k$, is given by

$$\mathring{\mathrm{R}}_k = \mathbb{R}^3 \setminus \dot{\mathrm{R}}_k = \{\boldsymbol{x} \in \mathbb{R}^3 : |\boldsymbol{x} - \mathbf{C}_k| > R\}. \tag{6}$$

By integrating separately over the ball and its complement, i.e., by using the identity

$$\int d\boldsymbol{r}\, \Omega_k(\boldsymbol{r}) \ldots = \int_{\dot{\mathrm{R}}_k} d\boldsymbol{r}\, \Omega_k(\boldsymbol{r}) \ldots + \int_{\mathring{\mathrm{R}}_k} d\boldsymbol{r}\, \Omega_k(\boldsymbol{r}) \ldots, \tag{7}$$

we are able to generate equations for determining R large enough that the second term in (7) is always negligible. The equations that arise, which are given in Sec. II B, have very simple dependence on the operators present and the number and type of distributions involved so that the procedure can be applied very generically. By solving them, one obtains *maximal extents*, E_k, for each distribution and guarantees that only the first term of (7) must be taken into account for screening purposes. Such spatially restricted terms can be easily bound from above in a scaling-consistent manner.

A second partitioning can be used to obtain still tighter bounds. By defining *intermediate extents*, e_k, such that $0 < e_k \leq \mathrm{E}_k$, we can partition the first term in (7) according to

$$\int_{\dot{\mathrm{E}}_k} d\boldsymbol{r}\, \Omega_k(\boldsymbol{r}) \ldots = \int_{\dot{e}_k} d\boldsymbol{r}\, \Omega_k(\boldsymbol{r}) \ldots + \int_{\dot{\mathrm{E}}_k \setminus \dot{e}_k} d\boldsymbol{r}\, \Omega_k(\boldsymbol{r}) \ldots. \tag{8}$$

With the right choice of e_k, this allows for a more accurate description of the space closest to the center $\mathbf{C}_k$, where most of the charge is concentrated.

A graphical representation of the spaces that result from our partitioning procedure for the distribution Ω_k is shown in Fig. 1.

2. Separable bounds

Partitioning leads to integrals of the form

$$(k_1; \mathrm{X}_1 | \cdots | k_\mathrm{N}; \mathrm{X}_\mathrm{N})_{\Pi \mathrm{F}_{ij}} = \int_{\mathrm{X}_1} d\boldsymbol{r}_1\, \Omega_{k_1} \ldots \int_{\mathrm{X}_\mathrm{N}} d\boldsymbol{r}_\mathrm{N}\, \Omega_{k_\mathrm{N}} \prod_{\substack{i=1 \\ j<i}}^{\mathrm{N}} \mathrm{F}_{ij}, \tag{9}$$

with integration restricted to subspaces $\mathrm{X}_1, \ldots, \mathrm{X}_\mathrm{N} \subseteq \mathbb{R}^3$. In Appendix D 2, we derive separable bounds for this integral for any set of operators.

Here and in the sequel, we assume that $\mathrm{N} \geq 2$ and that all operators excepting F_{12} are bounded-type, while F_{12} may be either bounded- or Coulomb-type. When this is the case, the bounds can be simplified to the following form:

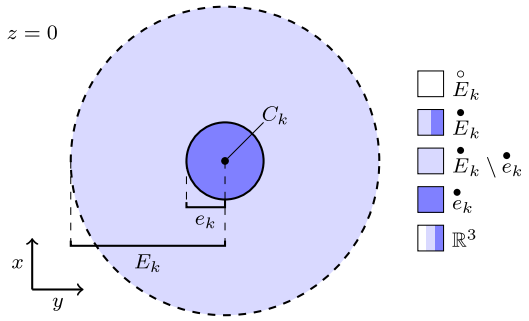


FIG. 1. Representation of the integration spaces used for a distribution Ω_k centered C_k in the *xy*-plane.

Published under license by AIP Publishing

$$|(k_1;\mathrm{X}_1|\cdots|k_\mathrm{N};\mathrm{X}_\mathrm{N})_{\Pi\mathrm{F}_{ij}}| \le \mathcal{U}_{k_1}^{\mathrm{X}_2\mathrm{X}_1}\prod_{j=2}^{\mathrm{N}}\mathcal{S}_{k_j}^{\mathrm{X}_j}\prod_{i<j}\mathcal{D}_{ij}^{\mathrm{X}_i,\mathrm{X}_j}, \tag{10}$$

with

$$\mathcal{U}_{k_j}^{\mathrm{XY}} = \begin{cases}\mathcal{S}_{k_j}^{\mathrm{Y}}, & d(\mathrm{X},\mathrm{Y}) \ge d_{\min}^{\mathrm{F}_{12}}\\ \mathcal{V}_{k_j}^{\mathrm{XY}} & \text{otherwise,}\end{cases} \tag{11}$$

$$\mathcal{D}_{ij}^{\mathrm{XY}} = \begin{cases}\mathrm{F}_{ij}(d(\mathrm{X},\mathrm{Y})), & d(\mathrm{X},\mathrm{Y}) \ge d_{\min}^{\mathrm{F}_{ij}}\\ \mathrm{F}_{ij}^{b}(d(\mathrm{X},\mathrm{Y})) & \text{otherwise,}\end{cases} \tag{12}$$

$$\mathcal{S}_{k_j}^{\mathrm{X}} = \mathcal{S}^{\mathrm{X}}(\Omega_{k_j}) = \int_{\mathrm{X}} d\boldsymbol{r}|\Omega_{k_j}(\boldsymbol{r})|, \tag{13}$$

$$\mathcal{V}_{k_j}^{\mathrm{XY}} = \mathcal{V}^{\mathrm{XY}}(\Omega_{k_j}) = \max_{\boldsymbol{x}\in\mathrm{X}}\int_{\mathrm{Y}} d\boldsymbol{r}\frac{|\Omega_{k_j}(\boldsymbol{r})|}{|\boldsymbol{r}-\boldsymbol{x}|}. \tag{14}$$

The set distance $d(\mathrm{X},\mathrm{Y}) \ge 0$ is defined as

$$d(\mathrm{X},\mathrm{Y}) = d(\mathrm{Y},\mathrm{X}) = \min_{\boldsymbol{x}\in\mathrm{X},\boldsymbol{y}\in\mathrm{Y}}|\boldsymbol{x}-\boldsymbol{y}|,$$

while $d_{\min}^{\mathrm{F}_{ij}}$ is a cutoff distance that is zero for bounded-type and unity for Coulomb-type F_{ij}.

We note that for any balls $\dot{\mathrm{R}}_1$ and $\dot{\mathrm{R}}_2$, with centers $\mathbf{C}_1$ and $\mathbf{C}_2$ and radii R_1 and R_2, the set distance is given by the simple formula

$$d(\dot{\mathrm{R}}_1,\dot{\mathrm{R}}_2) = \max\{0, |\mathbf{C}_1-\mathbf{C}_2| - \mathrm{R}_1 - \mathrm{R}_2\}. \tag{15}$$

B. Extent equations

As shown in Appendix D 1, the maximal extent E_{k_n} for Ω_{k_n} is determined by bounding

$$\mathrm{I}_\mathrm{N}^{\circ k_n} = (\overline{k_1}|\cdots|\overline{k_n};\mathring{\mathrm{R}}_{k_n}|\cdots|\overline{k_\mathrm{N}})_{\Pi\mathrm{F}_{ij}}, \tag{16}$$

where the bars over distribution indices indicate integration over the absolute values of the distributions, and solving the equation $(1+\mathrm{N})\mathrm{I}_\mathrm{N}^{\circ k_n} = \vartheta$ for R_{k_n}. Applying (10) leads to the separated form

$$(\mathrm{N}+1)\mathring{\mathcal{S}}_{k_n}^{\mathring{\mathrm{R}}_{k_n}}\mathcal{H}_{k_1}^{\mathrm{F}_{12}}\Big(\prod_{\substack{j=2\\ j\ne n}}^{\mathrm{N}}\mathcal{S}_{k_j}\Big)\prod_{\substack{j=2\\ i<j}}^{\mathrm{N}}\mathrm{F}_{ij}^{b}(0) = \vartheta, \tag{17}$$

where

$$\mathcal{H}_{k_1}^{\mathrm{F}_{12}} = \begin{cases}\mathcal{S}_{k_1} & \mathrm{F}_{12}\text{ is bounded-type}\\ \mathcal{V}_{k_1} & \mathrm{F}_{12}\text{ is Coulomb-type,}\end{cases} \tag{18}$$

$\mathcal{S}_{k_j} = \mathcal{S}_{k_j}^{\mathbb{R}^3}$, and $\mathcal{V}_{k_1} = \mathcal{V}_{k_1}^{\mathbb{R}^3\mathbb{R}^3}$. Taking the maximal values

$$\mathcal{S}_{k_i} \le \mathcal{S}_{\max} = \max_{i\in\mathbb{D}}\mathcal{S}_i,\ \mathcal{V}_{k_i} \le \mathcal{V}_{\max} = \max_{i\in\mathbb{D}}\mathcal{V}_i$$

leads to a form that is independent of all other distributions

$$(\mathrm{N}+1)\mathring{\mathcal{S}}_{k_n}^{\mathring{\mathrm{R}}_{k_n}}\mathcal{H}_{\max}^{\mathrm{F}_{12}}\mathcal{S}_{\max}^{\mathrm{N}-2}\prod_{\substack{i,j=1\\ j<i}}^{\mathrm{N}}\mathrm{F}_{ij}^{b}(0) = \vartheta. \tag{19}$$

In practice, one can reduce the size of the extents E_{k_n} when Ω_{k_n} is combined in integrals with distributions that have small absolute overlap, as evidenced by (17). For this we also take the minimum values $\mathcal{S}_{\min}$ and $\mathcal{V}_{\min}$ and perform a 10-point logarithmic interpolation between the values $\mathcal{H}_{\max}^{\mathrm{F}_{12}}\mathcal{S}_{\max}^{\mathrm{N}-2}$ and $\mathcal{H}_{\min}^{\mathrm{F}_{12}}\mathcal{S}_{\min}^{\mathrm{N}-2}$, which are typically on the order of unity and $\vartheta^{\mathrm{N}-1}$, respectively. For each point p in this interpolation, we solve (19) with the substitution $\mathcal{H}_{\max}^{\mathrm{F}_{12}}\mathcal{S}_{\max}^{\mathrm{N}-2}\to p$ and use the smallest extents applicable during screening, based on the size of the overlap factors for the particular distribution combination being screened.

Intermediate extents e_{k_n} are obtained as the smallest radii such that

$$\mathring{\mathcal{S}}_{k_n}^{e_{k_n}} \le \alpha\mathcal{S}_{k_n}, \tag{20}$$

with the parameter $\vartheta \ll \alpha < 1$. In this work, we use a fixed value of $\alpha = 0.1$ so that the intermediate extents give the radius outside of which at most 10% of the absolute charge remains. In some cases, the e_{k_n} determined in this way may be larger than E_{k_n}, in which case we set $e_{k_n} = \mathrm{E}_{k_n}$.

C. Scaling consistent bounds

1. Simple bound: IPB1

The bound that results from the single partitionings $\mathbb{R}^3 = \dot{\mathrm{E}}_{k_i}\sqcup\mathring{\mathrm{E}}_{k_i}$ and inequality (10) is

$$|(k_1|\cdots|k_\mathrm{N})_{\Pi\mathrm{F}_{ij}}| \le (1+\mathrm{N}_p)\dot{\mathcal{U}}_{k_1}^{\dot{\mathrm{E}}_{k_2}\dot{\mathrm{E}}_{k_1}}\prod_{j=2}^{\mathrm{N}}\dot{\mathcal{S}}_{k_j}^{\dot{\mathrm{E}}_{k_j}}\prod_{i<j}\dot{\mathcal{D}}_{ij}^{\dot{\mathrm{E}}_{k_i}\dot{\mathrm{E}}_{k_j}}, \tag{21}$$

where $0 \le \mathrm{N}_p \le \mathrm{N}$ is the number of distributions that have been partitioned. N_p is always equal to N in the last screening step, but lower values can be used in earlier screening stages (see Sec. II D).

The *overlap factors* $\dot{\mathcal{S}}_{k_j}^{\dot{\mathrm{E}}_{k_j}}$ and $\dot{\mathcal{V}}_{k_1}^{\dot{\mathrm{E}}_{k_2}\dot{\mathrm{E}}_{k_1}}$ present within (21) account for the overlap decay. For practical reasons, they are bound tightly by $\mathcal{S}_{k_j}$ and $\mathcal{V}_{k_1}$ during screening. The *distance factors* $\dot{\mathcal{D}}_{ij}^{\dot{\mathrm{E}}_{k_i}\dot{\mathrm{E}}_{k_j}}$ account for the distance decay of each operator, leading to general scaling-consistency.

2. Prescreening bound: IPB0

For distributions formed by basis function pairs, $\Omega_k = \Omega_{\mu\nu}$, with centers $\mathbf{C}_{\mu\nu}$ and maximal extents $\mathrm{E}_{\mu\nu}$, the numbers of significant factors $\mathcal{S}_{\mu\nu}$ and $\mathcal{V}_{\mu\nu}$ scale asymptotically linearly with small pre-factors and can thus be easily pre-computed. However, the $\mathcal{U}$ and $\mathcal{D}$ factors can depend on up to four basis-functions, making storage costly for large systems, despite asymptotic linear scaling. To avoid these storage costs and an expensive quadratic screening loop, we bound these factors from above using two-index values while retaining scaling consistency. For the $\mathcal{U}$ factors, we use

$$\dot{\mathcal{U}}_{k_1}^{\dot{\mathrm{E}}_{k_2}\dot{\mathrm{E}}_{k_1}} \le \bar{\mathcal{U}}_{k_1} = \begin{cases}\mathcal{S}_{k_1} & \mathrm{F}_{12}\text{ is bounded-type}\\ \max\{\mathcal{S}_{k_1},\mathcal{V}_{k_1}\} & \mathrm{F}_{12}\text{ is Coulomb-type.}\end{cases}$$

Published under license by AIP Publishing

For the $\mathcal{D}$ factors, we define single-function extents

$$\mathcal{E}_\mu = \max_{\nu\in\mathbb{O}(\mu)}\{|\mathbf{C}_{\mu\nu} - \boldsymbol{\mathcal{C}}_\mu| + \mathrm{E}_{\mu\nu}\}, \tag{22}$$

where $\boldsymbol{\mathcal{C}}_\mu$ is the center of χ_μ and $\mathbb{O}(\mu)$ is the set of functions that overlap significantly with χ_μ. This leads to the inequality

$$d(\dot{\mathrm{E}}_{\mu\nu}, \mathrm{Y}) \geq \max\{d(\dot{\mathcal{E}}_\mu, \mathrm{Y}), d(\dot{\mathcal{E}}_\nu, \mathrm{Y})\} \tag{23}$$

for any $\mathrm{Y} \subseteq \mathbb{R}^3$, and accordingly

$$\mathcal{D}_{ij}^{\dot{\mathrm{E}}_{\mu\nu}\mathrm{Y}} \leq \min\{\mathcal{D}_{ij}^{\dot{\mathcal{E}}_\mu \mathrm{Y}}, \mathcal{D}_{ij}^{\dot{\mathcal{E}}_\nu \mathrm{Y}}\}. \tag{24}$$

This allows the formulation of the bound

$$|(\mu\nu_1|\cdots|\mu\nu_\mathrm{N})_{\Pi\mathrm{F}_{ij}}| \leq (1+\mathrm{N}_p)\bar{\mathcal{U}}_{\mu\nu_1}\prod_{\mathrm{j}=2}^{\mathrm{N}}\mathcal{S}_{\mu\nu_\mathrm{j}}\prod_{\mathrm{i}<\mathrm{j}}\mathcal{D}_{ij}^{\dot{\mathcal{E}}_{\mu_\mathrm{i}}\dot{\mathcal{E}}_{\nu_\mathrm{j}}}, \tag{25}$$

which is not as tight as (21) but separates completely into factors that are easily pre-computed. It is especially useful for determining distance based significance lists $\mathbb{F}_{ij}$ for short-ranged operators F_{ij}, allowing for the design of very efficient screening algorithms that are inherently scaling consistent.

3. Final bound: IPB2

The tightest upper bound results from the partitionings $\mathbb{R}^3 = \dot{e}_{k_i} \sqcup \mathrm{S}_{k_i} \sqcup \mathring{\mathrm{E}}_{k_i}$, where $\mathrm{S}_{k_i} = \dot{\mathrm{E}}_{k_i} \setminus \dot{e}_{k_i}$. The terms containing $\mathring{\mathrm{E}}_{k_i}$ are negligible per construction and what remains is a sum of 2^N terms

$$\sum_{\mathrm{X}_1=\dot{e}_{k_1},\mathrm{S}_{k_1}} \cdots \sum_{\mathrm{X}_\mathrm{N}=\dot{e}_{k_\mathrm{N}},\mathrm{S}_{k_\mathrm{N}}} |(k_1;\mathrm{X}_1|\cdots|k_\mathrm{N};\mathrm{X}_\mathrm{N})_{\Pi\mathrm{F}_{ij}}|. \tag{26}$$

Each term is bound using inequality (10). The needed set distances can be calculated using (15) since, for all relevant $\mathrm{Y} \subseteq \mathbb{R}^3$, $d(\mathrm{Y}, \mathrm{S}_{k_i}) = d(\mathrm{Y}, \dot{\mathrm{E}}_{k_i})$. The overlap factors that result are bound according to

$$\mathcal{S}_{k_j}^{\dot{e}_{k_j}} \leq \mathcal{S}_{k_j} \qquad \mathcal{V}_{k_j}^{\mathrm{X_i}\dot{e}_{k_j}} \leq \mathcal{V}_{k_j}, \tag{27}$$

$$\mathcal{S}_{k_j}^{\mathrm{S}_{k_j}} \leq \alpha\mathcal{S}_{k_j} \qquad \mathcal{V}_{k_j}^{\mathrm{X_i}\mathrm{S}_{k_j}} \leq \mathcal{V}_{k_j}^{\mathbb{R}^3\mathring{e}_{k_j}}. \tag{28}$$

Here, α is the parameter used to determine the intermediate extents.

For N = 2 and bounded-type F_{12}, IPB2 becomes

$$\begin{aligned}|(k_1|\cdots|k_\mathrm{N})_{\Pi\mathrm{F}_{ij}}| \leq (1+\mathrm{N}_p)\mathcal{S}_{k_1}\mathcal{S}_{k_2}\Big[&\mathrm{F}_{12}(d(\dot{e}_{k_1},\dot{e}_{k_2}))\\ &+\alpha\left(\mathrm{F}_{12}(d(\dot{\mathrm{E}}_{k_1},\dot{e}_{k_2}))+\mathrm{F}_{12}(d(\dot{e}_{k_1},\dot{\mathrm{E}}_{k_2}))\right)\\ &+\alpha^2\mathrm{F}_{12}(d(\dot{\mathrm{E}}_{k_1},\dot{\mathrm{E}}_{k_2}))\Big].\end{aligned}$$

In general, the bound can be written as a sum of N + 1 terms $\mathrm{T}^0, \ldots, \mathrm{T}^\mathrm{N}$,

$$|(k_1|\cdots|k_\mathrm{N})_{\Pi\mathrm{F}_{ij}}| \leq (1+\mathrm{N}_p)\sum_{\mathrm{i}=0}^{\mathrm{N}}\mathrm{T}^\mathrm{i}(\alpha), \tag{29}$$

with each T^i containing $\binom{\mathrm{N}}{\mathrm{i}}$ sub-terms and having α dependence given by

$$\mathrm{T}^\mathrm{i}(\alpha) = \mathcal{O}(\alpha^\mathrm{i}). \tag{30}$$

These terms are given explicitly for up to N = 4 in the supplementary material.

4. Approximate bound: aIPB

In the limit $\alpha \to 0$, one has $e_{k_i} \to \mathrm{E}_{k_i}$, and the terms with $i > 0$ in (29) vanish. We have found in the two-electron case that even with rather large values of α, (29) is dominated by the T^0 term, which contains only the intermediate extents e_{k_i}. This has led us to test the approximate upper bound given by

$$|(k_1|\cdots|k_\mathrm{N})_{\Pi\mathrm{F}_{ij}}| \lesssim (1+\mathrm{N}_p)\mathcal{U}_{k_1}^{\dot{e}_{k_2}\dot{e}_{k_1}}\prod_{\mathrm{j}=2}^{\mathrm{N}}\mathcal{S}_{k_\mathrm{j}}^{\dot{e}_{k_\mathrm{j}}}\prod_{\mathrm{i}<\mathrm{j}}\mathcal{D}_{ij}^{\dot{e}_{k_\mathrm{i}}\dot{e}_{k_\mathrm{j}}}, \tag{31}$$

which we have found to be very effective. Our results show that with $\alpha = 0.1$, for which e_{k_i} is typically much smaller than E_{k_i}, (31) gives roughly the same error as IPB2 and *less* error than previously developed rigorous upper bounds for a given screening threshold, while being both simpler and tighter than IPB2.

5. Unit point-charge distributions

The bounds given above are easily applied to integrals containing point charge distributions. When Ω_{k_i} is a point charge at $\mathbf{R}_i$, i.e., $\Omega_{k_i} = \Omega_{\mathbf{R}_i} = \delta(\boldsymbol{r}_i - \mathbf{R}_i)$, one simply defines $\mathbf{C}_{\mathbf{R}_i} = \mathbf{R}_i$, $\mathrm{E}_{\mathbf{R}_i} = e_{\mathbf{R}_i} = 0$, and $\mathcal{S}_{\mathbf{R}_i} = 1$. These distributions do not require partitioning, and N_p is reduced by one for each point-charge present. If F_{12} is Coulomb-type, then some care must be taken for the distributions Ω_{k_1} and Ω_{k_2} because $\mathcal{V}_{k_i}$ becomes infinite for a point charge. If only one is a point charge it should be placed at the position of Ω_{k_2} so that it is not used in the $\mathcal{U}$ factor. This is always possible by switching electron labels $1 \leftrightarrow 2$ if necessary. If both are point charges, then the integral over $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ evaluates to $\mathrm{F}_{12}(|\mathbf{R}_1 - \mathbf{R}_2|)$, which is used in place of $\mathcal{U}_{k_1}^{\mathrm{X}_2\mathrm{X}_1}\mathcal{D}_{12}^{\mathrm{X}_1\mathrm{X}_2}$.

The most prominent example of such integrals is one-electron potential integrals, where one has

$$|(k_1;\mathrm{X}_1|\mathbf{R}_2)_{\mathrm{F}_{12}}| \leq 2\mathcal{U}_{k_1}^{\mathbb{R}^3\mathrm{X}_1}\mathcal{D}_{12}^{\mathrm{X}_1\mathbf{R}_2}. \tag{32}$$

The set distances are again calculated using (15), with $\mathbf{R}_2$ being interpreted as a ball centered at $\mathbf{R}_2$, with zero radius. In the case of Coulomb-type one-electron potential integrals, the extent equation,

$$2\mathcal{V}^{\mathring{\mathrm{R}}_{k_n}}\mathrm{F}_{12}^b(0) = \vartheta, \tag{33}$$

is used instead of (19) since there is only one distribution available to bind the Coulomb singularity.

Published under license by AIP Publishing

D. Intermediate bounds

Intermediate bounds are required in preliminary screening steps, in which only a subset of the N distributions has been specified. These bounds can be obtained by taking the maximal possible values over overlap and/or distance factors in the bounds given above, while a reduction in the value of N_p follows from the formal derivations in Appendix D.

For the overlap factors, one simply uses the maximum values over all distributions, $\mathcal{S}_{\max}$, $\mathcal{V}_{\max}$, and

$$\bar{\mathcal{U}}_{\max} = \begin{cases} \mathcal{S}_{\max} & \mathrm{F}_{12} \text{ is bounded-type} \\ \max\{\mathcal{S}_{max}, \mathcal{V}_{\max}\} & \mathrm{F}_{12} \text{ is Coulomb-type.} \end{cases}$$

Distance factor bounds are given by

$$\mathcal{D}_{ij}^{\mathrm{X_i X_j}} \leq \mathrm{F}_{ij}^{b}(0), \tag{34}$$

and N_p is reduced by one for each j for which (34) is used for all combinations ij. An important example is the one distribution overlap bound

$$|(k_1| \cdots |k_{\mathrm{N}})_{\Pi \mathrm{F}_{ij}}| \leq \mathcal{S}_{k_n} \mathcal{H}_{\max}^{\mathrm{F}_{12}} \mathcal{S}_{\max}^{\mathrm{N}-2} \prod_{\substack{j=2 \\ i<j}}^{\mathrm{N}} \mathrm{F}_{ij}^{b}(0) \tag{35}$$

for any $n \in \{1, \ldots, \mathrm{N}\}$, with

$$\mathcal{H}_{\max}^{\mathrm{F}_{12}} = \begin{cases} \mathcal{S}_{\max} & \mathrm{F}_{12} \text{ is bounded-type} \\ \mathcal{V}_{\max} & \mathrm{F}_{12} \text{ is Coulomb-type.} \end{cases}$$

The distribution Ω_{k_n} can be neglected completely when the right-hand side of (35) is less than ϑ, allowing one to setup a list of significant distributions based on this criterion.

Another important example is the basis-function pair distance bound

$$|(k_1| \cdots |k_{\mathrm{N}})_{\Pi \mathrm{F}_{ij}}| \leq 3 \mathcal{D}_{nm}^{\dot{\mathcal{E}}_\mu \dot{\mathcal{E}}_\nu} \bar{\mathcal{U}}_{\max} \mathcal{S}_{\max}^{\mathrm{N}-1} \prod_{\substack{j=2, i<j \\ (\mathrm{i},\mathrm{j}) \neq (n,m)}}^{\mathrm{N}} \mathrm{F}_{ij}^{b}(0), \tag{36}$$

where $1 \leq n < m \leq \mathrm{N}$ and $\mu, \nu \in \mathbb{B}$ are contained in k_n and k_m, respectively. This bound is used to setup distance-based significance lists for F_{nm}.

E. Density coupling

The significant number of integrals that contain an instance of the long ranged Coulomb operator $1/r_{12}$ scales quadratically in practice. However, coupling to the density matrix $\mathbf{P}$ can cause the scaling of the actual number of significant energy contributions to be linear for systems with significant HOMO-LUMO gaps. This allows, e.g., for the linear scaling calculation of exchange contributions in Hartree-Fock and Hybrid DFT for such systems.[32,33]

This coupling can be used to reduce the scaling to linear for all three- and four-electron integrals of F12 theory, in *both* direct and exchange forms. As an example, we consider the direct $\mathrm{V}_{\mathrm{D}}^{3e}$ and exchange $\mathrm{V}_{\mathrm{X}}^{3e}$ three-electron integral contributions to the so-called V intermediate energy

$$\mathrm{V}_{\mathrm{D}}^{3e} = \sum_{ijk} \langle ijk|\mathrm{C}_{12}\mathrm{F}_{23}|kji\rangle = \sum_{\substack{\mu\nu\lambda\sigma \\ \rho\eta}} \mathrm{P}_{\mu\eta}^{13} \mathrm{P}_{\nu\rho}^{13} \mathrm{P}_{\lambda\sigma}^{22} \langle \mu\lambda\rho|\mathrm{C}_{12}\mathrm{F}_{23}|\nu\sigma\eta\rangle, \tag{37}$$

$$\mathrm{V}_{\mathrm{X}}^{3e} = \sum_{ijk} \langle ijk|\mathrm{C}_{12}\mathrm{F}_{23}|kij\rangle = \sum_{\substack{\mu\nu\lambda\sigma \\ \rho\eta}} \mathrm{P}_{\mu\sigma}^{12} \mathrm{P}_{\lambda\eta}^{23} \mathrm{P}_{\rho\nu}^{13} \langle \mu\lambda\rho|\mathrm{C}_{12}\mathrm{F}_{23}|\nu\sigma\eta\rangle. \tag{38}$$

Here, i, j, and k denote molecular orbitals, F_{23} is a bounded-type correlation factor, and superscripts on density matrix elements highlight which electrons are coupled. Although the number of significant integrals scales quadratically, the number of significant energy contributions scales linearly when density decay is present.

To capture this decay, a final preparatory step involves calculating $\mathrm{P}_{\max} = \max_{\mu \in \mathbb{B}} |\mathrm{P}_{\mu\mu}|$, and creating, for each basis function μ, a density based significance list $\mathbb{P}_{\mathrm{ns}}(\mu)$ based on the criterion

$$|\mathrm{P}_{\mu\nu}| \mathrm{P}_{\max}^{\mathrm{N}-1} \mathcal{H}_{\max}^{\mathrm{F}_{12}} \mathcal{S}_{\max}^{\mathrm{N}-1} \prod_{\substack{\mathrm{i,j}=1 \\ j<i}}^{\mathrm{N}} \mathrm{F}_{ij}^{b}(0) < \vartheta. \tag{39}$$

The lists $\mathbb{P}_{\mathrm{ns}}(\mu)$ cannot be used directly in algorithms that utilize basis function pair symmetry to only calculate integrals over pairs $|\mu\nu)$ in which $\mu \leq \nu$, which reduces the computational work by a factor of roughly 2N. Symmetry respecting lists $\mathbb{P}(\mu)$ are obtained through the procedure given in Algorithm 1. During integral calculation the scaling-consistent and intermediate bounds given above are simply multiplied by the appropriate absolute density matrix elements (or powers of $\mathrm{P}_{\max}$) during screening.

F. Screening procedure and costs

For practical reasons, basis functions that share the same exponents, coefficients, and total angular momenta are grouped into *shells* with indices $\tilde{\mu} \in \mathbb{S}$, and entire sub-tensors of integrals corresponding to shell-tuplets are bound simultaneously. Accordingly, one aims to bound the maximum norm $\|\cdot\|$ over shell-based integral tensors

$$||(\tilde{\mu}_1\tilde{\nu}_1| \cdots |\tilde{\mu}_{\mathrm{N}}\tilde{\nu}_{\mathrm{N}})_{\Pi \mathrm{F}_{ij}}|| = \max_{\substack{\mu_i \in \tilde{\mu}_i, \nu_i \in \tilde{\nu}_i \\ \mathrm{i}=1,\ldots,\mathrm{N}}} |(\mu_1\nu_1| \cdots |\mu_{\mathrm{N}}\nu_{\mathrm{N}})_{\Pi \mathrm{F}_{ij}}|.$$

Algorithm 1. Symmetry respecting density significance lists.

```
for all μ ∈ 𝔹 do
  for all ν ∈ 𝕆(μ) do
    for all λ ∈ ℙ_ns(ν) do
      for all σ ∈ 𝕆(λ) do
        Add σ to ℙ(μ)
      end for
    end for
  end for
end for
```


Published under license by AIP Publishing

Our formulas for calculating the factors needed in the IPBs give values that bound all functions within a shell or shell-pair automatically. They are calculated once per shell or shell-pair using the equations in Appendix A and are directly usable in shell-based bounds, which are obtained through the substitutions $\mu \to \tilde{\mu}$ for any basis function and $\mathbb{B} \to \mathbb{S}$ in the bounds and factors defined above. This allows, e.g., for the construction of shell-based overlap significance lists $\mathbb{O}(\tilde{\mu})$.

For a screening including density coupling, shell-based bounds are obtained by taking maximum values within shell-pairs of density matrix elements

$$\mathrm{P}_{\tilde{\mu}\tilde{\nu}} = \max_{\mu\in\tilde{\mu},\nu\in\tilde{\nu}} |\mathrm{P}_{\mu\nu}|. \tag{40}$$

Shell-based lists $\mathbb{P}_{\mathrm{ns}}(\tilde{\mu})$ are obtained by replacing $|\mathrm{P}_{\mu\nu}|$ with $\mathrm{P}_{\tilde{\mu}\tilde{\nu}}$ in (39). Symmetry respecting lists $\mathbb{P}(\tilde{\mu})$ are obtained with Algorithm 1 using the shell-pair lists $\mathbb{O}(\tilde{\mu})$ and $\mathbb{P}_{\mathrm{ns}}(\tilde{\mu})$.

In Secs. II F 1 and II F 2, we give an overview of the algorithms used for screening integrals over *contracted* Gaussian shell-pairs; the procedure for single-shell distributions that occur in density fitted integrals is obtained through simple modifications. We do not distinguish between Cartesian and spherical shells at this point; the procedure is the same and only the values of the overlap factors and extents change slightly. Additionally, if one wishes to perform a finer screening of primitive contributions, the bounds given above are applicable and can be used with smaller primitive overlap factors and extents. The formulas needed for these various cases are given in Appendix A.

Algorithm 2. Two-electron integral screening procedure.

for all $\tilde{\mu} \in \mathbb{S}$ **do**
  for all $\tilde{\nu} \in \mathbb{O}(\tilde{\mu})$ **do**
    for all $\tilde{\lambda} \in \mathbb{F}_{12}(\tilde{\nu})$ **do**
      if $3\bar{\mathcal{U}}_{\tilde{\mu}\tilde{\nu}}\dot{\mathcal{D}}_{12}^{\dot{\mathcal{E}}_{\tilde{\mu}}\dot{\mathcal{E}}_{\tilde{\lambda}}}\mathcal{S}_{\max} < \vartheta$ **then** Continue
      end If
      for all $\tilde{\sigma} \in \mathbb{O}(\tilde{\lambda})$ **do**
        Calc $\mathrm{T}_{\mathrm{init}} = \dot{\mathcal{D}}_{12}^{\dot{\mathrm{E}}_{\tilde{\mu}\tilde{\nu}}\dot{\mathrm{E}}_{\tilde{\lambda}\tilde{\sigma}}}\dot{\mathcal{U}}_{\tilde{\mu}\tilde{\nu}}^{\dot{\mathrm{E}}_{\tilde{\lambda}\tilde{\sigma}}\dot{\mathrm{E}}_{\tilde{\mu}\tilde{\nu}}}$
        if $3\,\mathcal{S}_{\tilde{\lambda}\tilde{\sigma}}\,\mathrm{T}_{\mathrm{init}} < \vartheta$ **then** Continue
        end if
        Calc $\mathrm{T}^2 = \alpha\,\dot{\mathcal{D}}_{12}^{\mathrm{S}_{\tilde{\mu}\tilde{\nu}}\mathrm{S}_{\tilde{\lambda}\tilde{\sigma}}}\dot{\mathcal{U}}_{\tilde{\mu}\tilde{\nu}}^{\dot{\mathrm{E}}_{\tilde{\lambda}\tilde{\sigma}}\mathrm{S}_{\tilde{\mu}\tilde{\nu}}}$
        Calc $\mathrm{T}_1^1 = \dot{\mathcal{D}}_{12}^{\dot{e}_{\tilde{\mu}\tilde{\nu}}\dot{\mathrm{E}}_{\tilde{\lambda}\tilde{\sigma}}}\dot{\mathcal{U}}_{\tilde{\mu}\tilde{\nu}}^{\dot{e}_{\tilde{\lambda}\tilde{\sigma}}\mathrm{S}_{\tilde{\mu}\tilde{\nu}}}$
        Calc $\mathrm{T}_2^1 = \alpha\,\dot{\mathcal{D}}_{12}^{\dot{\mathrm{E}}_{\tilde{\mu}\tilde{\nu}}\dot{e}_{\tilde{\lambda}\tilde{\sigma}}}\dot{\mathcal{U}}_{\tilde{\mu}\tilde{\nu}}^{\dot{\mathrm{E}}_{\tilde{\lambda}\tilde{\sigma}}\dot{e}_{\tilde{\mu}\tilde{\nu}}}$
        Calc $\mathrm{T}^0 = \dot{\mathcal{D}}_{12}^{\dot{e}_{\tilde{\mu}\tilde{\nu}}\dot{e}_{\tilde{\lambda}\tilde{\sigma}}}\dot{\mathcal{U}}_{\tilde{\mu}\tilde{\nu}}^{\dot{e}_{\tilde{\lambda}\tilde{\sigma}}\dot{e}_{\tilde{\mu}\tilde{\nu}}}$
        $\mathrm{T}_{\mathrm{fin}} = \mathrm{T}^0 + \mathrm{T}_1^1 + \mathrm{T}_2^1 + \mathrm{T}^2$
        if $3\,\mathrm{T}_{\mathrm{fin}}\mathcal{S}_{\tilde{\lambda}\tilde{\sigma}} \geq \vartheta$ **then**
          Calculate integral sub-tensor $(\tilde{\mu}\tilde{\nu}|F_{12}|\tilde{\lambda}\tilde{\sigma})$
        end if
      end for
    end for
  end for
end for

1. Preparatory work

The first step is calculating $\mathcal{S}_{\max}$ for each set of distributions used. If F_{12} is of Coulomb-type, $\mathcal{V}_{\max}$ is also required. For shell-pair distributions, we note that

$$\mathcal{S}_{\max} = \max_{\mu\in\mathbb{B}} \mathcal{S}_{\mu\mu} = 1, \quad \mathcal{V}_{\max} = \max_{\mu\in\mathbb{B}} \mathcal{V}_{\mu\mu} \tag{41}$$

due to the Cauchy-Schwarz inequality and the normalization of the basis functions. Then, the overlap factors, $\mathcal{S}_{\tilde{\mu}\tilde{\nu}}$ and, if needed, $\mathcal{V}_{\tilde{\mu}\tilde{\nu}}$ and $\bar{\mathcal{U}}_{\tilde{\mu}\tilde{\nu}}$, are calculated and, if significant according to bound (35), the maximal extents $\mathrm{E}_{\tilde{\mu}\tilde{\nu}}$ [solve (19) or (33) for each interpolation point p], intermediate extents $e_{\tilde{\mu}\tilde{\nu}}$ [solve (20)], and single-shell extents $\mathcal{E}_{\tilde{\mu}}$ [Eq. (22)] are calculated. Significant pairs are stored as overlap significance lists $\mathbb{O}(\tilde{\mu})$ containing significant partners $\tilde{\nu} \leq \tilde{\mu}$ for each shell $\tilde{\mu}$.

The next preparatory step is the calculation of the shell-pair distance factors $\dot{\mathcal{D}}_{ij}^{\dot{\mathcal{E}}_{\tilde{\mu}}\dot{\mathcal{E}}_{\tilde{\nu}}}$ for each short-ranged operator involved. These values are checked for significance according to bound (36), and the linear scaling number of significant pairs are stored as distance based significance lists $\mathbb{F}_{ij}(\tilde{\mu})$.

Algorithm 3. $\mathrm{V}_{\mathrm{D}}^{3e}$ and $\mathrm{V}_{\mathrm{X}}^{3e}$ screening procedure.

for all $\tilde{\lambda} \in \mathbb{S}$ **do**
  for all $\tilde{\sigma} \in \mathbb{O}(\tilde{\lambda})$ **do**
    for all $\tilde{\rho} \in \mathbb{F}_{23}(\tilde{\sigma})$ **do**
      if $3\mathcal{S}_{\tilde{\lambda}\tilde{\sigma}}\dot{\mathcal{D}}_{23}^{\dot{\mathcal{E}}_{\tilde{\lambda}}\dot{\mathcal{E}}_{\tilde{\rho}}}\mathcal{S}_{\max}\mathcal{V}_{\max}\mathrm{P}_{\max}^3 < \vartheta$ **then**
        Continue
      end if
      for all $\tilde{\eta} \in \mathbb{O}(\tilde{\rho})$ **do**
        Calc $\mathrm{D}_{23} = \dot{\mathcal{D}}_{23}^{\dot{\mathrm{E}}_{\tilde{\lambda}\tilde{\sigma}}\dot{\mathrm{E}}_{\tilde{\rho}\tilde{\eta}}}$
        $\mathrm{P}_{\mathrm{int}} = \max\{\mathrm{P}_{\tilde{\lambda}\tilde{\sigma}}, \mathrm{P}_{\tilde{\lambda}\tilde{\eta}}, \mathrm{P}_{\tilde{\lambda}\tilde{\rho}}, \mathrm{P}_{\tilde{\sigma}\tilde{\eta}}, \mathrm{P}_{\tilde{\sigma}\tilde{\rho}}\}$
        if $3\mathcal{S}_{\tilde{\lambda}\tilde{\sigma}}\mathcal{S}_{\tilde{\rho}\tilde{\eta}}\mathrm{D}_{23}\mathrm{P}_{\mathrm{int}}\mathcal{V}_{\max}\mathrm{P}_{\max}^2 < \vartheta$ **then**
          Continue
        end if
        for all $\tilde{\mu} \in \mathbb{P}(\tilde{\eta})$ **do**
          for all $\tilde{\nu} \in \mathbb{O}(\tilde{\mu})$ **do**
            Calc $\mathrm{D}_{12} = \dot{\mathcal{D}}_{12}^{\dot{\mathrm{E}}_{\tilde{\mu}\tilde{\nu}}\dot{\mathrm{E}}_{\tilde{\lambda}\tilde{\sigma}}}$
            Calc $\mathrm{U}_{\tilde{\mu}\tilde{\nu}} = \dot{\mathcal{U}}_{\tilde{\mu}\tilde{\nu}}^{\dot{\mathrm{E}}_{\tilde{\lambda}\tilde{\sigma}}\dot{\mathrm{E}}_{\tilde{\mu}\tilde{\nu}}}$
            $\mathrm{S} = 4\,\mathrm{U}_{\tilde{\mu}\tilde{\nu}}\mathcal{S}_{\tilde{\lambda}\tilde{\sigma}}\mathcal{S}_{\tilde{\rho}\tilde{\eta}}$
            $\mathrm{D} = \mathrm{D}_{12}\mathrm{D}_{23}$
            $\mathrm{P}_{\mathrm{D}} = \hat{\mathcal{P}}_{\tilde{\mu}\tilde{\nu}}\hat{\mathcal{P}}_{\tilde{\lambda}\tilde{\sigma}}\hat{\mathcal{P}}_{\tilde{\rho}\tilde{\eta}}(\mathrm{P}_{\tilde{\lambda}\tilde{\sigma}}\mathrm{P}_{\tilde{\mu}\tilde{\rho}}\mathrm{P}_{\tilde{\nu}\tilde{\eta}})$
            $\mathrm{P}_{\mathrm{X}} = \hat{\mathcal{P}}_{\tilde{\mu}\tilde{\nu}}\hat{\mathcal{P}}_{\tilde{\lambda}\tilde{\sigma}}\hat{\mathcal{P}}_{\tilde{\rho}\tilde{\eta}}(\mathrm{P}_{\tilde{\mu}\tilde{\lambda}}\mathrm{P}_{\tilde{\nu}\tilde{\rho}}\mathrm{P}_{\tilde{\sigma}\tilde{\eta}})$
            if $\mathrm{SDP}_{\mathrm{D}} \geq \vartheta$ or $\mathrm{SDP}_{\mathrm{X}} \geq \vartheta$ **then**
              Calculate integral sub-tensor
              $\langle\tilde{\mu}\tilde{\lambda}\tilde{\rho}|C_{12}F_{23}|\tilde{\nu}\tilde{\sigma}\tilde{\eta}\rangle$
              Contract with **P** elements
            end if
          end for
        end for
      end for
    end for
  end for
end for

Published under license by AIP Publishing

If density coupling is to be taken into account, the last step is to calculate P_{max}, setup the simple density decay lists $\mathbb{P}_{ns}(\tilde{\mu})$ using (39), and then the symmetry respecting lists $\mathbb{P}(\tilde{\mu})$ according to Algorithm 1. The restriction $\tilde{\nu} \leq \tilde{\mu}$ for $\tilde{\nu} \in \mathbb{P}(\tilde{\mu})$ is not made here or for the distance based lists in general.

All of these factors and lists are inexpensive to compute, depend on at most two shells, and require low pre-factor, asymptotic linear scaling storage. The expressions needed are given in Appendix A. The cost of their calculation is always negligible and timings demonstrating this are given in the supplementary material.

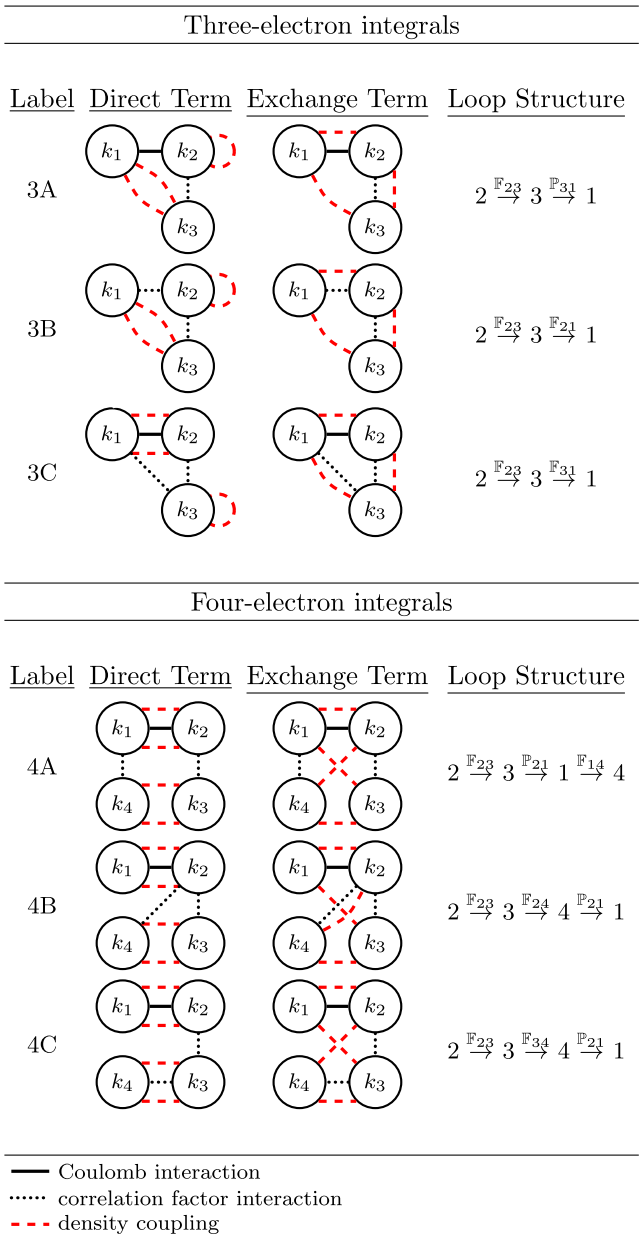


FIG. 2. Graphic representations of the three- and four-electron integrals that arise in F12 theory. Electronic distributions (circles) are coupled through Coulomb interactions (solid lines), the F12 correlation factor (dotted lines), and density elements (red dashed lines). Scaling consistent loop structures are given for each integral type.

2. Integral screening

Using the significance lists described above, it is simple to iterate through a prescreened set of shell combinations that is inherently scaling consistent and takes advantage of shell-pair symmetry. We give the procedure for a general two-electron integral in Algorithm 2 using the bounds IPB0, IPB1 in intermediate forms, and IPB2 as the final bound. This procedure scales linearly for short ranged operators and quadratically for long ranged ones. The procedure for using aIPB is obtained by neglecting T_{init} and the bound containing it and using $T_{fin} = T^0$.

In Algorithm 3, we give an outline of the procedure used for screening the three-electron contributions V_D^{3e} [Eq. (37)] and V_X^{3e} [Eq. (38)] using only the bounds IPB0 and IPB1 for brevity. Here we use the maximal permutation operator $\hat{\mathcal{P}}_{\tilde{\mu}\tilde{\nu}}$ defined as

$$\hat{\mathcal{P}}_{\tilde{\mu}\tilde{\nu}} X_{\tilde{\mu}\tilde{\nu}} = \max\{X_{\tilde{\mu}\tilde{\nu}}, X_{\tilde{\nu}\tilde{\mu}}\}.$$

Density coupling is used to reduce the scaling from quadratic to linear for sparse density matrices. In both direct and exchange forms, the second and third electrons are coupled by the short-range operator F_{23} and the first and third electrons are coupled by at least one density matrix element, which allows for a common scaling-consistent loop structure for both terms. We loop over the distributions in the order $2 \to 3 \to 1$ because this leads to the strongest intermediate bounds for weak density decay.

In Fig. 2, representations of each type of three- and four-electron integral that arise in F12 theory are given that highlight the operator and density couplings involved. Scaling-consistent loop structures are given in a format that details the loop order and the significance lists used to ensure scaling consistency. The same loop structure is applicable to both direct and exchange integrals in each case. Type 3A in Fig. 2 corresponds to the terms V_D^{3e} and V_X^{3e}, which are screened according to Algorithm 3. The other terms are screened analogously using the given loop structures and the proper intermediate and final bounds. The procedure for a four-electron integral (Type 4A) is given in detail in the supplementary material.

III. COMPUTATIONAL DETAILS AND RESULTS

We test our bounds based on their performance in screening two-electron integrals in various settings. We compare to the distance-independent Schwarz bound (QQ), given by

$$||(\tilde{\mu}\tilde{\nu}|\tilde{\lambda}\tilde{\sigma})_F|| \leq Q_{\tilde{\mu}\tilde{\nu}} Q_{\tilde{\lambda}\tilde{\sigma}} ,$$

where $Q_{\tilde{\mu}\tilde{\nu}} = \max_{\mu\in\tilde{\mu},\nu\in\tilde{\nu}} \sqrt{(\mu\nu|\mu\nu)_F}$, and our recently developed distance-including combined Schwarz bound (CSB)[14]

Published under license by AIP Publishing

$$||(\tilde{\mu}\tilde{\nu}|\tilde{\lambda}\tilde{\sigma})_F|| \leq \min\{Q_{\tilde{\mu}\tilde{\nu}}Q_{\tilde{\lambda}\tilde{\sigma}}, M_{\tilde{\mu}\tilde{\lambda}}M_{\tilde{\nu}\tilde{\sigma}}, M_{\tilde{\mu}\tilde{\sigma}}M_{\tilde{\nu}\tilde{\lambda}}\},$$

where $M_{\tilde{\mu}\tilde{\lambda}} = \max_{\mu\in\tilde{\mu},\lambda\in\tilde{\lambda}}\sqrt{(\mu\mu|\lambda\lambda)_F}$.

We use six different integral operators in our analysis: the very long ranged Coulomb operator C_{12}, two shorter ranged operators used in DFT calculations employing screened Hartree-Fock exchange,[2,4] and three very short ranged operators that are needed in explicitly correlated F12 theories employing a Slater-type geminal (STG).

We quantify the range of each of these operators by comparing the distances, R_{max}, at which they fall below the threshold 10^{-12}. The abbreviated names used in this work for each of the six operators are given in Table II, along with their functional forms and R_{max} values. Plots of the operators are shown in Fig. 3.

We implemented and tested our estimates in the quantum chemistry program package FermiONs++[34–36] developed in our group. The molecular structures used in our tests are part of our integral screening test suite, which are available for download online, see Ref. 16.

A. Direct integral screening

In this section, we compare the bounds based on a screening of the entire tensor of four-center two-electron integrals. Entire shell-quartets are neglected when the respective estimates are below the screening threshold. The number of significant quartets remaining is used to measure the speedup possible with each estimate, while the accuracy is measured by the unsigned sum of discarded integrals, which we denote W. This allows for the comparison of the estimates for various operators independent of any particular theory. We also compare to an *a posteriori* exact integral screening in which only integrals are discarded whose exact shell-quartet-maxima are below the threshold.

As test systems we used linear alkanes (C_nH_{2n+2}) of increasing length up to $n = 100$. A screening threshold of 10^{-10} is used for the exact screening. For the other estimates, and for each different operator, we use thresholds chosen such that the W errors are the same as for the exact screening for the largest alkane (the errors for the smaller alkanes are also roughly equal). This is done to ease comparison of the overall effectiveness of each estimate. The decadic logarithms of the thresholds used in each case are given in Table III. For all but the long ranged C_{12} operator, the IPB-type estimates, including aIPB, give smaller W errors and are therefore shifted to looser thresholds.

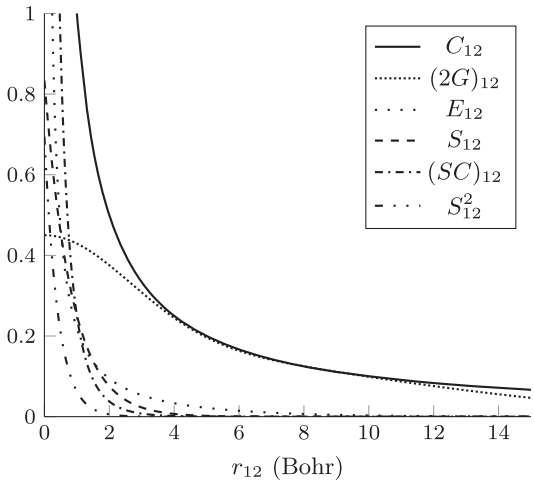


FIG. 3. Plots of the six operators listed in Table II.

The results are given in Fig. 4, where the difference between the number of significant shell quartets N_{SQ} and the exact number N_{SQ}^{Exact} is given for each combination of estimate, operator, and alkane size n. For the C_{12} operator, there is little difference between the estimates for all but the largest linear alkanes, where the IPBs show improvement compared to the QQ and CSB bounds. The small difference is not surprising due to the long-ranged nature of the Coulomb operator. As the range of the operators decreases, we see that the distance-including bounds significantly outperform the QQ bound. The best performance is clearly given by aIPB, while the completely rigorous bound IPB2 performs consistently well and is only slightly outperformed by CSB for the very short ranged operator S^2_{12}. It is of note that although aIPB is only an approximate upper bound, no underestimations were observed in these calculations.

All of the distance-dependent bounds show the correct scaling of the number of integrals with increasing n, which is essentially linear for all but the operator C_{12}. This scaling consistency is evidenced in Table IV, where we give the approximate scaling exponent λ in the equation $N_{SQ}(n) = \mathcal{O}(n^{\lambda})$. The

TABLE II. Operators used in this work and their R_{max} values. R_{max} is the distance (in bohr), after which the operator value falls below 10^{-12}.

Short name	Operator	R_{max}
C_{12}	$1/r_{12}$	10^{12}
$(2G)_{12}$	$0.27e^{-0.075r_{12}^2} + 0.18e^{-0.006r_{12}^2}$	65.7
E_{12}	$0.25\ \mathrm{erfc}(0.11r_{12})/r_{12}$	42.1
S_{12}	$(1/1.2)e^{-1.2r_{12}}$	22.9
$(SC)_{12}$	$(1/1.2)e^{-1.2r_{12}}/r_{12}$	20.4
S^2_{12}	$(1/(1.2)^2)e^{-2.4r_{12}}$	11.4

TABLE III. Decadic logarithms of the thresholds that ensure $W = W_{exact}$ for each estimate-operator pair in the direct screenings in Fig. 4.

	$\log_{10}(\vartheta)$					
	C_{12}	$(2G)_{12}$	E_{12}	S_{12}	$(SC)_{12}$	S^2_{12}
QQ	−8.78	−9.06	−8.79	−8.83	−8.76	−8.92
CSB	−8.78	−9.08	−8.83	−8.87	−8.81	−8.99
IPB0	−9.20	−8.47	−8.12	−7.49	−7.51	−6.98
IPB1	−9.37	−8.63	−8.18	−7.49	−7.51	−6.98
IPB2	−9.37	−8.73	−8.30	−7.54	−7.53	−6.99
aIPB	−9.39	−8.35	−8.35	−7.62	−7.60	−7.02

Published under license by AIP Publishing

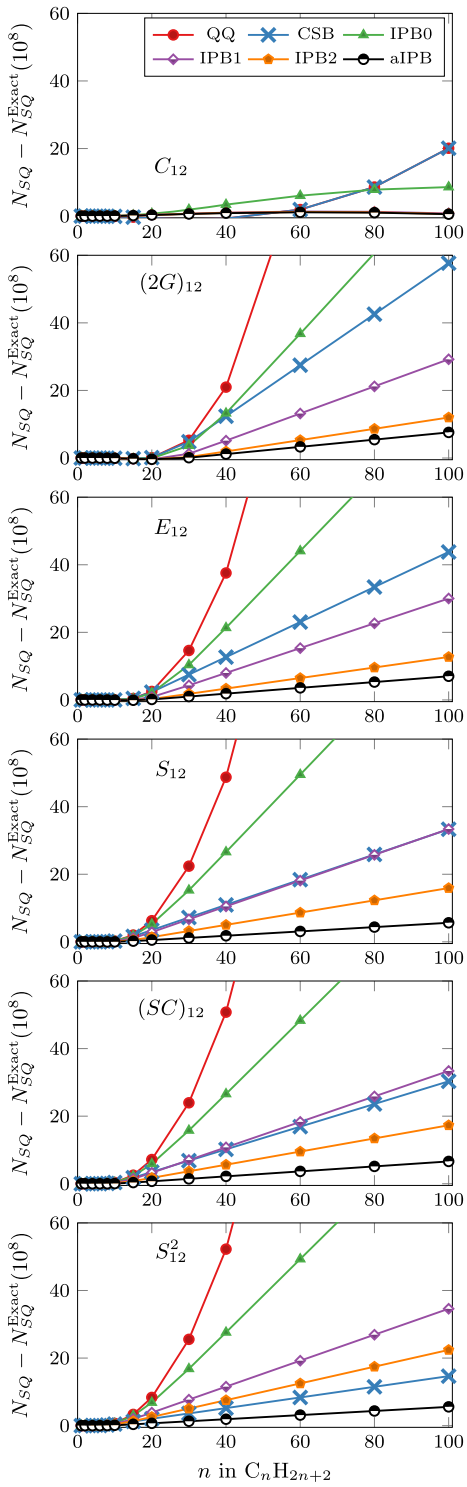


FIG. 4. Difference between estimated and exact number of significant shell quartets for a series of linear alkanes with one to 100 carbon atoms. The basis set aug-cc-pVDZ was used, and the individual integral thresholds were chosen such that for, $C_{100}H_{202}$, the W errors are the same as for an exact screening with a threshold of 10^{-10}.

TABLE IV. Significant integral scaling behaviour for each estimate/operator combination when increasing the number of carbon atoms in a linear alkane from 80 to 100. The scaling exponents λ are calculated according to Eq. (42).

	Scaling exponent λ					
	C_{12}	$(2G)_{12}$	E_{12}	S_{12}	$(SC)_{12}$	S_{12}^2
Exact	1.974	1.132	1.085	1.053	1.048	1.038
QQ	2.036	2.036	2.036	2.034	2.034	2.033
CSB	2.036	1.180	1.113	1.073	1.065	1.048
IPB0	1.947	1.184	1.129	1.099	1.092	1.086
IPB1	1.957	1.166	1.111	1.081	1.074	1.067
IPB2	1.966	1.143	1.092	1.060	1.055	1.049
aIPB	1.967	1.140	1.089	1.056	1.051	1.046

λ values are approximated using the values for the last two alkanes (n = 80 and n = 100) in Fig. 4,

$$\lambda = \frac{\ln(N_{SQ}(100)/N_{SQ}(80))}{\ln(100/80)}. \tag{42}$$

As expected, the Schwarz estimate does not show the correct asymptotic scaling behaviour because it lacks distance dependence.

B. Explicitly correlated MP2-F12 calculations

We tested our bounds and estimates within the context of the explicitly correlated F12 correction[5,37,38] to the second-order Møller-Plesset correlation energy, which improves the description of the wave function at electron-electron cusp regions, significantly accelerating basis set convergence. We employ a Slater type geminal (STG) correlation factor,[7,39] given by the operator S_{12}, and Ten-no's fixed amplitudes[40] within the 3*C variant[41] of the F12 correction. Standard within the F12 method is the use of resolution of the identity approximations (RI) to break down the formally required three- and four-electron integrals into sums of products of two-electron integrals over five different operators. Four of these are the C_{12}, S_{12}, $(SC)_{12}$, and S_{12}^2 operators defined in Table II. The remaining operator is the so-called *double commutator* operator, which, for the explicit STG correlation factor, is just S_{12}^2 multiplied by a constant[7] so that only integrals over the former four operators are needed. We use the CABS+ method of Valeev[42] for constructing the RI basis from an auxiliary basis set $\{\mu'\}$ and density fitting to further break down the four-center two-electron integrals into two- and three-center two-electron integrals, which requires a second auxiliary basis set $\{\mu''\}$. In total, the following sets of integrals are needed:

$$\{(\mu\nu|\lambda'')_{C_{12}}\}\ \{(\mu\nu|\lambda'')_{S_{12}}\}\ \{(\mu\nu|\lambda'')_{S_{12}^2}\}\ \{(\mu\nu|\lambda'')_{(SC)_{12}}\}$$
$$\{(\mu\nu'|\lambda'')_{C_{12}}\}\ \{(\mu\nu'|\lambda'')_{S_{12}}\}\ \{(\mu\nu'|\lambda'')_{S_{12}^2}\}.$$

Due to the use of density fitting, integral calculation is not the most expensive step in the presented calculations. Hence, speedups in integral computation do not lead to large decreases in total calculation times here. For this reason,

Published under license by AIP Publishing

we view these results as a proof of concept, which will be indispensable for some proposed variants of explicitly correlated methods. For example, an efficient implementation of the completely atomic orbital based F12 correction as proposed by Hollman *et al.*[43] would be possible using our estimates. In addition, an explicitly correlated F12 method that forgoes the usual RI insertions and thus requires large numbers of three- and four-electron integrals (as considered, e.g., in Refs. 31, 43, and 44) would greatly benefit from our estimates, as the vast majority of the arising integrals are expected to be insignificant.

The CSB bound cannot be used for these calculations because its formulation for three-center integrals is not straightforward. Therefore, we compared our new bounds IPB2 and aIPB to the QQ bound and the non-rigorous generalized QQR estimate,[14,16] which also takes distance-dependence into account, but uses non-rigorous extents designed for the Coulomb operator. The bounds IPB0 and IPB1 are used for pre-screening purposes but are always less tight than IPB2 and are thus not included in these results. We implemented our bounds within the RI-MP2-F12 routines available in FermiONs++.[34–36]

Because the bounds perform similarly for the long range Coulomb operator C_{12}, we give results based on screening of the integrals over the other three operators. We tested the three molecular systems shown in Fig. 5 and used a series of four screening thresholds ϑ from 10^{-10} to 10^{-7}.

In Table V, we present the screening errors given as the difference to the F12 correction of the reference (no integral screening), and the speedups in integral calculation given as the ratio of the total number of integral quartets of the reference divided by this number for the respective screenings. All three molecules were calculated using the aug-cc-pVDZ basis set in the SCF reference calculation and aug-cc-pVDZ-RI for both auxiliary basis sets.

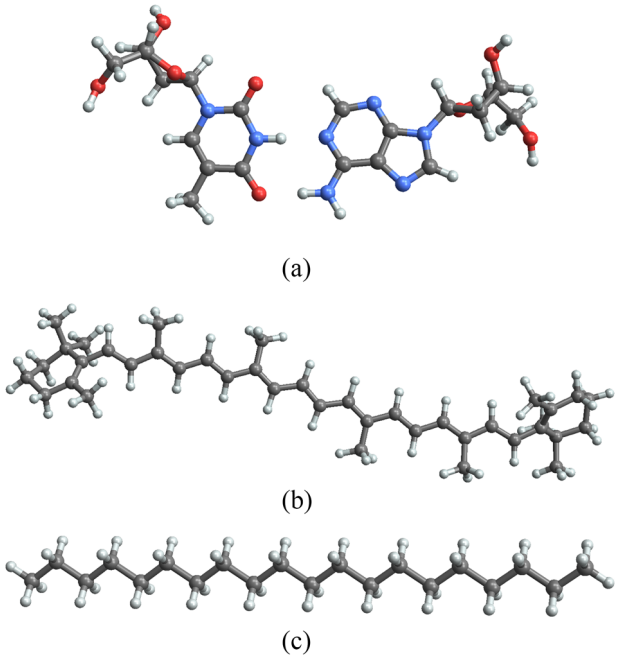


FIG. 5. Molecules used in MP2-F12 calculations. (a) Adenine-thymine base pair, (b) β-carotene, and (c) linear alkane ($C_{20}H_{42}$).

Our bounds show both significant speedups and reduced errors compared to the QQ bound. The approximate bound aIPB outperforms IPB2 in each case, with significantly better

TABLE V. Error in F12 energy correction and speedup (SU) calculated as a number of significant integral quartets of the reference divided by the respective value for each threshold (ϑ) and estimate. The aug-cc-pVDZ basis set was used in combination with the aug-cc-pVDZ-RI basis for both required auxiliary basis sets.

		QQ		IPB2		aIPB		QQR[a]	
	ϑ	Error (H)	SU	Error (H)	SU	Error (H)	SU	Error (H)	SU
Adenine-thymine base pair	10^{-10}	2×10^{-10}	1.09	$<10^{-10}$	1.50	1×10^{-10}	1.76	1×10^{-6}	1.74
	10^{-9}	-7×10^{-9}	1.16	2×10^{-10}	1.62	6×10^{-10}	1.94	3×10^{-6}	1.93
	10^{-8}	1×10^{-7}	1.23	1×10^{-8}	1.76	2×10^{-8}	2.18	2×10^{-5}	2.16
	10^{-7}	9×10^{-6}	1.33	1×10^{-6}	1.95	2×10^{-6}	2.49	6×10^{-5}	2.46
β-carotene	10^{-10}	-1×10^{-10}	1.09	$<10^{-10}$	2.05	4×10^{-10}	2.49	8×10^{-6}	2.50
	10^{-9}	1×10^{-8}	1.14	1×10^{-9}	2.21	-2×10^{-10}	2.73	1×10^{-5}	2.74
	10^{-8}	5×10^{-7}	1.21	1×10^{-8}	2.40	3×10^{-8}	3.03	3×10^{-5}	3.03
	10^{-7}	4×10^{-5}	1.29	9×10^{-7}	2.63	1×10^{-6}	3.41	4×10^{-4}	3.41
Linear alkane $C_{20}H_{42}$	10^{-10}	1×10^{-9}	1.08	1×10^{-10}	1.54	2×10^{-10}	1.84	2×10^{-7}	1.85
	10^{-9}	2×10^{-8}	1.13	9×10^{-10}	1.65	1×10^{-9}	2.00	3×10^{-6}	2.01
	10^{-8}	3×10^{-7}	1.19	7×10^{-9}	1.78	4×10^{-9}	2.22	3×10^{-5}	2.21
	10^{-7}	4×10^{-6}	1.26	8×10^{-8}	1.94	7×10^{-8}	2.48	4×10^{-5}	2.48

[a]Due to increased error and for ease of comparison, the thresholds used in the QQR screening are three orders of magnitude tighter than those given in the ϑ column.

Published under license by AIP Publishing

speedups and very similar errors. The QQR estimate suffers from reduced accuracy for these operators and much tighter thresholds are needed. The error to speedup ratio for QQR is worse than for both IPB2 and aIPB.

C. Screened Hartree-Fock exchange

In this section, we present the results of DFT calculations using the HSE06 functional,[2] which employs a screened Hartree-Fock exchange matrix that is calculated using integrals over the E_{12} operator. We use our estimates to screen such integrals within the LinK[32,33] screening algorithm. This method allows for linear scaling calculations of the exchange matrix for systems with sparse density matrices, which is accomplished by pre-sorting and pre-screening in a way that takes this sparsity into account.

As an SCF convergence criterion, the error calculated in the direct inversion of the iterative subspace method (DIIS error) is required to be below 10^{-7}. A superposition of atomic densities is used as an initial guess.

In Fig. 6, we plot the errors of the respective screenings against the speedups attained for screening thresholds increasing from $\vartheta = 10^{-11}$ to $\vartheta = 10^{-8}$ and using the basis set cc-pVDZ. The system used is a DNA strand of 16 adenine-thymine base pairs, containing 1052 atoms. Here, IPB2 and aIPB perform very similarly and again show reduced error with increased speed compared to the other bounds at each threshold tested.

In Fig. 7, we show the analogous plot for the aug-cc-pVDZ basis set and a chain of 16 amylose sugar molecules, containing 339 atoms. Screening thresholds from $\vartheta = 10^{-12}$ to $\vartheta = 10^{-9}$ are used here. Tighter thresholds are used because SCF convergence is more difficult with larger, more diffuse basis sets. In fact, the calculations using the QQ and CSB bounds with $\vartheta = 10^{-9}$ do not converge so that these points are missing from the plot. These calculations converge in

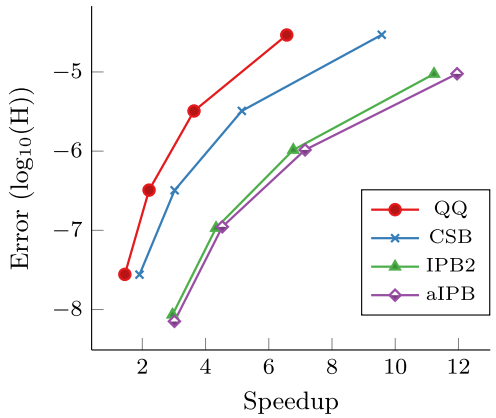


FIG. 6. Speedup vs. SCF energy error for DFT/HSE06 screened exchange matrix calculations for a strand of 16 adenine-thymine DNA base pairs using the basis set cc-pVDZ. Data points correspond to four different screening thresholds from $\vartheta = 10^{-11}$ (bottom left) to $\vartheta = 10^{-8}$ (top right). Error and speedup are calculated with respect to reference calculation using the QQ bound and $\vartheta = 10^{-12}$. The speedup is given as the ratio of significant integral quartets when calculating the screened exchange matrix using the converged density.

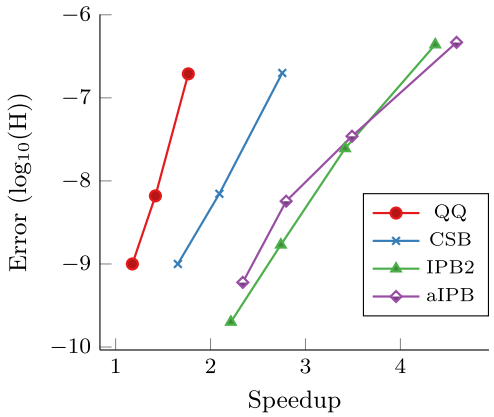


FIG. 7. Speedup vs. SCF energy error for DFT/HSE06 screened exchange matrix calculations for a chain of 16 amylose sugar molecules using the basis set aug-cc-pVDZ. Data points correspond to four different screening thresholds from $\vartheta = 10^{-12}$ (bottom left) to $\vartheta = 10^{-9}$ (top right). The calculations using the QQ and CSB bounds did not converge in combination with $\vartheta = 10^{-9}$ so that these data points are missing. Error and speedup are calculated with respect to reference calculation using the QQ bound and $\vartheta = 10^{-13}$. The speedup is given as the ratio of significant integral quartets when calculating the screened exchange matrix using the converged density.

TABLE VI. Error and speedups for DFT/HSE06 calculations of various systems using our bounds for screening exchange integrals. A screening threshold of $\vartheta = 10^{-9}$ and the basis set cc-pVDZ are used throughout. Errors are given as the difference of final SCF energy to the reference calculation using the QQ bound with a screening threshold of 10^{-12}. Speedups are given as the reference number of significant integral quartets divided by the number for the respective estimate.

		QQ	CSB	IPB2	aIPB
DNA_{16}	Error (μH)	3.21	3.22	1.04	1.04
	Speedup	3.64	5.15	6.78	7.15
$(H_2O)_{569}$	Error (μH)	2.72	2.73	1.00	1.00
	Speedup	5.19	8.21	12.5	13.4
$(S_8)_{20}$	Error (μH)	1.21	1.22	0.38	0.38
	Speedup	3.08	3.81	4.84	5.11
$Amylose_{64}$	Error (μH)	2.02	2.02	0.81	0.79
	Speedup	2.64	3.29	3.82	3.95
Angiotensin	Error (μH)	0.25	0.25	0.09	0.09
	Speedup	2.40	2.67	3.00	3.10
Angiotensin deprotonated	Error (μH)	0.25	0.25	0.09	0.09
	Speedup	2.39	2.66	2.98	3.08
β-carotene	Error (μH)	0.06	0.06	0.02	0.02
	Speedup	1.78	2.62	2.82	2.89
Graphite $C_{96}H_{24}$	Error (μH)	−0.21	−0.21	0.03	0.05
	Speedup	1.90	2.00	2.46	2.53
Polyyne $C_{1024}H_2$	Error (μH)	9.13	12.5	0.81	0.76
	Speedup	1.85	23.9	24.6	25.1

Published under license by AIP Publishing

combination with IPB2 and aIPB, which again screen a larger number of integrals while incurring less error.

We tested various other systems from our screening test suite[16] in combination with the cc-pVDZ basis set and a fixed screening threshold of $\vartheta = 10^{-9}$. These results are given in Table VI. Here, we see the same general trends with speed increasing in the order QQ, CSB, IPB2, and aIPB. The IPB2 and aIPB show very similar errors and both produce between one-third and one-tenth smaller errors than the other two bounds.

IV. CONCLUSION AND OUTLOOK

We have derived completely rigorous, distance-including, and scaling consistent upper bounds (IPB0, IPB1, and IPB2). They can be used for integrals over any number of electrons and involving a wide range of distance based multiplicative operators. The bounds are based on a partitioning of the integration space that allows for the determination of rigorous extents for neglecting contributions far away from distribution centers. They can be used in a wide variety of chemical theories and are particularly well-suited to those involving short-ranged operators, for example, in explicitly correlated theories.

We have also introduced an approximate bound, aIPB, that is almost never an underestimate in practice. The advantage of aIPB over the tightest rigorous bound IPB2 is that it uses a simplified distance factor and is tighter in general, while virtually no error increase is observed. We have found that both IPB2 and aIPB lead to reduced errors and higher speedups for a given screening threshold when compared to the Schwarz inequality bound QQ and our previously developed distance-including combined Schwarz bound CSB.

The high separability of IPB0 allows one to screen very efficiently by determining distance-based significance lists, leading to scaling consistent loop structures. We regard this as a crucial advantage when developing fast and accurate methods for large systems.

We expect that the IPB bounds will be useful in many branches of quantum chemistry due to their generality, simplicity of use, and exceptional tightness. We are currently working on the development of efficient quantum chemical methods based on their application.

SUPPLEMENTARY MATERIAL

See supplementary material for further data as indicated in the text.

ACKNOWLEDGMENTS

The authors thank Matthias Beuerle (LMU Munich), Laurens Peters (LMU Munich), and Henryk Laqua (LMU Munich) for useful discussions.

Financial support from the DFG funding initiatives Oc35/4-1, SFB749 (C7), and the Excellence Cluster EXC114 (CIPSM) is acknowledged.

APPENDIX A: EXPRESSIONS FOR GAUSSIAN BASIS FUNCTIONS

The contracted Gaussian-type orbital (cGTO) basis functions used in practice are given by

$$\chi_\mu^{lm}(\mathbf{r};\mathbf{A}) = \mathrm{N}_\mu^l \sum_{a=1}^{k_\mu} c_{\mu,a}\phi_{\lambda_{\mu,a}}^{lm}(\mathbf{r}_\mathbf{A}). \tag{A1}$$

Here, $\mathbf{A}$ is the center of the cGTO, l is the angular quantum number, N_μ^l is the normalization constant, k_μ is the degree of contraction, and $\mathbf{r}_\mathbf{A} = \mathbf{r} - \mathbf{A}$. In the case of spherical-harmonic Gaussians (SHGs), m is an integer with $-l \le m \le l$, while for Cartesian Gaussians (CGs) $m = (m_x, m_y, m_z)$ is a three-dimensional vector with $m_x + m_y + m_z = l$. The primitive Gaussian-type orbital (pGTO) $\phi_{\lambda_{\mu,a}}^{lm}$ is defined as

$$\phi_{\lambda_{\mu,a}}^{lm}(\mathbf{r}) = \mathrm{Y}_{lm}^{\lambda_{\mu,a}}(\theta,\varphi)\, r^l \exp(-\lambda_{\mu,a} r^2). \tag{A2}$$

For SHGs, on has

$$\mathrm{Y}_{lm}^{\lambda_{\mu,a}}(\theta,\varphi) = \tilde{\mathrm{A}}_{lm}\mathrm{P}_l^{|m|}(\cos\theta)\begin{cases}\cos(m\varphi) & m \ge 0\\ \sin(|m|\varphi) & m < 0,\end{cases}$$

where $\tilde{\mathrm{A}}_{lm} = \frac{\sqrt{2-\delta_{m0}}}{(-1)^m}\sqrt{\frac{(l-|m|)!}{(l+|m|)!}}$ and $\mathrm{P}_l^{|m|}$ is an associated Legendre polynomial. For CGs, one has instead

$$\mathrm{Y}_{lm}^{\lambda_{\mu,a}}(\theta,\varphi) = \mathrm{A}_m^{\lambda_{\mu,a}}(\sin\theta\cos\varphi)^{m_x}(\sin\theta\sin\varphi)^{m_y}(\cos\theta)^{m_z},$$

with $\mathrm{A}_m^{\lambda_{\mu,a}} = \prod_{s=m_x,m_y,m_z}\sqrt{\frac{(2\lambda_{\mu,a})^{s+1/2}}{\Gamma(s+1/2)}}$. In both cases, (A2) can be bound [see Ref. 45 (SHG) and Ref. 23 (CG)] in a radial, m-independent manner using the inequality

$$|\mathrm{Y}_{lm}^{\lambda_{\mu,a}}| \le \mathrm{B}_l^{\lambda_{\mu,a}} = \begin{cases}1 & \text{SHG}\\ \sqrt{\frac{(2\lambda_{\mu,a})^{l+3/2}}{\sum_{i=0}^{2}\Gamma(l\frac{i}{3}]+\frac{1}{2})}} & \text{CG.}\end{cases} \tag{A3}$$

The most complicated distributions we are concerned with are absolute products of cGTOs

$$|\Omega_{\mu\nu}| = |\chi_\mu^{lm}(\mathbf{r};\mathbf{A})\chi_\nu^{jk}(\mathbf{r};\mathbf{B})|, \tag{A4}$$

$$\le \mathrm{N}_\mu^l \mathrm{N}_\nu^j \sum_{a=1}^{k_\mu}\sum_{b=1}^{k_\nu} |c_{\mu,a}c_{\nu,b}|\,|\phi_{\lambda_{\mu,a}}^{lm}\phi_{\lambda_{\nu,b}}^{jk}|. \tag{A5}$$

Inequality (A5) is used to bound integrals over cGTOs by sums of integrals over pGTOs. In the following, we use the definitions

$$I_{ab} = 2\pi \mathrm{B}_l^{\lambda_{\mu,a}}\mathrm{B}_j^{\lambda_{\nu,b}}, \tag{A6}$$

$$p_{ab} = \lambda_{\mu,a} + \lambda_{\nu,b}, \tag{A7}$$

$$\mathrm{K}_{ab} = \exp\left(-\frac{\lambda_{\mu,a}\lambda_{\nu,b}}{p_{ab}}|\mathbf{AB}|^2\right), \tag{A8}$$

$$\mathbf{P}_{ab} = (\lambda_{\mu,a}\mathbf{A} + \lambda_{\nu,b}\mathbf{B})/p_{ab}, \tag{A9}$$

$$R_{ab} = \max(0, R - |\mathbf{P}_{ab} - \mathbf{C}_{\mu\nu}|). \tag{A10}$$

The cGTO product center $\mathbf{C}_{\mu\nu}$ is defined as

Published under license by AIP Publishing

$$\mathbf{C}_{\mu\nu} = \mathbf{P}_{\min},$$

where $\mathbf{P}_{\min}$ is the center $\mathbf{P}_{ab}$ of the primitive pair with the smallest combined exponent p_{ab}. This choice is motivated by the fact that the description of the total distribution at large radial distances is dominated by this primitive pair, and we wish to describe this outer region as accurately as possible.

The bounds given below are in terms of integrals over pGTO pairs. They are combined with (A5) to bound integrals over pairs of cGTOs. If one wishes to bound primitive contributions individually, tighter primitive bounds are obtained by neglecting the terms corresponding to all other primitive pairs in (A5) and letting $R_{ab} \to R$ in the inequalities given below.

1. Absolute tail integrals

To calculate extents, one must be able to calculate the quantities

$$\mathcal{S}^{\mathrm{R}}_{\mu\nu} = \mathring{\mathcal{S}}^{\mathrm{R}_{\mu\nu}}_{\mu\nu} = \int_{|\mathbf{r}-\mathbf{C}_{\mu\nu}|>\mathrm{R}} |\Omega_{\mu\nu}(\mathbf{r})| d\mathbf{r},$$

$$\mathcal{V}^{\mathrm{R}}_{\mu\nu} = \mathring{\mathcal{V}}^{\mathbb{R}^3\mathrm{R}_{\mu\nu}}_{\mu\nu} = \max_{\mathbf{x}\in\mathbb{R}^3} \int_{|\mathbf{r}-\mathbf{C}_{\mu\nu}|>\mathrm{R}} |\Omega_{\mu\nu}(\mathbf{r})| / |\mathbf{r}-\mathbf{x}|.$$

In both cases, we first insert the definition (A1), use inequality (A5), and bound the resulting integrals over primitive pairs.

In the SHG case, if the two contracted functions share the same center, i.e., ($\mathbf{A} = \mathbf{B}$), or if the charge distribution is formed by only one basis function (e.g., for density fitted integrals), integration in spherical coordinates gives

$$\mathcal{S}^{\mathrm{R}}(\phi^{lm}_{\lambda_{\mu,a}}\phi^{jk}_{\lambda_{\nu,b}}) \le \tilde{\mathcal{M}}_{l,j} \mathrm{I}_{ab} \frac{\Gamma(\frac{3+l+j}{2}, p_{ab}\mathrm{R}^2)}{2(p_{ab})^{\frac{3+l+j}{2}}}, \tag{A11}$$

where Γ is the upper incomplete gamma function, and $0 < \tilde{\mathcal{M}}_{l,j} \le 4\pi$ is a factor that depends only on the two non-negative integers i and j and bounds the angular part of the distribution

$$\tilde{\mathcal{M}}_{l,j} = \max_{\substack{l\le m\le l\\ -j\le k\le j}} \{\mathcal{M}_{lm,jk}\},$$

$$\mathcal{M}_{lm,jk} = \int_0^{\pi} d\theta \sin\theta \int_0^{2\pi} d\varphi |\mathrm{Y}_{lm}(\theta,\varphi)\mathrm{Y}_{jk}(\theta,\varphi)|.$$

The numerically calculated values of $\tilde{\mathcal{M}}_{l,j}$ for $0 \le i, j \le 8$ are given in the supplementary material.

In all other cases, we first use the triangle inequality

$$|\mathbf{r}_{\mathbf{A}}| \le |\mathbf{r} - \mathbf{P}_{ab}| + |\mathbf{P}_{ab} - \mathbf{A}|$$

to get

$$|\mathbf{r}_{\mathbf{A}}|^l \le \sum_{t=0}^{l} \binom{l}{t} |\mathbf{P}_{ab} - \mathbf{A}|^{l-t} |\mathbf{r}_{\mathbf{P}_{ab}}|^t. \tag{A12}$$

This bound combined with (A3) and the Gaussian product theorem[46] (GPT) lead to

$$|\phi^{lm}_{\lambda_a}(\mathbf{r}_{\mathbf{A}})\phi^{jk}_{\lambda_b}(\mathbf{r}_{\mathbf{B}})| \le \mathrm{I}_{ab}\mathrm{K}_{ab} \sum_{t=0}^{l}\sum_{u=0}^{j} \mathrm{F}^{lj}_{tu} |\mathbf{r}_{\mathbf{P}_{ab}}|^{t+u} e^{-p_{ab}\mathbf{r}^2_{\mathbf{P}_{ab}}},$$

with

$$\mathrm{F}^{lj}_{tu} = \binom{l}{t}\binom{j}{u} |\mathbf{P}_{ab} - \mathbf{A}|^{l-t} |\mathbf{P}_{ab} - \mathbf{B}|^{j-u}.$$

We use the inclusion

$$\{\mathbf{r} : |\mathbf{r} - \mathbf{C}_{\mu\nu}| > \mathrm{R}\} \subseteq \{\mathbf{r} : |\mathbf{r} - \mathbf{P}_{ab}| > \mathrm{R}_{ab}\} \tag{A13}$$

to shift the center of integration to P_{ab}, simplifying integration and giving

$$\mathcal{S}^{\mathrm{R}}(\phi^{lm}_{\lambda_{\mu,a}}\phi^{jk}_{\lambda_{\nu,b}}) \le \mathrm{I}_{ab}\mathrm{K}_{ab} \sum_{t=0}^{l}\sum_{u=0}^{j} \mathrm{F}^{lj}_{tu} \frac{\Gamma(\frac{3+t+u}{2}, p_{ab}\mathrm{R}^2_{ab})}{(p_{ab})^{\frac{3+t+u}{2}}}. \tag{A14}$$

The potential integral requires the additional application of Newton's Shell theorem,[47,48] which states that for any spherically symmetric function $\mathcal{S}$ and any $\mathbf{r}' \in \mathbb{R}^3$,

$$\int d\mathbf{r} \frac{\mathcal{S}(\mathbf{r})}{|\mathbf{r}' - \mathbf{r}|} = \frac{1}{|\mathbf{r}'|} \int_{|\mathbf{r}|\le|\mathbf{r}'|} d\mathbf{r}\mathcal{S}(\mathbf{r}) + \int_{|\mathbf{r}|>|\mathbf{r}'|} d\mathbf{r} \frac{\mathcal{S}(\mathbf{r})}{|\mathbf{r}|}. \tag{A15}$$

This leads after further analysis to the bound

$$\mathcal{V}^{\mathrm{R}}(\phi^{lm}_{\lambda_{\mu,a}}\phi^{jk}_{\lambda_{\nu,b}}) \le \mathrm{I}_{ab}\mathrm{K}_{ab} \sum_{t=0}^{l}\sum_{u=0}^{j} \mathrm{F}^{lj}_{tu} \frac{\Gamma(1+\frac{t+u}{2}, p_{ab}\mathrm{R}^2_{ab})}{(p_{ab})^{1+\frac{t+u}{2}}}. \tag{A16}$$

2. Absolute integrals

It remains to give the equations for the two absolute integral types $\mathcal{S}_{\mu\nu}$ and $\mathcal{V}_{\mu\nu}$. While both can be obtained by letting $\mathrm{R} \to 0$ in the bounds for $\mathcal{S}^{\mathrm{R}}_{\mu\nu}$ and $\mathcal{V}^{\mathrm{R}}_{\mu\nu}$, tighter bounds are achievable. For $\mathcal{S}_{\mu\nu}$ in the case of SHGs with $\mathbf{A} = \mathbf{B}$, the bound (A11) with $\mathrm{R} = 0$ is best. For the other cases, the integrals over contracted functions are bound by integrals over primitive functions as in (A5), which are bound further using (A3). Afterwards, the Cauchy-Schwarz inequality can be used to get

$$\mathcal{S}(\phi^{lm}_{\lambda_{\mu,a}}\phi^{jk}_{\lambda_{\nu,b}}) \le \sqrt{\mathcal{S}(\phi^{l_+0}_{\lambda_{\mu,a}}\phi^{j_-0}_{\lambda_{\nu,b}})\mathcal{S}(\phi^{l_-0}_{\lambda_{\mu,a}}\phi^{j_+0}_{\lambda_{\nu,b}})}, \tag{A17}$$

$$\mathcal{V}(\phi^{lm}_{\lambda_{\mu,a}}\phi^{jk}_{\lambda_{\nu,b}}) \le \sqrt{\mathcal{V}(\phi^{l_+0}_{\lambda_{\mu,a}}\phi^{j_-0}_{\lambda_{\nu,b}})\mathcal{V}(\phi^{l_-0}_{\lambda_{\mu,a}}\phi^{j_+0}_{\lambda_{\nu,b}})}, \tag{A18}$$

where

$$l_{\pm} = \begin{cases} l & l \text{ even} \\ l \pm 1 & l \text{ odd} \end{cases} \tag{A19}$$

is always a non-negative, even integer and equality holds when l and j are even. For non-negative and even integers r and s, we can use the derivative representation $\phi^{r0}_{\lambda_{\mu,a}} = (-1)^{r/2}\partial^{r/2}_{\lambda_{\mu,a}}\phi^{00}_{\lambda_{\mu,a}}$ and the GPT to get the following expression:

$$\begin{aligned}|\phi^{r0}_{\lambda_{\mu,a}}(\mathbf{r}_{\mathbf{A}})\phi^{s0}_{\lambda_{\nu,b}}(\mathbf{r}_{\mathbf{B}})| &= (-1)^{\frac{r+s}{2}} \mathrm{I}_{ab}\partial^{r/2}_{\lambda_{\mu,a}}\partial^{s/2}_{\lambda_{\nu,b}}\mathrm{K}_{ab}\exp(-p_{ab}\mathbf{r}^2_{\mathbf{P}_{ab}}) \\ &= \mathrm{I}_{ab}\sum_{t=0}^{r/2}\sum_{u=0}^{s/2}(-1)^{t+u}\mathrm{E}^{rs}_{tu}\partial^{t}_{\lambda_{\mu,a}}\partial^{u}_{\lambda_{\nu,b}}\exp(-p_{ab}\mathbf{r}^2_{\mathbf{P}_{ab}}).\end{aligned}$$

The factors E^{rs}_{tu}, whose efficient recursive calculation is outlined in Appendix C, are given by

Published under license by AIP Publishing

$$E^{rs}_{tu} = (-1)^{\frac{r+s}{2}-t-u}\binom{\frac{r}{2}}{t}\binom{\frac{s}{2}}{u}\partial_{\lambda_a}^{\frac{r}{2}-t}\partial_{\lambda_b}^{\frac{s}{2}-u}K_{ab}. \quad \text{(A20)}$$

Performing the integration and then carrying out the derivatives give

$$\mathcal{S}(\phi^{r0}_{\lambda_{\mu,a}}\phi^{s0}_{\lambda_{\nu,b}}) = I_{ab}\sum_{t=0}^{r/2}\sum_{u=0}^{s/2}E^{rs}_{tu}\frac{\Gamma(\frac{3}{2}+t+u)}{(p_{ab})^{\frac{3}{2}+t+u}}, \quad \text{(A21)}$$

$$\mathcal{V}(\phi^{r0}_{\lambda_{\mu,a}}\phi^{s0}_{\lambda_{\nu,b}}) = I_{ab}\sum_{t=0}^{r/2}\sum_{u=0}^{s/2}E^{rs}_{tu}\frac{\Gamma(1+t+u)}{(p_{ab})^{1+t+u}}, \quad \text{(A22)}$$

where (A15) is needed in the second case.

APPENDIX B: SOLVING EXTENT EQUATIONS

Here, we discuss our method for solving extent equations for pairs of cGTO basis functions. Because of the similarity of the three integral bounds (A14), (A16), and (A11), we discuss now only the use of (A14) for calculating extents; the other expressions can be used analogously. Given a target T, which represents the maximal amount of charge that should be left outside of the extent to be solved for and depends, for maximal extents, on the integral threshold ϑ, we must solve the following equation:

$$N_\mu^l N_\nu^j \sum_{a=1}^{k_\mu}\sum_{b=1}^{k_\nu}|c_{\mu,a}c_{\nu,b}|I_{ab}K_{ab}\sum_{t=0}^{l}\sum_{u=0}^{j}F^{lj}_{tu}\frac{\Gamma(\frac{3+t+u}{2}, p_{ab}R^2_{ab})}{(p_{ab})^{\frac{3+t+u}{2}}} = T, \quad \text{(B1)}$$

for R. This is done numerically using Newton-Raphson optimization. The needed derivatives of (B1) can be calculated using

$$\frac{\partial}{\partial R}\frac{\Gamma(x, p_{ab}R^2_{ab})}{p^x_{ab}} = \begin{cases} -2\frac{\exp(-p_{ab}R^2_{ab})}{R_{ab}^{-(2x-1)}} & R_{ab} > 0 \\ 0 & R_{ab} = 0. \end{cases}$$

Because this derivative is never positive, the left side of (B1) is decreasing in R and, if it less than T for R = 0, then (B1) has no solution and the cGTO shell-pair can be neglected.

To obtain an initial guess, we solve a simplified version of (B1) by only retaining the term corresponding to $t = l$, $u = j$ and the smallest exponents λ_a and λ_b, with corresponding coefficients $c_{\mu,\min}$, $c_{\nu,\min}$, combined exponent $p_{\min}$, and factors $I_{\min}$ and $K_{\min}$. The resulting equation is

$$\frac{\Gamma(\frac{3+l+j}{2}, p_{\min}R^2)}{\Gamma(\frac{3+l+j}{2})} = \frac{T|c_{\mu,\min}c_{\nu,\min}|^{-1}(p_{\min})^{\frac{3+l+j}{2}}}{I_{\min}N_\mu^l N_\nu^j K_{\min}\Gamma(\frac{3+l+j}{2})}, \quad \text{(B2)}$$

where we divided both sides of the equation by $\Gamma(\frac{3+l+j}{2})$ and isolated the resulting *regularized* incomplete gamma function given by $\Gamma(x, y)/\Gamma(x)$ on the left hand side.

This allows us to solve for R using routines for inverting the regularized incomplete gamma function.[49] We note that (B2) gives the exact solution for uncontracted functions on the same center. If (B2) has a solution, it is typically a very good guess and Newton-Raphson optimization quickly converges. If (B1) has a solution but (B2) does not, which occurs when the right hand side of (B2) is larger than unity, then the extent is quite small and bracketing solvers[49] can be used effectively.

APPENDIX C: CALCULATION OF E^{rs}_{tu} FACTORS

For fixed λ_a, λ_b, and $AB = |\mathbf{A} - \mathbf{B}|$, we write E^{rs}_{tu} as

$$E^{rs}_{tu} = \binom{\frac{r}{2}}{t}\binom{\frac{s}{2}}{u}\Theta^{\frac{r}{2}-t,\frac{s}{2}-u}, \quad \text{(C1)}$$

where we define

$$\Theta^{kl} = (-1)^{k+l}\partial_a^k\partial_b^l e^{-\frac{\lambda_a\lambda_b}{\lambda_a+\lambda_b}AB^2}. \quad \text{(C2)}$$

For $k = l = 0$, we simply have

$$\Theta^{00} = e^{-\mu_{ab}AB^2} = K_{ab}. \quad \text{(C3)}$$

Using the general Leibniz formula for derivatives, one obtains the recursion formulas

$$\Theta^{k(l+1)} = AB^2\sum_{p=0}^{k}\sum_{q=0}^{l}\binom{k}{p}\binom{l}{q}\Theta^{(k-p)(l-q)}\left[(-1)^{p+q}\partial_{\lambda_a}^p\partial_{\lambda_b}^q\frac{\lambda_a^2}{(\lambda_a+\lambda_b)^2}\right], \quad \text{(C4)}$$

$$\Theta^{(k+1)l} = AB^2\sum_{p=0}^{k}\sum_{q=0}^{l}\binom{k}{p}\binom{l}{q}\Theta^{(k-p)(l-q)}\left[(-1)^{p+q}\partial_{\lambda_a}^p\partial_{\lambda_b}^q\frac{\lambda_b^2}{(\lambda_a+\lambda_b)^2}\right]. \quad \text{(C5)}$$

The remaining derivatives are given by

$$(-1)^{p+q}\partial_{\lambda_a}^p\partial_{\lambda_b}^q\frac{\lambda_a^2}{(\lambda_a+\lambda_b)^2} = \frac{1}{(\lambda_a+\lambda_b)^{p+q}}\sum_{i=0}^{\min(p,2)}\binom{p}{i}(-1)^i \times\frac{(p+q-i+1)!}{(\lambda_a+\lambda_b)^{2-i}}\partial_{\lambda_a}^i\lambda_a^2, \quad \text{(C6)}$$

$$(-1)^{p+q}\partial_{\lambda_a}^p\partial_{\lambda_b}^q\frac{\lambda_b^2}{(\lambda_a+\lambda_b)^2} = \frac{1}{(\lambda_a+\lambda_b)^{p+q}}\sum_{i=0}^{\min(q,2)}\binom{q}{i}(-1)^i \times\frac{(p+q-i+1)!}{(\lambda_a+\lambda_b)^{2-i}}\partial_{\lambda_b}^i\lambda_b^2. \quad \text{(C7)}$$

APPENDIX D: INTEGRAL BOUNDS

We bound the absolute value of the generalized N-electron integral $(k_1;X_1|\cdots|k_N;X_N)_{\Pi F_{ij}}$ using the elementary inequality

$$|(k_1;X_1|\cdots|k_N;X_N)_{\Pi F_{ij}}| \le (\overline{k_1};X_1|\cdots|\overline{k_N};X_N)_{\Pi F_{ij}},$$

where bars over distribution indices indicate integration over the corresponding absolute distribution $\Omega_{\overline{k_n}} = |\Omega_{k_n}|$.

1. Partition bounds

As in Sec. II A 1, we consider a set of charge distributions $\Omega_k, k \in \mathbb{D}$, with centers $\mathbf{C}_k$ and partitionings of three-dimensional space into balls, $\overset{\bullet}{R}_k$, centered at $\mathbf{C}_k$ with radius R_k, and their complements $\overset{\circ}{R}_k$ such that

Published under license by AIP Publishing

$$\int d\boldsymbol{r}\,\Omega_k(\boldsymbol{r})\ldots = \int_{\overset{\bullet}{\mathrm{R}}_k} d\boldsymbol{r}\,\Omega_k(\boldsymbol{r})\ldots + \int_{\overset{\circ}{\mathrm{R}}_k} d\boldsymbol{r}\,\Omega_k(\boldsymbol{r})\ldots. \tag{D1}$$

We define $\mathcal{N} = \{1,\ldots,\mathrm{N}\}$ as the set of electronic indices. We denote with $\mathcal{P}(\mathcal{N})$ the set of all subsets of $\mathcal{N}$. Each subset $\mathcal{I} \in \mathcal{P}(\mathcal{N})$ is characterised by its cardinality $0 \leq |\mathcal{I}| \leq \mathrm{N}$ and the particular set of indices $\mathcal{I} = \{i_1,\ldots,i_{|\mathcal{I}|}\}$, which we assume to be in ascending order.

We define, for any $n \in \mathcal{N}$, the shorthand notation

$$(\overline{k_1}|\cdots|\overset{\bullet}{\overline{k_n}}|\cdots|\overline{k_\mathrm{N}})_{\Pi\mathrm{F}_{ij}} = (\overline{k_1}|\cdots|\overline{k_n};\overset{\bullet}{\mathrm{R}}_{k_n}|\cdots|\overline{k_\mathrm{N}})_{\Pi\mathrm{F}_{ij}},$$
$$(\overline{k_1}|\cdots|\overset{\circ}{\overline{k_n}}|\cdots|\overline{k_\mathrm{N}})_{\Pi\mathrm{F}_{ij}} = (\overline{k_1}|\cdots|\overline{k_n};\overset{\circ}{\mathrm{R}}_{k_n}|\cdots|\overline{k_\mathrm{N}})_{\Pi\mathrm{F}_{ij}}.$$

We choose a subset of indices $\mathcal{I} \in \mathcal{P}(\mathcal{N})$ and partition the integral over $|\Omega_{k_{i_1}}|$, giving

$$(\overline{k_1}|\cdots|\overline{k_\mathrm{N}})_{\Pi\mathrm{F}_{ij}} = (\overline{k_1}|\cdots|\overset{\bullet}{\overline{k_{i_1}}}|\cdots|\overline{k_\mathrm{N}})_{\Pi\mathrm{F}_{ij}} + (\overline{k_1}|\cdots|\overset{\circ}{\overline{k_{i_1}}}|\cdots|\overline{k_\mathrm{N}})_{\Pi\mathrm{F}_{ij}}. \tag{D2}$$

We then partition $|\Omega_{k_{i_2}}|$ in the first term in (D2),

$$(\overline{k_1}|\cdots|\overset{\bullet}{\overline{k_{i_1}}}|\cdots|\overline{k_\mathrm{N}})_{\Pi\mathrm{F}_{ij}} = (\overline{k_1}|\cdots|\overset{\bullet}{\overline{k_{i_1}}}|\overset{\bullet}{\overline{k_{i_2}}}|\cdots|\overline{k_\mathrm{N}})_{\Pi\mathrm{F}_{ij}} + (\overline{k_1}|\cdots|\overset{\bullet}{\overline{k_{i_1}}}|\overset{\circ}{\overline{k_{i_2}}}|\cdots|\overline{k_\mathrm{N}})_{\Pi\mathrm{F}_{ij}}, \tag{D3}$$

and bound the resulting second term

$$(\overline{k_1}|\cdots|\overset{\bullet}{\overline{k_{i_1}}}|\overset{\circ}{\overline{k_{i_2}}}|\cdots|\overline{k_\mathrm{N}})_{\Pi\mathrm{F}_{ij}} \leq (\overline{k_1}|\cdots|\overset{\circ}{\overline{k_{i_2}}}|\cdots|\overline{k_\mathrm{N}})_{\Pi\mathrm{F}_{ij}}. \tag{D4}$$

The notation used here highlights the integration spaces used for the distributions in the middle but is meant to make no restrictions on their positions within the integral. In particular, there may be indices between i_1 and i_2 for which the corresponding distributions have not been partitioned. Inequality (D4) is used to obtain a term that depends on the radius $\mathrm{R}_{k_{i_2}}$ only and is very tight when $\overset{\bullet}{\mathrm{R}}_{k_{i_1}}$ contains most of the absolute charge of $\Omega_{k_{i_1}}$.

The partitioning and bounding procedure of Eqs. (D3) and (D4) is repeated successively for the remaining indices in $\mathcal{I}$ leading to the following result:

$$(\overline{k_1}|\cdots|\overline{k_\mathrm{N}})_{\Pi\mathrm{F}_{ij}} \leq (\overline{k_1}|\cdots|\overset{\bullet}{\overline{k_{i_1}}}|\cdot|\overset{\bullet}{\overline{k_{i_{|\mathcal{I}|}}}}|\cdots|\overline{k_\mathrm{N}})_{\Pi\mathrm{F}_{ij}} + \sum_{n=1}^{|\mathcal{I}|}(\overline{k_1}|\cdots|\overset{\circ}{\overline{k_{i_n}}}|\cdots|\overline{k_\mathrm{N}})_{\Pi\mathrm{F}_{ij}} \tag{D5}$$

$$= \mathrm{I}_\mathrm{N}^{\bullet\mathcal{I}} + \sum_{n=1}^{|\mathcal{I}|}\mathrm{I}_\mathrm{N}^{\circ k_{i_n}}, \tag{D6}$$

where the notation $|\overset{\bullet}{\overline{k_{i_1}}}|\cdot|\overset{\bullet}{\overline{k_{i_{|\mathcal{I}|}}}}|$ implies the partitioning of all indices in $\mathcal{I}$. If each term in (D6) is less than $\vartheta/(|\mathcal{I}|+1)$, it follows that $(\overline{k_1}|\cdots|\overline{k_\mathrm{N}})_{\Pi\mathrm{F}_{ij}} \leq \vartheta$ and the integral can be neglected.

An important special case of (D6) occurs for $\mathcal{I} = \mathcal{N}$, $|\mathcal{I}| = \mathrm{N}$, giving the fully partitioned form

$$(\overline{k_1}|\cdots|\overline{k_\mathrm{N}})_{\Pi\mathrm{F}_{ij}} \leq (\overset{\bullet}{\overline{k_1}}|\cdots|\overset{\bullet}{\overline{k_\mathrm{N}}})_{\Pi\mathrm{F}_{ij}} + \sum_{n=1}^{\mathrm{N}}(\overline{k_1}|\cdots|\overset{\circ}{\overline{k_n}}|\cdots|\overline{k_\mathrm{N}})_{\Pi\mathrm{F}_{ij}} \tag{D7}$$

$$= \mathrm{I}_\mathrm{N}^{\bullet} + \sum_{n=1}^{\mathrm{N}}\mathrm{I}_\mathrm{N}^{\circ k_n}. \tag{D8}$$

This form is used for determining radii R_{k_n} such that the *tail integrals* $\mathrm{I}_\mathrm{N}^{\circ k_n}$ are less than $(\vartheta/(\mathrm{N}+1))$ or equivalently such that $((\mathrm{N}+1)\mathrm{I}_\mathrm{N}^{\circ k_n} \leq \vartheta)$. This results in maximal extents E_{k_n} for each distribution. Because $|\mathcal{I}| \leq \mathrm{N}$, these extents also ensure that the tail integrals are negligible in the more general bound (D6). They are then used in the *principle integrals* $\mathrm{I}_\mathrm{N}^{\bullet\mathcal{I}}$, which can be bound straightforwardly in a distant-dependent manner due to the restriction to finite balls around the distribution centers. The different forms of $\mathcal{I}$ lead to the final and intermediate bounds of Secs. II C and II D.

2. Separable bounds

The principle and tail integrals given above are not separable and are difficult to compute directly. Here we present our *ansatz* for obtaining separable bounds containing easily computable factors. This procedure is independent of the partitioning, and it is the combination of the two that gives an efficient and scaling-consistent bound.

We write the integral $(\overline{k_1};\mathrm{X}_1|\cdots|\overline{k_\mathrm{N}};\mathrm{X}_\mathrm{N})_{\Pi\mathrm{F}_{ij}}$ over the general spaces $\mathrm{X}_1,\ldots,\mathrm{X}_\mathrm{N}$ in the following form:

$$\int_{\mathrm{X}_1} d\boldsymbol{r}_1\,|\Omega_{k_1}|\int_{\mathrm{X}_2} d\boldsymbol{r}_2\,|\Omega_{k_2}|\mathrm{F}_{12}\ldots\int_{\mathrm{X}_\mathrm{N}} d\boldsymbol{r}_\mathrm{N}\,|\Omega_{k_\mathrm{N}}|\prod_{j<\mathrm{N}}\mathrm{F}_{j\mathrm{N}}. \tag{D9}$$

With the shorthand notation $(\boldsymbol{r})_\mathrm{i} = (\boldsymbol{r}_1,\ldots,\boldsymbol{r}_\mathrm{i})$, we define the function

$$\mathrm{U}_{k_m}^{\mathrm{X}_m}\big((\boldsymbol{r})_{m-1}\big) = \int_{\mathrm{X}_m} d\boldsymbol{r}_m\,|\Omega_{k_m}|\prod_{j<m}\mathrm{F}_{jm}(|\boldsymbol{r}_j-\boldsymbol{r}_m|). \tag{D10}$$

The restriction of this function to $\mathcal{X}_{m-1} = \mathrm{X}_1\times\cdots\times\mathrm{X}_{m-1}$ is bound from above by the general maximal potential integral

$$\mathrm{U}_{k_m}^{\mathrm{X}_m}\big|_{\mathcal{X}_{m-1}} \leq \max_{(\boldsymbol{r})_{m-1}\in\mathcal{X}_{m-1}}\mathrm{U}_{k_m}^{\mathrm{X}_m}\big((\boldsymbol{r})_{m-1}\big) = \tilde{\mathrm{U}}_{k_m}^{\mathrm{X}_m;\mathcal{X}_{m-1}}.$$

Inserting this bound successively into (D9) gives the inequality

$$(\overline{k_1};\mathrm{X}_1|\cdots|\overline{k_\mathrm{N}};\mathrm{X}_\mathrm{N})_{\Pi\mathrm{F}_{ij}} \leq \prod_{i=1}^{\mathrm{N}}\tilde{\mathrm{U}}_{k_i}^{\mathrm{X}_i;\mathcal{X}_{i-1}}, \tag{D11}$$

which is a product of terms that depend only on a single distribution directly, with parametric dependence on the integration spaces of the other distributions. We note that it is always possible to simplify $\tilde{\mathrm{U}}_{k_m}^{\mathrm{X}_m;\mathcal{X}_{m-1}}$ further by taking maxima over $\boldsymbol{r}_m$ for each F_{jm} to get

Published under license by AIP Publishing

$$\tilde{U}^{X_m;\mathcal{X}_{m-1}}_{k_m} \le \mathcal{S}^{X_m}_{k_m} \prod_{j<m} \max_{\substack{\boldsymbol{r}_m\in X_m \\ \boldsymbol{r}_j\in X_j}} F_{jm}(|\boldsymbol{r}_j - \boldsymbol{r}_m|) \tag{D12}$$

$$= \mathcal{S}^{X_m}_{k_m} \prod_{j<m} F_{jm}(d(X_j, X_m)) , \tag{D13}$$

where $\mathcal{S}^{X_m}_{k_m} = \int_{X_m} d\boldsymbol{r} |\Omega_{k_m}|$. This is a significant simplification because $d(X_j, X_m)$ is always straightforward to calculate for the sets used in practice. However, for unbounded F_{jm} and overlapping X_j and X_m, one has

$$F_{jm}(d(X_j, X_m)) = F_{jm}(0) = \infty. \tag{D14}$$

For each j where this is the case, we take the corresponding maximum over $\boldsymbol{r}_m$ only for the bounded part of the operator F^b_{jm} in order to obtain a practically useful bound.

If we restrict our formulation to the case that all F_{ij} are bounded-type, except for F_{12}, which may be bounded or Coulomb-type, then we can safely use (D13) for all terms except $\tilde{U}^{X_2;\mathcal{X}_1}_{k_2}$ where we use the following inequality instead:

$$\tilde{U}^{X_2;\mathcal{X}_1}_{k_2} \le \mathcal{U}^{X_1,X_2}_{k_2} \mathcal{D}^{X_1,X_2}_{12}. \tag{D15}$$

These factors are defined in Sec. II A 2. The result is the bound

$$(\overline{k_1}; X_1 | \cdots | \overline{k_N}; X_N)_{\Pi F_{ij}} \le \mathcal{S}^{X_1}_{k_1} \mathcal{U}^{X_1X_2}_{k_2} \mathcal{D}^{X_1,X_2}_{12} \prod_{j=3}^{N} \mathcal{S}^{X_j}_{k_j} \prod_{i<j} \mathcal{D}^{X_i,X_j}_{ij} \tag{D16}$$

$$= \mathcal{U}^{X_2X_1}_{k_1} \prod_{j=2}^{N} \mathcal{S}^{X_j}_{k_j} \prod_{i<j} \mathcal{D}^{X_i,X_j}_{ij}. \tag{D17}$$

The second form is obtained by exchanging electronic indices $1 \leftrightarrow 2$ and using the symmetry of $\mathcal{D}^{X_1,X_2}_{12}$. Applying (D17) to the integrals $I^{\circ k_n}_N$, $I^{\bullet\mathcal{I}}_N$ and those that result from the partitionings $\dot{E}_{k_i} = \dot{e}_{k_i} \sqcup S_{k_i}$ lead to the extent equations, the IPB1, and the IPB2 bounds, respectively.

REFERENCES

[1] J. Heyd, G. E. Scuseria, and M. Ernzerhof, J. Chem. Phys. **118**, 8207 (2003).
[2] A. V. Krukau, O. A. Vydrov, A. F. Izmaylov, and G. E. Scuseria, J. Chem. Phys. **125**, 224106 (2006).
[3] J.-W. Song, K. Yamashita, and K. Hirao, J. Chem. Phys. **135**, 071103 (2011).
[4] J.-W. Song and K. Hirao, J. Chem. Phys. **143**, 144112 (2015).
[5] W. Kutzelnigg and W. Klopper, J. Chem. Phys. **94**, 1985 (1991).
[6] K.-C. Pan and H. F. King, J. Chem. Phys. **53**, 4397 (1970).
[7] S. Ten-no, Chem. Phys. Lett. **398**, 56 (2004).
[8] T. Yanai and T. Shiozaki, J. Chem. Phys. **136**, 084107 (2012).
[9] B. J. Persson and P. R. Taylor, J. Chem. Phys. **105**, 5915 (1996).
[10] A. J. May and F. R. Manby, J. Chem. Phys. **121**, 4479 (2004).
[11] E. F. Valeev, J. Chem. Phys. **125**, 244106 (2006).
[12] J. L. Whitten, J. Chem. Phys. **58**, 4496 (1973).
[13] M. Häser and R. Ahlrichs, J. Comput. Chem. **10**, 104 (1989).
[14] T. H. Thompson and C. Ochsenfeld, J. Chem. Phys. **147**, 144101 (2017).
[15] S. A. Maurer, D. S. Lambrecht, J. Kussmann, and C. Ochsenfeld, J. Chem. Phys. **138**, 014101 (2013).
[16] S. A. Maurer, D. S. Lambrecht, D. Flaig, and C. Ochsenfeld, J. Chem. Phys. **136**, 144107 (2012).
[17] D. S. Hollman, H. F. Schaefer, and E. F. Valeev, J. Chem. Phys. **142**, 154106 (2015).
[18] A. F. Izmaylov, G. E. Scuseria, and M. J. Frisch, J. Chem. Phys. **125**, 104103 (2006).
[19] M. Beuerle, J. Kussmann, and C. Ochsenfeld, J. Chem. Phys. **146**, 144108 (2017).
[20] T. B. Adler, H.-J. Werner, and F. R. Manby, J. Chem. Phys. **130**, 054106 (2009).
[21] G. M. J. Barca and P. M. W. Gill, J. Chem. Theory Comput. **12**, 4915 (2016).
[22] G. M. J. Barca and P.-F. Loos, J. Chem. Phys. **147**, 024103 (2017).
[23] G. M. J. Barca, "Single-determinant theory of electronic excited states and many-electron integrals for explicitly correlated," Ph.D. thesis, Research School of Chemistry, The Australian National University, 2017.
[24] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[25] M. J. Field, P. A. Bash, and M. Karplus, J. Comput. Chem. **11**, 700 (1990).
[26] H. M. Senn and W. Thiel, Angew. Chem., Int. Ed. **48**, 1198 (2009).
[27] R. A. Friesner, Chem. Phys. Lett. **116**, 39 (1985).
[28] F. Neese, F. Wennmohs, A. Hansen, and U. Becker, Chem. Phys. **356**, 98 (2009).
[29] H. Laqua, J. Kussmann, and C. Ochsenfeld, J. Chem. Theory Comput. **14**, 3451 (2018).
[30] B. J. Persson and P. R. Taylor, Theor. Chem. Acc. **97**, 240 (1997).
[31] A. Komornicki and H. F. King, J. Chem. Phys. **134**, 244115 (2011).
[32] C. Ochsenfeld, C. A. White, and M. Head-Gordon, J. Chem. Phys. **109**, 1663 (1998).
[33] C. Ochsenfeld, Chem. Phys. Lett. **327**, 216 (2000).
[34] J. Kussmann and C. Ochsenfeld, J. Chem. Phys. **138**, 134114 (2013).
[35] J. Kussmann and C. Ochsenfeld, J. Chem. Theory Comput. **11**, 918 (2015).
[36] J. Kussmann and C. Ochsenfeld, J. Chem. Theory Comput. **13**, 3153 (2017).
[37] C. Hättig, W. Klopper, A. Köhn, and D. P. Tew, Chem. Rev. **112**, 4 (2011).
[38] L. Kong, F. A. Bischoff, and E. F. Valeev, Chem. Rev. **112**, 75 (2011).
[39] S. Ten-no, J. Chem. Phys. **126**, 014108 (2007).
[40] S. Ten-no, J. Chem. Phys. **121**, 117 (2004).
[41] H.-J. Werner, T. B. Adler, and F. R. Manby, J. Chem. Phys. **126**, 164102 (2007).
[42] E. F. Valeev, Chem. Phys. Lett. **395**, 190 (2004).
[43] D. S. Hollman, J. J. Wilke, and H. F. Schaefer, J. Chem. Phys. **138**, 064107 (2013).
[44] G. M. J. Barca, P.-F. Loos, and P. M. W. Gill, J. Chem. Theory Comput. **12**, 1735 (2016).
[45] G. Lohöfer, J. Approx. Theory **95**, 178 (1998).
[46] T. Helgaker, J. Olsen, and P. Jørgensen, *Molecular Electronic-Structure Theory* (Wiley-Blackwell, 2013).
[47] I. Newton, T. Leseur, and F. Jacquier, *Philosophiae Naturalis Principia Mathematica*, *Philosophiae Naturalis Principia Mathematica* (G. Brookman, 1833), Vols. 1 and 2.
[48] E. H. Lieb and R. Seiringer, *The Stability of Matter in Quantum Mechanics* (Cambridge University Press, 2009).
[49] Boost C++ libraries, Version 1.60.0, https://www.boost.org.

Published under license by AIP Publishing

SUPPLEMENTARY MATERIAL

Integral partition bounds for fast and effective screening of general one-, two-, and many-electron integrals

Travis H. Thompson and Christian Ochsenfeld

Chair of Theoretical Chemistry, Department of Chemistry,
University of Munich (LMU),
Butenandtstr. 7, D-81377 Munich, Germany
and
Center for Integrated Protein Science Munich (CIPSM) at the Department of Chemistry,
University of Munich (LMU), Butenandtstr. 5-13,
D-81377 Munich, Germany

I. IPB2 TERMS

The IPB2 bound is given by

$$|\,(k_1|\cdots|k_N)_{\Pi F_{ij}}\,| \leq (1+N_p)\sum_{i=0}^{N} T^i(\alpha)\ , \tag{1}$$

We define

$$D_{ij}^{ee} = \mathcal{D}_{ij}^{\dot{e}_{k_i}\dot{e}_{k_j}}, D_{ij}^{Ee} = \mathcal{D}_{ij}^{\dot{E}_{k_i}\dot{e}_{k_j}}, D_{ij}^{eE} = \mathcal{D}_{ij}^{\dot{e}_{k_i}\dot{E}_{k_j}}, D_{ij}^{EE} = \mathcal{D}_{ij}^{\dot{E}_{k_i}\dot{E}_{k_j}}$$
$$U_{k_1}^{ee} = \mathcal{U}_{k_1}^{\dot{e}_{k_1}\dot{e}_{k_2}}, U_{k_1}^{eS} = \mathcal{U}_{k_1}^{\dot{e}_{k_2}S_{k_1}}, U_{k_1}^{Ee} = \mathcal{U}_{k_1}^{\dot{E}_{k_2}\dot{e}_{k_1}}, U_{k_1}^{ES} = \mathcal{U}_{k_1}^{\dot{E}_{k_2}S_{k_1}},$$

and note that as $\alpha \to 0$, $U_{k_1}^{eS} = \mathcal{O}(\alpha)$ and $U_{k_1}^{ES} = \mathcal{O}(\alpha)$, while $U_{k_1}^{ee}$ and $U_{k_1}^{Ee}$ are independent of α. When F_{12} is bounded type the factors simplify to $U_{k_1}^{eS} = U_{k_1}^{ES} = \alpha\mathcal{S}_{k_1}$ and $U_{k_1}^{ee} = U_{k_1}^{Ee} = \mathcal{S}_{k_1}$.

For $N = 2$ one calculates the factors,

$$U_{k1}^{ee}, U_{k1}^{Ee}, U_{k1}^{eS}, U_{k1}^{ES} \quad D_{12}^{ee}, D_{12}^{Ee}, D_{12}^{eE}, D_{12}^{EE}\ ,$$

The T^i terms are:

$$\begin{aligned}
T^0 =& \mathcal{S}_{k_2} U_{k_1}^{ee} D_{12}^{ee} \\
T^1 =& \mathcal{S}_{k_2}\left(U_{k_1}^{eS} D_{12}^{Ee} + \alpha\, U_{k_1}^{Ee} D_{12}^{eE}\right) \\
T^3 =& \mathcal{S}_{k_2}\alpha\, U_{k_1}^{ES} D_{12}^{EE}
\end{aligned}$$

For $N = 3$ one calculates the factors,

$$\begin{aligned}
&U_{k1}^{ee}, U_{k1}^{Ee}, U_{k1}^{eS}, U_{k1}^{ES} \quad D_{12}^{ee}, D_{12}^{Ee}, D_{12}^{eE}, D_{12}^{EE} \quad D_{13}^{ee}, D_{13}^{Ee}, D_{13}^{eE}, D_{13}^{EE} \\
&D_{23}^{ee}, D_{23}^{Ee}, D_{23}^{eE}, D_{23}^{EE}\ ,
\end{aligned}$$

as early as possible in the loop structure. For the integral types 3A and 3B of the main paper, $F_{13} = 1$, and the D_{13} factors are not needed. The T^i terms in the general case are:

$$\begin{aligned}
T^0 =& \mathcal{S}_{k_2}\mathcal{S}_{k_3} U_{k_1}^{ee} D_{12}^{ee} D_{13}^{ee} D_{23}^{ee} \\
T^1 =& \mathcal{S}_{k_2}\mathcal{S}_{k_3}\left(U_{k_1}^{eS} D_{12}^{Ee} D_{13}^{Ee} D_{23}^{ee} + \alpha\, U_{k_1}^{Ee} D_{12}^{eE} D_{13}^{ee} D_{23}^{Ee} + \alpha\, U_{k_1}^{ee} D_{12}^{ee} D_{13}^{eE} D_{23}^{eE}\right) \\
T^2 =& \mathcal{S}_{k_2}\mathcal{S}_{k_3}\left(\alpha\, U_{k_1}^{ES} D_{12}^{EE} D_{13}^{Ee} D_{23}^{Ee} + \alpha\, U_{k_1}^{eS} D_{12}^{Ee} D_{13}^{EE} D_{23}^{eE} + \alpha^2\, U_{k_1}^{Ee} D_{12}^{eE} D_{13}^{eE} D_{23}^{EE}\right) \\
T^3 =& \mathcal{S}_{k_2}\mathcal{S}_{k_3}\alpha^2\, U_{k_1}^{ES} D_{12}^{EE} D_{13}^{EE} D_{23}^{EE}
\end{aligned}$$

For $N = 4$ one needs

$$\begin{array}{lll} U^{ee}_{k1}, U^{Ee}_{k1}, U^{eS}_{k1}, U^{ES}_{k1} & D^{ee}_{12}, D^{Ee}_{12}, D^{eE}_{12}, D^{EE}_{12} & D^{ee}_{13}, D^{Ee}_{13}, D^{eE}_{13}, D^{EE}_{13} \\ D^{ee}_{23}, D^{Ee}_{23}, D^{eE}_{23}, D^{EE}_{23} & D^{ee}_{14}, D^{Ee}_{14}, D^{eE}_{14}, D^{EE}_{14} & D^{ee}_{24}, D^{Ee}_{24}, D^{eE}_{24}, D^{EE}_{24} \\ D^{ee}_{34}, D^{Ee}_{34}, D^{eE}_{34}, D^{EE}_{34} \ , & & \end{array}$$

where in the F12 integral types 4A, 4B, and 4C, only three of the six operators are present and only the D factors corresponding to these are needed.

The T^i terms are in the general case:

$$\begin{aligned} T^0 &= \mathcal{S}_{k_2}\mathcal{S}_{k_3}\mathcal{S}_{k_4} U^{ee}_{k_1} D^{ee}_{12} D^{ee}_{13} D^{ee}_{23} D^{ee}_{14} D^{ee}_{24} D^{ee}_{34} \\ T^1 &= \mathcal{S}_{k_2}\mathcal{S}_{k_3}\mathcal{S}_{k_4} \big(U^{eS}_{k_1} D^{Ee}_{12} D^{Ee}_{13} D^{ee}_{23} D^{Ee}_{14} D^{ee}_{24} D^{ee}_{34} + \alpha\, U^{Ee}_{k_1} D^{eE}_{12} D^{ee}_{13} D^{Ee}_{23} D^{ee}_{14} D^{Ee}_{24} D^{ee}_{34} \\ &\qquad + \alpha\, U^{ee}_{k_1} D^{ee}_{12} D^{eE}_{13} D^{eE}_{23} D^{ee}_{14} D^{ee}_{24} D^{Ee}_{34} + \alpha\, U^{ee}_{k_1} D^{ee}_{12} D^{ee}_{13} D^{ee}_{23} D^{eE}_{14} D^{eE}_{24} D^{eE}_{34} \big) \\ T^2 &= \mathcal{S}_{k_2}\mathcal{S}_{k_3}\mathcal{S}_{k_4} \big(\alpha\, U^{ES}_{k_1} D^{EE}_{12} D^{Ee}_{13} D^{Ee}_{23} D^{Ee}_{14} D^{Ee}_{24} D^{ee}_{34} + \alpha\, U^{eS}_{k_1} D^{Ee}_{12} D^{EE}_{13} D^{eE}_{23} D^{Ee}_{14} D^{ee}_{24} D^{Ee}_{34} \\ &\qquad + \alpha^2\, U^{ee}_{k_1} D^{eE}_{12} D^{eE}_{13} D^{EE}_{23} D^{ee}_{14} D^{Ee}_{24} D^{Ee}_{34} + \alpha\, U^{eS}_{k_1} D^{Ee}_{12} D^{Ee}_{13} D^{ee}_{23} D^{EE}_{14} D^{eE}_{24} D^{eE}_{34} \\ &\qquad + \alpha^2\, U^{ee}_{k_1} D^{eE}_{12} D^{ee}_{13} D^{Ee}_{23} D^{eE}_{14} D^{EE}_{24} D^{eE}_{34} + \alpha^2\, U^{ee}_{k_1} D^{ee}_{12} D^{eE}_{13} D^{eE}_{23} D^{eE}_{14} D^{eE}_{24} D^{EE}_{34} \big) \\ T^3 &= \mathcal{S}_{k_2}\mathcal{S}_{k_3}\mathcal{S}_{k_4} \big(\alpha^2\, U^{ES}_{k_1} D^{EE}_{12} D^{EE}_{13} D^{EE}_{23} D^{Ee}_{14} D^{Ee}_{24} D^{Ee}_{34} + \alpha^2\, U^{ES}_{k_1} D^{EE}_{12} D^{Ee}_{13} D^{Ee}_{23} D^{EE}_{14} D^{EE}_{24} D^{eE}_{34} \\ &\qquad + \alpha^3\, U^{Ee}_{k_1} D^{eE}_{12} D^{eE}_{13} D^{EE}_{23} D^{eE}_{14} D^{EE}_{24} D^{EE}_{34} + \alpha^2\, U^{eS}_{k_1} D^{Ee}_{12} D^{EE}_{13} D^{eE}_{23} D^{EE}_{14} D^{eE}_{24} D^{EE}_{34} \big) \\ T^4 &= \mathcal{S}_{k_2}\mathcal{S}_{k_3}\mathcal{S}_{k_4} \alpha^3\, U^{ES}_{k_1} D^{EE}_{12} D^{EE}_{13} D^{EE}_{23} D^{EE}_{14} D^{EE}_{24} D^{EE}_{34} \ . \end{aligned}$$

In practice, the terms are simpler due to the fact that three of the six D factors are equal to unity in each term.

II. SCREENING PROCEDURE FOR A FOUR-ELECTRON INTEGRAL

Here we present the screening algorithm for the four-electron integral type 4A from the main paper. The energies that correspond to these terms are

$$E^{4A}_D = \sum_{ijkl} \langle ijkl|C_{12}F_{23}F_{14}|jikl\rangle = \sum_{\substack{\mu\nu\ \lambda\sigma \\ \rho\eta\ \delta\zeta}} P^{12}_{\mu\sigma} P^{12}_{\lambda\nu} P^{34}_{\rho\zeta} P^{34}_{\delta\eta} \langle \mu\lambda\rho\delta|C_{12}F_{23}F_{14}|\nu\sigma\eta\zeta\rangle \tag{2}$$

$$E^{4A}_X = \sum_{ijkl} \langle ijkl|C_{12}F_{23}F_{14}|kilj\rangle = \sum_{\substack{\mu\nu\ \lambda\sigma \\ \rho\eta\ \delta\zeta}} E^{12}_{\mu\sigma} P^{13}_{\rho\nu} P^{24}_{\lambda\zeta} P^{34}_{\delta\eta} \langle \mu\lambda\rho\delta|C_{12}F_{23}F_{14}|\nu\sigma\eta\zeta\rangle. \tag{3}$$

Here, C_{12} is the Coulomb potential and F_{23} and F_{14} are bounded-type correlation factors. The procedure for screening the integrals using IPB0 and IPB1 is given in Algorithm 1 below.

Algorithm 1 E_D^{4A} and E_X^{4A} screening procedure

for all $\tilde{\lambda} \in \mathbb{S}$ **do**
  for all $\tilde{\sigma} \in \mathbb{O}(\tilde{\lambda})$ **do**
    for all $\tilde{\rho} \in \mathbb{F}_{23}(\tilde{\sigma})$ **do**
      if $3\mathcal{S}_{\tilde{\lambda}\tilde{\sigma}}\mathcal{D}_{23}^{\dot{\mathcal{E}}_{\tilde{\lambda}}\dot{\mathcal{E}}_{\tilde{\rho}}} F_{14}(0)\mathcal{S}_{\max}^2\mathcal{V}_{\max}P_{max}^4 < \vartheta$ **then**
        Continue
      end if
      for all $\tilde{\eta} \in \mathbb{O}(\tilde{\rho})$ **do**
        Calc $D_{23} = \mathcal{D}_{23}^{\dot{E}_{\tilde{\lambda}\tilde{\sigma}}\dot{E}_{\tilde{\rho}\tilde{\eta}}}$
        if $3\mathcal{S}_{\tilde{\lambda}\tilde{\sigma}}\mathcal{S}_{\tilde{\rho}\tilde{\eta}}D_{23}F_{14}(0)\mathcal{S}_{\max}\mathcal{V}_{\max}P_{\max}^4 < \vartheta$ **then**
          Continue
        end if
        for all $\tilde{\mu} \in \mathbb{P}(\tilde{\sigma})$ **do**
          for all $\tilde{\nu} \in \mathbb{O}(\tilde{\mu})$ **do**
            Calc $D_{12} = \mathcal{D}_{12}^{\dot{E}_{\tilde{\mu}\tilde{\nu}}\dot{E}_{\tilde{\lambda}\tilde{\sigma}}}$
            Calc $U_{\tilde{\mu}\tilde{\nu}} = \mathcal{U}_{\tilde{\mu}\tilde{\nu}}^{\dot{E}_{\tilde{\lambda}\tilde{\sigma}}\dot{E}_{\tilde{\mu}\tilde{\nu}}}$
            $P_{\text{int}} = \max\{\hat{\mathcal{P}}_{\tilde{\mu}\tilde{\nu}}P_{\tilde{\mu}\tilde{\sigma}}P_{\tilde{\nu}\tilde{\lambda}}, \hat{\mathcal{P}}_{\tilde{\mu}\tilde{\nu}}P_{\tilde{\mu}\tilde{\sigma}}P_{\tilde{\rho}\tilde{\nu}}, \hat{\mathcal{P}}_{\tilde{\mu}\tilde{\nu}}P_{\tilde{\mu}\tilde{\lambda}}P_{\tilde{\rho}\tilde{\nu}}, \hat{\mathcal{P}}_{\tilde{\mu}\tilde{\nu}}P_{\tilde{\mu}\tilde{\sigma}}P_{\tilde{\eta}\tilde{\nu}}, \hat{\mathcal{P}}_{\tilde{\mu}\tilde{\nu}}P_{\tilde{\mu}\tilde{\lambda}}P_{\tilde{\eta}\tilde{\nu}}\}$
            if $4U_{\tilde{\mu}\tilde{\nu}}\mathcal{S}_{\tilde{\lambda}\tilde{\sigma}}\mathcal{S}_{\tilde{\rho}\tilde{\eta}}D_{12}D_{23}F_{14}(0)P_{\text{int}}\mathcal{S}_{\max}P_{\max}^2 < \vartheta$ **then**
              Continue
            end if
            for all $\tilde{\delta} \in \mathbb{F}_{14}(\tilde{\mu})$ **do**
              if $5\,U_{\tilde{\mu}\tilde{\nu}}\mathcal{S}_{\tilde{\lambda}\tilde{\sigma}}\mathcal{S}_{\tilde{\rho}\tilde{\eta}}D_{12}D_{23}\mathcal{D}_{14}^{\dot{\mathcal{E}}_{\tilde{\lambda}}\dot{\mathcal{E}}_{\tilde{\rho}}}P_{\text{int}}\mathcal{S}_{\max}P_{max}^2 < \vartheta$ **then** Continue
              end if
              for all $\tilde{\zeta} \in \mathbb{O}(\tilde{\delta})$ **do**
                Calc $D_{14} = \mathcal{D}_{14}^{\dot{E}_{\tilde{\mu}\tilde{\nu}}\dot{E}_{\tilde{\delta}\tilde{\zeta}}}$
                $P_X = \hat{\mathcal{P}}_{\tilde{\mu}\tilde{\nu}}\hat{\mathcal{P}}_{\tilde{\lambda}\tilde{\sigma}}\hat{\mathcal{P}}_{\tilde{\rho}\tilde{\eta}}\hat{\mathcal{P}}_{\tilde{\rho}\tilde{\eta}}(P_{\tilde{\mu}\tilde{\sigma}}P_{\tilde{\lambda}\tilde{\nu}}P_{\tilde{\rho}\tilde{\zeta}}P_{\tilde{\delta}\tilde{\eta}})$
                $P_D = \hat{\mathcal{P}}_{\tilde{\mu}\tilde{\nu}}\hat{\mathcal{P}}_{\tilde{\lambda}\tilde{\sigma}}\hat{\mathcal{P}}_{\tilde{\rho}\tilde{\eta}}\hat{\mathcal{P}}_{\tilde{\rho}\tilde{\eta}}(P_{\tilde{\mu}\tilde{\sigma}}P_{\tilde{\rho}\tilde{\nu}}P_{\tilde{\lambda}\tilde{\zeta}}P_{\tilde{\delta}\tilde{\eta}})$
                $P_{\text{fin}} = \max\{P_X, P_D\}$
                if $5\,U_{\tilde{\mu}\tilde{\nu}}\mathcal{S}_{\tilde{\lambda}\tilde{\sigma}}\mathcal{S}_{\tilde{\rho}\tilde{\eta}}\mathcal{S}_{\tilde{\delta}\tilde{\zeta}}D_{12}D_{23}D_{14}P_{\text{fin}} \geq \vartheta$ **then**
                  Calculate integral sub-tensor $\langle\tilde{\mu}\tilde{\lambda}\tilde{\rho}\tilde{\delta}|C_{12}F_{23}F_{14}|\tilde{\nu}\tilde{\sigma}\tilde{\eta}\tilde{\zeta}\rangle$
                  Contract with $\boldsymbol{P}$ elements
                end if
              end for
            end for
          end for
        end for
      end for
    end for
  end for
end for

III. COST OF PRE-FACTOR CALCULATION

In Table I we compare the timings for IPB pre-factor and Schwarz inequality (QQ) pre-factor calculation. The timings were performed on a single node containing two Intel E5-2667 CPU's (2*8 cores) for strands of adenine-thymine DNA base pairs of increasing length with the cc-pVDZ basis set. The times are also given as percentages of the time needed for one exchange matrix build using the converged density matrix of the respective system. Here we see that IPB pre-factor calculation time, which is dominated by the solution of the extent equations, is only slightly more costly than that of the Schwarz factors in absolute terms. Pre-factor calculation time is negligible in both cases and the slight increase for the IPB's is compensated for many times over by the savings seen in integral calculation, which are detailed in the main paper.

TABLE I: Pre-factor calculation times in seconds for strands of DNA base pairs with cc-pVDZ basis set. The times are also given as percentages of the time needed for one exchange matrix build (%K) using a QQ screening and $\vartheta = 10^{-9}$.

Prefactor Calculation Times					
		QQ		IPB2	
Base pairs	No. of atoms	Absolute Time (s)	%K	Absolute Time (s)	%K
1	62	0.09	0.4%	0.3	1.8%
2	128	0.3	0.2%	0.9	0.8%
4	260	0.6	0.1%	2.1	0.4%
8	524	1.3	0.07%	4.6	0.2%
16	1052	2.6	0.07%	9.6	0.2%

IV. SPHERICAL INTEGRAL FACTORS $\tilde{\mathcal{M}}_{l,j}$

The $\tilde{\mathcal{M}}_{l,j}$ values given in Table II were calculated using the computational software Wolfram Mathematica [1].

TABLE II: Numerically calculated values of $\tilde{\mathcal{M}}_{l,j}$

$l\backslash j$	0	1	2	3	4	5	6	7	8
0	12.56637								
1	6.283185	4.188790							
2	4.836774	2.720699	2.513274						
3	4.084063	2.336123	1.756204	1.795196					
4	3.601415	1.985678	1.576410	1.298730	1.396263				
5	3.257724	1.831955	1.433464	1.184905	1.030835	1.142397			
6	2.997066	1.659836	1.287233	1.108474	0.948599	0.854958	0.9666439		
7	2.790302	1.561038	1.177159	1.033207	0.890517	0.797688	0.7313304	0.8377580	
8	2.621252	1.454301	1.127299	0.949797	0.854748	0.752337	0.6883817	0.6389458	0.7391983

[1] Wolfram Research, Inc., Mathematica, Version 10.2, Champaign, IL (2015).

4.3 Publication III

A Schwarz inequality for complex basis function methods in non-Hermitian quantum chemistry

T. H. Thompson, C. Ochsenfeld, and T.-C. Jagau,
J. Chem. Phys. **151**, 184104 (2019).

Abstract:

A generalization of the Schwarz bound employed to reduce the scaling of quantum-chemical calculations is introduced in the context of non-Hermitian methods employing complex-scaled basis functions. Non-Hermitian methods offer a treatment of molecular metastable states in terms of L^2-integrable wave functions with complex energies, but until now, an efficient upper bound for the resulting electron-repulsion integrals has been unavailable due to the complications from non-Hermiticity. Our newly formulated bound allows us to inexpensively and rigorously estimate the sparsity in the complex-scaled two-electron integral tensor, providing the basis for efficient integral screening procedures. We have incorporated a screening algorithm based on the new Schwarz bound into the state-of-the-art complex basis function integral code by White, Head-Gordon, and McCurdy [J. Chem. Phys. **142**, 054103 (2015)]. The effectiveness of the screening is demonstrated through non-Hermitian Hartree-Fock calculations of the static field ionization of the 2-pyridoxine 2-aminopyridine molecular complex.

The following article is reproduced in agreement with its publisher (AIP Publishing LLC) and can be found online at: https://doi.org/10.1063/1.5123541

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

A Schwarz inequality for complex basis function methods in non-Hermitian quantum chemistry

Cite as: J. Chem. Phys. 151, 184104 (2019); doi: 10.1063/1.5123541
Submitted: 9 August 2019 • Accepted: 14 October 2019 •
Published Online: 11 November 2019

View Online Export Citation CrossMark

Travis H. Thompson,[1,2] Christian Ochsenfeld,[1,2] and Thomas-C. Jagau[1,a)]

AFFILIATIONS

[1] Department of Chemistry, University of Munich (LMU), Butenandtstr. 5-13, D-81377 Munich, Germany
[2] Center for Integrated Protein Science Munich (CIPSM) at the Department of Chemistry, University of Munich (LMU), Butenandtstr. 5-13, D-81377 Munich, Germany

[a)] Electronic mail: th.jagau@lmu.de

ABSTRACT

A generalization of the Schwarz bound employed to reduce the scaling of quantum-chemical calculations is introduced in the context of non-Hermitian methods employing complex-scaled basis functions. Non-Hermitian methods offer a treatment of molecular metastable states in terms of L^2-integrable wave functions with complex energies, but until now, an efficient upper bound for the resulting electron-repulsion integrals has been unavailable due to the complications from non-Hermiticity. Our newly formulated bound allows us to inexpensively and rigorously estimate the sparsity in the complex-scaled two-electron integral tensor, providing the basis for efficient integral screening procedures. We have incorporated a screening algorithm based on the new Schwarz bound into the state-of-the-art complex basis function integral code by White, Head-Gordon, and McCurdy [J. Chem. Phys. **142**, 054103 (2015)]. The effectiveness of the screening is demonstrated through non-Hermitian Hartree-Fock calculations of the static field ionization of the 2-pyridoxine 2-aminopyridine molecular complex.

Published under license by AIP Publishing. https://doi.org/10.1063/1.5123541

I. INTRODUCTION

Electronic resonances are metastable states embedded in the continuum. Such states are involved in a variety of chemical processes where unbound electrons occur.[1,2] For example, resonances are important in the chemistry of molecules exposed to electric and electromagnetic fields with strengths comparable with the intermolecular forces. These occur, for example, in molecular high-harmonic generation,[3] laser-induced electron diffraction,[4] and Coulomb explosion.[5]

An elegant description of such states can be achieved using complex-coordinate methods[1,6] in which the coordinates of the Hamiltonian are rotated into the complex-plane through an unbounded similarity transformation. The resulting Hamiltonian is no longer Hermitian, its eigenfunctions are square integrable and provide a mathematically rigorous description of the resonance states. The corresponding eigenvalues are the so-called Siegert energies,[2,6,7] whose real and imaginary parts give the positions $\mathcal{E}$ and widths Γ of resonance states, respectively,

$$E = \mathcal{E} - i\Gamma/2. \tag{1}$$

The use of this unbounded similarity transformation is generally known as complex scaling (CS), and its main advantage is that it allows one to apply the efficient quantum-chemical methods developed for bound states to resonances. The mathematical basis for CS is provided by the Balslev-Combes theorem[8–10] and its subsequent extensions for resonances in external fields.[11–13] A problem that arises, however, is that the original CS formulation is not compatible with the Born-Oppenheimer approximation,[14,15] which initially limited applications to atoms.[16] A mathematically rigorous unification of CS with the Born-Oppenheimer picture is given by exterior complex scaling,[14] where only the coordinates outside a given radius are complex scaled, but its direct application is not practical in the context of electronic-structure theory using Gaussian basis functions.

One suggestion to overcome the practical difficulties involved was proposed by McCurdy and Rescigno,[15] who noted that the important asymptotic effects of CS could be captured by complex-scaling the exponents of the most diffuse basis functions. The Gaussian basis functions used in such calculations take the form

$$\chi_\mu(\boldsymbol{r}, \boldsymbol{A}) = N_\mu S_\mu(\boldsymbol{r}_A) \exp\left[-\alpha_\mu e^{-2i\theta_\mu} \boldsymbol{r}_A^2\right], \tag{2}$$

Published under license by AIP Publishing

where N_μ is a normalization constant, $\boldsymbol{A}$ is an atomic center, $\boldsymbol{r_A} = \boldsymbol{r} - \boldsymbol{A}$, and S_μ is a real polynomial in the components of $\boldsymbol{r_A}$ that depends on the angular quantum numbers of χ_μ. Each basis function is either strictly real, $\theta_\mu = 0$, or scaled using a global complex scaling factor $\theta_\mu = \theta \in \mathbb{R}$. We note that the use of complex θ has been explored in the context of the closely related approach of analytic continuation of the Hamiltonian matrix elements.[17]

The method of complex basis functions had been restricted to small systems[18–24] due to the need for calculating nonstandard two-electron integrals over complex basis functions. Recently, however, renewed interest has led to efficient implementations of complex basis function methods in the non-Hermitian Hartree-Fock (HF) and static exchange (SE) approximations.[25,26] This allowed the application of the method to resonances of medium-size polyatomic molecules for the first time. This implementation was then extended to include electron correlation at the second-order Møller-Plesset (MP2) and coupled-cluster singles and doubles (CCSD) levels of theory,[27] allowing for accurate descriptions of the resonances of small molecules.

In Ref. 26, it has been identified that one barrier to reaching the efficiency of bound-state calculations is the lack of a simple upper bound on the value of the electron-repulsion integrals (ERIs) over complex basis functions. In bound-state calculations, a key step for efficiency was the introduction of the Schwarz bound for screening ERIs over real basis functions,[28,29] which in Mulliken notation takes the form

$$|(\mu\nu|\lambda\sigma)| \leq (\mu\nu|\mu\nu)^{1/2}(\lambda\sigma|\lambda\sigma)^{1/2}. \tag{3}$$

This already reduces the asymptotic scaling of the integral work involved from quartic to quadratic and, in combination with density matrix screening methods for the exchange part, can even lead to linear scaling.[30–32]

The application of the Schwarz bound to complex-valued electron-repulsion integrals is straightforward as long as they conform to a true inner product. Such integrals arise, for example, when describing finite magnetic fields by means of London orbitals[33,34] or when working with a mixed basis set of Gaussian functions and plane waves.[35–38] The use of the latter type of basis is appropriate, for example, for describing electron-molecule scattering.[36,39]

However, in complex basis function methods for molecular electronic resonances that are the subject of this article, the bound (3) cannot be used directly because the ERIs are not positive semidefinite. Here, we show that a simple and effective Schwarz-type bound is nevertheless available and can be used analogously to (3), making the efficient screening methods of Hermitian quantum chemistry available to the non-Hermitian, complex basis function case, and eliminating this barrier to large-scale calculations.

In this work, we present results pertaining to the application of the new Schwarz bound to Stark resonances, which are induced by a static electric field. The application of the method of complex basis functions to atomic and molecular Stark resonances was first detailed in Refs. 40 and 41 for coupled-cluster wave functions. Very recently, an implementation at the resolution-of-the-identity MP2 level of theory has been reported.[42] Here, we focus on the complex basis function HF part and use the Schwarz bound to screen the complex ERIs required.

II. THEORY

A. Non-Hermitian SCF with complex basis functions

The working equations for non-Hermitian self-consistent field (SCF) methods employing complex basis functions are of essentially the same form as their bound-state counterparts. Details of exact formulations can be found in Refs. 19, 20, 25, and 26. Here, we focus on how the features of the theory affect the two-electron integrals involved.

The goal of non-Hermitian SCF methods is to find a stationary solution of the complex energy functional,[43,44]

$$E[\tilde{\Psi}] = \frac{\langle\tilde{\Psi}|\hat{H}|\tilde{\Psi}\rangle_c}{\langle\tilde{\Psi}|\tilde{\Psi}\rangle_c}, \tag{4}$$

where $\langle\cdot|\cdot\rangle_c$ denotes the c-product,[45] in which the bra is not complex-conjugated, $\hat{H}$ is the many-body Hamiltonian, and $\tilde{\Psi}$ is a complex trial wave function for which $\langle\tilde{\Psi}|\tilde{\Psi}\rangle_c < \infty$. $\tilde{\Psi}$ is approximated as a Slater-determinant of complex molecular orbitals which are expanded in a basis of atomic orbitals of the form (2). Thus, analogous to the case of a real trial wave function, non-Hermitian SCF is an iterative orbital optimization procedure, with an added dependency on the global complex scaling angle $0 < \theta < \pi/4$, which is contained completely in the one-electron basis. For an exact solution, the corresponding Siegert energy is independent of θ above the critical value θ_c required for square integrability.[6] However, this is not the case in an incomplete basis set, and it is necessary to determine the optimal value θ_{opt} for a particular system and basis set, which is given as a stationary point along a calculated θ-trajectory.[43,46] The use of the c-product leads to complex ERIs of the form

$$\begin{aligned}(\mu\nu|\lambda\sigma) &= \int d\boldsymbol{r}_1\chi_\mu(\boldsymbol{r}_1)\chi_\nu(\boldsymbol{r}_1)\int d\boldsymbol{r}_2\,\chi_\lambda(\boldsymbol{r}_2)\chi_\sigma(\boldsymbol{r}_2)\,r_{12}^{-1}\\ &:= \mathfrak{C}(\mu\nu,\lambda\sigma) \in \mathbb{C},\end{aligned} \tag{5}$$

where $r_{12} = |\boldsymbol{r}_1 - \boldsymbol{r}_2|$. The number of such integrals scales quartically with the size of the basis set and their calculation forms a large portion of the work to be done. The algorithms for their evaluation are similar to the case of real basis functions and require the evaluation of the complex version of the Boys function.[26] As we show in Sec. II C, the magnitudes of the complex ERIs are increasing on the interval $\theta \in [0, \pi/4]$ so that the value of θ has an impact on the number of integrals which can be neglected.

B. Schwarz bound for complex ERIs

The mapping $\mathfrak{C}$ defined in Eq. (5) assigns a complex energy to two complex one-electron functions. We denote the map describing the real ERIs of bound-state calculations as $\mathfrak{R}$, and it is just the restriction of $\mathfrak{C}$ to real functions. Thus, for any real one-electron functions, f and g, we have

$$\mathfrak{C}(f,g) = \mathfrak{R}(f,g) = \int d\boldsymbol{r}_1 f(\boldsymbol{r}_1)\int d\boldsymbol{r}_2\,g(\boldsymbol{r}_2)\,r_{12}^{-1} \in \mathbb{R}. \tag{6}$$

It is clear that $\mathfrak{R}$ is a symmetric bilinear form, and one can also show that it is positive semidefinite,[28,47,48] i.e., $\mathfrak{R}(f,f) \geq 0$ for any real f. Thus, all conditions are met for employing the Cauchy-Schwarz inequality,

$$|\mathfrak{R}(f,g)|^2 \leq \mathfrak{R}(f,f)\mathfrak{R}(g,g). \tag{7}$$

Published under license by AIP Publishing

This is just a reformulation of bound (3), which is of great importance in quantum chemistry, as it provides an inexpensive and simple test for determining *a priori* if an ERI, or class of ERIs, can be neglected. Integral screening techniques based on inequality (7) take advantage of the inherent sparsity in molecular calculations and lead to great computational savings.[29,49]

In contrast, the map $\mathfrak{C}$ is not a positive semidefinite Hermitian form and thus does not fulfill the complex version of inequality (7). For this reason, the efficient screening techniques developed on the basis of the Schwarz bound have, until now, evaded use in complex basis function methods.[26] However, we now show that one can relate $\mathfrak{C}$ to a map $\mathfrak{C}^*$ which is Hermitian and positive semidefinite, giving access to a complex Schwarz bound that can be employed in non-Hermitian complex basis function methods.

We define $\mathfrak{C}^*$ as the following Hermitian generalization of $\mathfrak{R}$ for complex functions:

$$\mathfrak{C}^*(z,w) = \int d\boldsymbol{r}_1 z^*(\boldsymbol{r}_1) \int d\boldsymbol{r}_2\, w(\boldsymbol{r}_2)\, r_{12}^{-1}. \tag{8}$$

The restriction of $\mathfrak{C}^*$ to real functions again coincides with $\mathfrak{R}$, i.e., $\mathfrak{C}^*(f,g) = \mathfrak{R}(f,g)$ for real f and g, which allows us to deduce the positive semidefiniteness of $\mathfrak{C}^*$ from the real case,

$$\begin{aligned}\mathfrak{C}^*(z,z) &= \mathfrak{R}(\mathbf{Re}(z),\mathbf{Re}(z)) + i\,\mathfrak{R}(\mathbf{Re}(z),\mathbf{Im}(z)) \\ &\quad - i\,\mathfrak{R}(\mathbf{Im}(z),\mathbf{Re}(z)) - i^2\,\mathfrak{R}(\mathbf{Im}(z),\mathbf{Im}(z)) \qquad (9)\\ &= \mathfrak{R}(\mathbf{Re}(z),\mathbf{Re}(z)) + \mathfrak{R}(\mathbf{Im}(z),\mathbf{Im}(z)) \geq 0, \qquad (10)\end{aligned}$$

where we have written z as a sum of real and imaginary parts, $z = \mathbf{Re}(z) + i\mathbf{Im}(z)$, which are real functions. Accordingly, $\mathfrak{C}^*$ fulfills the Cauchy-Schwarz inequality,

$$|\mathfrak{C}^*(z,w)|^2 \leq \mathfrak{C}^*(z,z)\mathfrak{C}^*(w,w), \tag{11}$$

in analogy to the real case. In addition, from Eq. (10), we conclude that $\mathfrak{C}^*(z,z) = \mathfrak{C}^*(z^*,z^*)$, and using the trivial relation $\mathfrak{C}(z,w) = \mathfrak{C}^*(z^*,w)$, we obtain the bound

$$\begin{aligned}|\mathfrak{C}(z,w)|^2 &= |\mathfrak{C}^*(z^*,w)|^2 \leq \mathfrak{C}^*(z^*,z^*)\mathfrak{C}^*(w,w) \\ &= \mathfrak{C}^*(z,z)\mathfrak{C}^*(w,w) = \mathfrak{C}(z^*,z)\mathfrak{C}(w^*,w). \qquad (12)\end{aligned}$$

Thus, in computational practice, the real, non-negative diagonals of the map $\mathfrak{C}^*$ can be used to bound the norms of the complex ERIs from above. For integrals over products of complex scaled basis functions, the bound takes the form

$$|(\mu\nu|\lambda\sigma)| \leq (\mu^*\nu^*|\mu\nu)^{1/2}(\lambda^*\sigma^*|\lambda\sigma)^{1/2}, \tag{13}$$

in direct analogy to Eq. (3).

The integrals $(\mu^*\nu^*|\mu\nu)$ are not required in the non-Hermitian SCF calculation itself but can be precomputed using the same machinery used for the regular complex ERIs. The complex conjugates are given by

$$\chi_\mu^*(\boldsymbol{r},\boldsymbol{A}) = N_\mu S_\mu(\boldsymbol{r}_A)\exp[-\alpha_\mu e^{+2i\theta_\mu}\boldsymbol{r}_A^2] \tag{14}$$

so that the estimates are calculated by simply substituting θ_μ with its negative in the bra.

C. Theta dependence

The complex ERIs depend on the parameter θ, and calculations are usually performed for a range of θ values. For screening purposes, the dependence of the magnitudes of the ERIs on θ is important, and here, we take a closer look at this dependence analytically.

The magnitude of a product of normalized, complex-scaled basis functions is given by

$$\begin{aligned}&|\chi_\mu(\boldsymbol{r},\boldsymbol{A})\chi_\nu(\boldsymbol{r},\boldsymbol{B})| \\ &\quad = |S_\mu(\boldsymbol{r}_A)S_\nu(\boldsymbol{r}_B)|\exp[-(\alpha_\mu\cos 2\theta_\mu \boldsymbol{r}_A^2 + \alpha_\nu\cos 2\theta_\nu \boldsymbol{r}_B^2)],\end{aligned} \tag{15}$$

where we have used Euler's formula and the fact that $|e^{ix}| = 1$ for any real-valued x. The polynomial functions in Eq. (15) can be bound[50] according to

$$|S_\mu(\boldsymbol{r}_A)| \leq C_\mu |\boldsymbol{r}_A|^{L_\mu}, \tag{16}$$

with a constant C_μ and total angular momentum L_μ. Defining $C_{\lambda\sigma}^{\mu\nu} = C_\mu C_\nu C_\lambda C_\sigma$, this leads to the following estimate for the magnitude of the complex-scaled ERI:

$$\begin{aligned}|(\mu\nu|\lambda\sigma)| \leq C_{\lambda\sigma}^{\mu\nu} &\int d\boldsymbol{r}_1 \int d\boldsymbol{r}_2 r_{12}^{-1} |\boldsymbol{r}_{1A}|^{L_\mu}|\boldsymbol{r}_{1B}|^{L_\nu}|\boldsymbol{r}_{2C}|^{L_\lambda}|\boldsymbol{r}_{2D}|^{L_\sigma} \\ &\times e^{-(\alpha_\mu\cos 2\theta_\mu \boldsymbol{r}_{1A}^2 + \alpha_\nu\cos 2\theta_\nu \boldsymbol{r}_{1B}^2 + \alpha_\lambda\cos 2\theta_\lambda \boldsymbol{r}_{2C}^2 + \alpha_\sigma\cos 2\theta_\sigma \boldsymbol{r}_{2D}^2)}.\end{aligned} \tag{17}$$

This estimate includes the important radial decay effects inherent to integrals over Gaussian basis functions, which tend to zero exponentially at distances far from the Gaussian centers. This decay is responsible for the convergence of the integral which would otherwise become infinite in the limits $\alpha_\mu + \alpha_\nu \to 0$ or $\alpha_\lambda + \alpha_\sigma \to 0$. For scaling parameters larger than zero, the cosine terms cause a decrease in the effective Gaussian exponents, leading to more diffuse functions. The integrand in (17) strictly increases as the complex-scaling parameters increase. Accordingly, the magnitudes of integrals containing complex-scaled basis functions increase with the global parameter θ, so that the number of significant integrals increases and fewer integrals can be screened. The strictly real integrals are, of course, not affected by increases in θ, and the effect of larger θ values becomes more influential as the number of complex-scaled functions in the integral increases.

Due to the symmetry of the cosine function, the estimate (17) is unchanged for a complex-conjugated bra, i.e., for the integral $(\mu^*\nu^*|\lambda\sigma)$. For this reason, the θ-dependence is contained in the Schwarz estimate (12), and it performs similarly in terms of percentages of insignificant integrals screened over the complete range of θ values and integral types. This is demonstrated in the numerical results below.

III. COMPUTATIONAL DETAILS

We have implemented the integral screening routines based on Eq. (13) into a development version of the Q-Chem program package,[51] building on the implementation presented by White, Head-Gordon, and McCurdy in Ref. 26.

In order to test the performance of our generalized Schwarz inequality for complex basis functions, we examine the ionization

Published under license by AIP Publishing

of the 2-pyridoxine 2-aminopyridine (2p2a) complex (Fig. 1) from the S22 benchmark set[52] in a static electric field. At a large enough field strength F, the ground state of the neutral molecule becomes metastable with respect to electron detachment, and no bound states exist.

The complex energies are calculated at the HF level of theory for varying distances between the 2-pyridoxine and 2-aminopyridine units in order to test the performance of the screening at various degrees of integral tensor sparsity. The distance is varied by translating all atoms of the 2-pyridoxine molecule in positive x direction. The degree of translation is denoted Δx and is given in Å. The xyz coordinates of the reference geometry, $\Delta x = 0$ Å, are available for download online.[53]

Electron correlation has been shown to have a large impact on the calculated ionization rates[41] so that the numbers presented here do not necessarily provide a qualitatively correct description of the strong-field ionization of the 2p2a complex. This inherent inaccuracy in the mean-field treatment does not present a problem for this work, as we are concerned here with the general effectiveness of our integral screening, and the calculations presented are always the first step toward more accurate correlated calculations.

The electric field is applied in the positive z direction at a field strength of $F = 0.03$ a.u. The non-Hermitian SCF energies are corrected according to the following formula from Ref. 40:

$$E_{\mathrm{corr}}(F,\theta) = E(F,\theta) - E(F=0,\theta) + E(F=0,\theta=0). \tag{18}$$

This correction accounts for finite-basis effects, which result in a nonzero imaginary part of the bound-state energy even without the presence of an electric field ($F = 0$). We have performed calculations with a range of θ values between 0° and 30°, with a step size of 1°. Optimal values θ_{opt} are determined as the value at which the numerically calculated derivative $|dE_{\mathrm{corr}}/d\theta|$ obtains its minimum in this range.

All calculations are performed with aug-cc-pVDZ[54–56] as the unscaled ($\theta_\mu = 0$) basis set. This set is augmented with 3 diffuse complex-scaled shells ($\theta_\mu = \theta$) for each atom and shell type (+3s3p3d) such that the most diffuse unscaled shell and the 3 additional shells form an even-tempered set with a spacing of 2.0. This procedure for constructing the complex basis set is outlined in Refs. 26 and 41 and is common for the treatment of electronic resonances based on the method of complex basis functions.

As an SCF convergence criterion, the error in the direct inversion of the iterative subspace (DIIS) method of Pulay[57,58] is required to be below 10^{-8} for both non-Hermitian HF and preceding HF calculations.

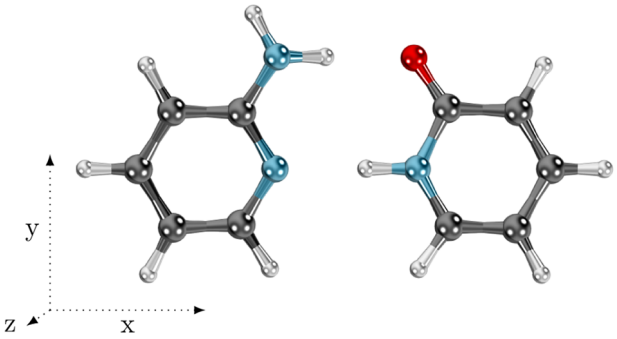


FIG. 1. 2-pyridoxine 2-aminopyridine (2p2a) complex in the xy-plane.

IV. RESULTS

A. Screening threshold dependence

As in the case of real ERIs, a screening threshold ζ is used to strike a balance between screening accuracy and efficiency. We tested the threshold dependence of the screening for the 2p2a complex with $\Delta x = 6$ Å, which is chosen because it contains a wide range of distances between basis function centers. A reference value is obtained using the very tight threshold $\zeta = 10^{-14}$. In Table I, the optimal θ values and differences in the real and imaginary parts of the energies are given for thresholds down to $\zeta = 10^{-9}$. For $\zeta = 10^{-8}$, the non-Hermitian SCF no longer converges, which is to be expected with the SCF convergence criterion used (see Sec. III). Integrals are screened in a shellwise manner; i.e., basis functions are grouped into shells with the same exponent and total angular momentum and the maximum estimate for the integrals within a shell-quartet is used to determine if the entire shell-quartet can be neglected or not. We give the percentages of shell-quartets that were screened (% Screened) and the percentages of shell-quartets that were not screened and whose magnitudes were determined to be less than the integral threshold after exact calculation (% Missed). In both cases, the percentages are calculated with respect to the formal number of shell-pair combinations,

TABLE I. Comparison of non-Hermitian HF calculations on the 2p2a complex ($\Delta x = 6$ Å) for various screening thresholds ζ. In the first row, the reference energies Re(E) and Im(E) are given, while the remaining rows show differences to the reference values for looser thresholds.

ζ	θ_{opt} (deg)	Re(E) (a.u.)	Im(E) (a.u.)	% Screened	% Missed
1×10^{-14}	6.0	−623.431 352 234 9	−0.000 239 962 7	15.54	0.29
1×10^{-13}	6.0	-3.1×10^{-12}	-5.5×10^{-12}	15.98	0.46
1×10^{-12}	6.0	-5.2×10^{-11}	2.2×10^{-11}	16.74	0.68
1×10^{-11}	6.0	4.6×10^{-9}	-1.1×10^{-9}	17.96	0.89
1×10^{-10}	6.0	7.7×10^{-9}	-1.4×10^{-8}	19.66	1.09
1×10^{-9}	6.0	-2.8×10^{-7}	2.8×10^{-8}	21.87	1.33

Published under license by AIP Publishing

TABLE II. Comparison of non-Hermitian HF calculations on the 2p2a complex using a screening threshold of $\zeta = 10^{-11}$ with reference calculations using $\zeta = 10^{-14}$ for various values of Δx.

Δx	$\theta_{\text{opt}}^{\text{ref}}$ (deg)	θ_{opt} (deg)	Re(E) (Error) (a.u.)	Im(E) (Error) (a.u.)	% Screened	% Missed
0	6.0	6.0	−623.452 100 1(2.1×10^{-10})	−0.000 219 8(−1.1×10^{-10})	3.66	0.42
3	9.0	9.0	−623.432 584 8(1.6×10^{-10})	−0.000 224 9(4.3×10^{-10})	9.93	0.63
6	6.0	6.0	−623.431 352 2(4.6×10^{-9})	−0.000 240 0(−1.1×10^{-9})	17.96	0.89
9	6.0	6.0	−623.431 200 2(7.7×10^{-10})	−0.000 239 4(6.5×10^{-10})	27.42	1.05
12	6.0	6.0	−623.431 187 2(−2.8×10^{-9})	−0.000 242 5(2.2×10^{-10})	36.86	1.21
15	6.0	6.0	−623.431 175 2(3.9×10^{-9})	−0.000 245 0(−8.4×10^{-10})	45.68	1.11
18	6.0	6.0	−623.431 192 6(3.2×10^{-9})	−0.000 246 5(1.9×10^{-9})	52.59	1.03
21	9.0	9.0	−623.431 123 6(3.3×10^{-9})	−0.000 267 3(−2.7×10^{-10})	57.29	0.93

$$\left(\frac{N_s + N_s^2}{2}\right)^2, \tag{19}$$

where N_s is the total number of shells. For the present system, this corresponds to a total of just over $4 \cdot 10^9$ shell-quartets. As expected, more integrals are screened with looser thresholds, while the magnitudes of the deviations from the reference also increases. Encouraging is the fact that the optimal value for θ remains the same for all thresholds, and the errors in the real and imaginary parts of the energies are always less than 10^{-6} a.u.

B. Δx and θ dependence

In order to test the performance of the Schwarz screening for a variety of Δx and θ values, we used a fixed threshold of $\zeta = 10^{-11}$ and calculated errors compared with corresponding reference calculations using $\zeta = 10^{-14}$. In Table II, the optimal θ values of the reference and test calculations are given for each Δx value tested, along with the deviations to the reference in the complex energies. The screening performance metrics % Screened and % Missed defined above are also listed. The optimal θ values are always the same as for the reference, and the errors are low in each case. As expected, the number of integrals that can be screened increases dramatically with the distance between the monomers, as the overlap in the basis functions diminishes. The small % Missed values are a testament to the quality of the Schwarz estimate for the complex ERI tensor.

In Fig. 2, the errors induced by the screening for all Δx and θ values are given in more detail. The errors tend to be slightly higher for larger separations between the monomers, which is expected due to the larger number of integrals screened. More interestingly, we see that the errors tend to increase with the value of θ despite the fact that fewer integrals are screened as θ increases. This is most likely due to the increased diffuseness of the complex scaled functions for larger θ values (see Sec. II C), which generally increases the magnitude of the integrals, also those that are discarded. This means that the discarded integrals are on average closer to the screening threshold in magnitude. While the errors remain acceptable, it makes sense to consider using tighter screening thresholds as θ increases.

In Fig. 3, the screening metrics for all Δx and θ values are shown. As expected, the number of integrals screened is higher for larger separations and smaller θ values. The low % Missed values show that the inequality performs quite well over these ranges of values.

C. Performance for different integral classes

When using complex-scaled basis functions, the arising integrals can be divided into six different classes depending on the number of complex-scaled functions in both bra and ket. For a pair of shells, one can distinguish between the number of complex-scaled shells present $N^C \in \{0, 1, 2\}$. This leads to nine categories of shell-quartets N_B^C-N_K^C based on the number of complex-scaled shells

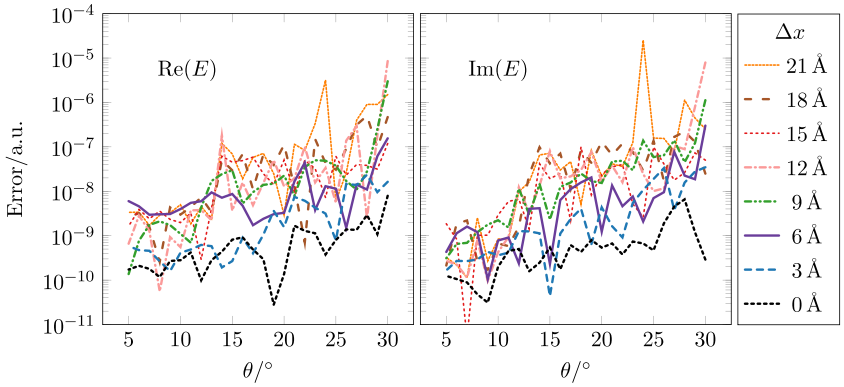


FIG. 2. Errors in the real and imaginary parts of the non-Hermitian HF energies for various Δx and θ values. A screening threshold of $\zeta = 10^{-11}$ is used, and errors are given with respect to the references with $\zeta = 10^{-14}$.

Published under license by AIP Publishing

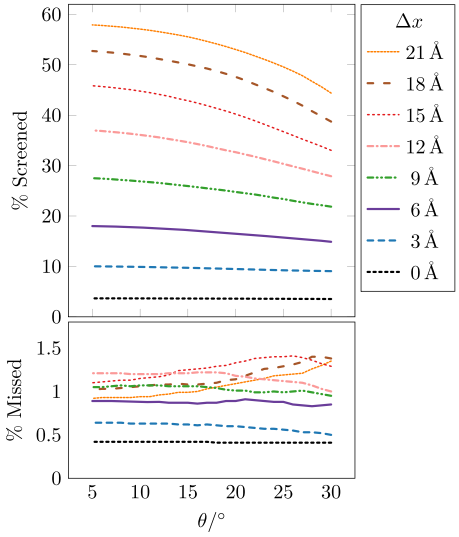


FIG. 3. Percentages of screened integrals (% Screened) and unscreened integrals that were actually negligible (% Missed) for various Δx and θ values. A screening threshold of $\zeta = 10^{-11}$ is used.

in the bra N_B^C and ket N_K^C. For example, the category 2-1 contains integrals with two complex functions in the bra and one in the ket. Due to symmetry, three of these categories are redundant and the six classes given in Table III result.

Because the complex-scaled functions are in general much more diffuse and become even more diffuse as θ increases, it is interesting to study the effects of the Schwarz screening for each class separately. In Fig. 4, screening performance data are given for each of the integral classes. In addition to the % Screened and % Missed metrics, the integrals that were not screened are broken down into percentages for each class. As expected from our analysis in Sec. II C, the more complex-scaled shells are present, the less screening is possible. As θ increases, the percentages of screened integrals decrease for all classes, except 0:0, because these completely real integrals are unaffected by θ. In the % Missed category, it is apparent that the screening performs worse when mixed real/complex shell-pairs are present. This may be due to an increased importance in operator-based distance decay for such shell-pairs, which will tend to be

TABLE III. Unique integral classes based on the numbers of complex-scaled shells. The number of shell-quartets is given for the 2p2a test system described in Sec. III.

Integral class	Included categories	No. of shell-quartets	% of total
0:0	0-0	196 784 784	4.7
1:0	1-0, 0-1	899 587 584	21.5
2:0	2-0, 0-2	519 821 568	12.5
1:1	1-1	1 028 100 096	24.6
2:1	2-1, 1-2	1 188 163 584	28.5
2:2	2-2	343 286 784	8.2

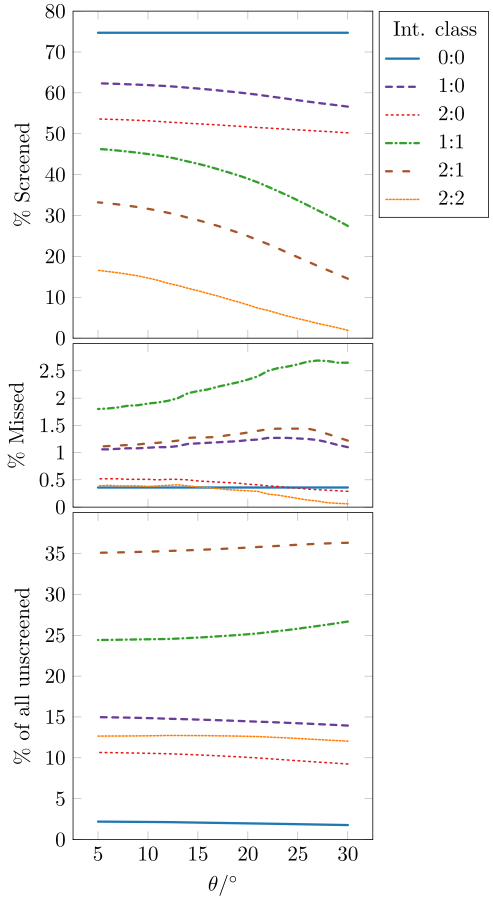


FIG. 4. Percentages of screened integrals (% Screened), percentages of unscreened integrals that were actually negligible (% Missed), and percentages that each integral class makes up in the unscreened integrals (% of all unscreened) for the six classes of integrals (see discussion in Sec. IV C) and various θ values. A screening threshold of $\zeta = 10^{-11}$ is used and $\Delta x = 15$ Å.

mostly described by the much less diffuse real function. For such integrals, a distance-dependent treatment[48,50,59] may be warranted, although for SCF calculations with the Coulomb operator $1/r_{12}$ used here, the performance of the Schwarz inequality is still quite good. The percentage of all unscreened data given at the bottom of Fig. 4 show that because the integrals containing fewer complex-scaled functions are more often negligible, the use of an efficient screening shifts the class percentages to those classes containing more complex functions (compare with the formal class percentages in Table III). This indicates that increasing the efficiency of the calculation of complex-scaled integrals should be an important goal of future work.

V. CONCLUSION AND OUTLOOK

We have presented an extension of the Schwarz bound that is applicable to complex basis functions that appear in non-Hermitian extensions of quantum mechanics to treat electronic resonances. We have implemented a screening procedure based on this bound and

Published under license by AIP Publishing

shown that it performs well when used to screen the complex ERIs with various distances between basis function centers and various values of the complex-scaling parameter θ.

In future work, we anticipate that the combination of this bound with density-matrix based screening techniques will afford the reduction of the asymptotic scaling of the integral work in non-Hermitian SCF methods that employ complex-scaled basis functions to linear, as has already been accomplished for the corresponding bound-state methods.[30–32,49] Such developments will become increasingly important for transforming the ideas of non-Hermitian quantum mechanics[2] into electronic-structure methods[1] suitable for the treatment of resonances in large systems.

ACKNOWLEDGMENTS

This work has been supported by the DFG Excellence Cluster EXC2111 (Munich Center for Quantum Science and Technology—MCQST) (T. H. Thompson and C. Ochsenfeld) and through the DFG Grant No. JA 2794/1-1 within the Emmy Noether program (T.-C. Jagau).

REFERENCES

[1] T.-C. Jagau, K. B. Bravaya, and A. I. Krylov, Annu. Rev. Phys. Chem. **68**, 525 (2017).
[2] N. Moiseyev, *Non-Hermitian Quantum Mechanics*, 1st ed. (Cambridge University Press, 2011).
[3] J. P. Marangos, J. Phys. B: At., Mol. Opt. Phys. **49**, 132001 (2016).
[4] T. Zuo, A. Bandrauk, and P. Corkum, Chem. Phys. Lett. **259**, 313 (1996).
[5] Z. Vager, R. Naaman, and E. P. Kanter, Science **244**, 426 (1989).
[6] N. Moiseyev, Phys. Rep. **302**, 212 (1998).
[7] A. J. F. Siegert, Phys. Rev. **56**, 750 (1939).
[8] E. Balslev and J. M. Combes, Commun. Math. Phys. **22**, 280 (1971).
[9] J. Aguilar and J. M. Combes, Commun. Math. Phys. **22**, 269 (1971).
[10] B. Simon, Commun. Math. Phys. **27**, 1 (1972).
[11] I. W. Herbst and B. Simon, Phys. Rev. Lett. **41**, 67 (1978).
[12] I. W. Herbst, Commun. Math. Phys. **64**, 279 (1979).
[13] I. W. Herbst and B. Simon, Commun. Math. Phys. **80**, 181 (1981).
[14] B. Simon, Phys. Lett. A **71**, 211 (1979).
[15] C. W. McCurdy and T. N. Rescigno, Phys. Rev. Lett. **41**, 1364 (1978).
[16] K. B. Bravaya, D. Zuev, E. Epifanovsky, and A. I. Krylov, J. Chem. Phys. **138**, 124106 (2013).
[17] N. Moiseyev and C. Corcoran, Phys. Rev. A **20**, 814 (1979).
[18] T. N. Rescigno, A. E. Orel, and C. W. McCurdy, J. Chem. Phys. **73**, 6347 (1980).
[19] C. W. McCurdy, T. N. Rescigno, E. R. Davidson, and J. G. Lauderdale, J. Chem. Phys. **73**, 3268 (1980).
[20] M. Mishra, Y. Öhrn, and P. Froelich, Phys. Lett. A **84**, 4 (1981).
[21] C. W. McCurdy and R. C. Mowrey, Phys. Rev. A **25**, 2529 (1982).
[22] J. G. Lauderdale, C. W. McCurdy, and A. U. Hazi, J. Chem. Phys. **79**, 2200 (1983).
[23] M. Honigmann, R. J. Buenker, and H.-P. Liebermann, J. Chem. Phys. **125**, 234304 (2006).
[24] M. Honigmann, R. J. Buenker, and H.-P. Liebermann, J. Chem. Phys. **131**, 034303 (2009).
[25] A. F. White, C. W. McCurdy, and M. Head-Gordon, J. Chem. Phys. **143**, 074103 (2015).
[26] A. F. White, M. Head-Gordon, and C. W. McCurdy, J. Chem. Phys. **142**, 054103 (2015).
[27] A. F. White, E. Epifanovsky, C. W. McCurdy, and M. Head-Gordon, J. Chem. Phys. **146**, 234107 (2017).
[28] J. L. Whitten, J. Chem. Phys. **58**, 4496 (1973).
[29] M. Häser and R. Ahlrichs, J. Comput. Chem. **10**, 104 (1989).
[30] C. Ochsenfeld, C. A. White, and M. Head-Gordon, J. Chem. Phys. **109**, 1663 (1998).
[31] C. Ochsenfeld, Chem. Phys. Lett. **327**, 216 (2000).
[32] J. Kussmann and C. Ochsenfeld, J. Chem. Phys. **138**, 134114 (2013).
[33] A. S. Erik, I. Tellgren, and T. Helgaker, J. Chem. Phys. **129**, 154114 (2008).
[34] S. Stopkowicz, J. Gauss, K. K. Lange, E. I. Tellgren, and T. Helgaker, J. Chem. Phys. **143**, 074110 (2015).
[35] P. Čársky and M. Polášek, J. Comput. Phys. **143**, 266 (1998).
[36] C. Winstead and V. McKoy, Comput. Phys. Commun. **128**, 386 (2000).
[37] M. Tachikawa and M. Shiga, Phys. Rev. E **64**, 056706 (2001).
[38] L. Füsti-Molnar and P. Pulay, J. Phys. Chem. **116**, 7795 (2002).
[39] C. Winstead and V. McKoy, in *Modern Electronic Structure Theory: Part II*, edited by D. R. Yarkony (World Scientific, 1995), Chap. 22, pp. 1375–1462.
[40] T.-C. Jagau, J. Chem. Phys. **145**, 204115 (2016).
[41] T.-C. Jagau, J. Chem. Phys. **148**, 204102 (2018).
[42] M. Hernandez Vera and T.-C. Jagau, J. Chem. Phys. **151**, 111101 (2019).
[43] N. Moiseyev, P. Certain, and F. Weinhold, Mol. Phys. **36**, 1613 (1978).
[44] N. Moiseyev, Mol. Phys. **47**, 585 (1982).
[45] In non-Hermitian quantum mechanics, it is common practice to indicate the c-product by using parentheses instead of chevrons. To avoid confusion with Mulliken notation, we do not follow this practice.
[46] N. Moiseyev, P. R. Certain, and F. Weinhold, Int. J. Quantum Chem. **14**, 727 (1978).
[47] E. Lieb and M. Loss, *Analysis*, CRM Proceedings & Lecture Notes (American Mathematical Society, 2001).
[48] T. H. Thompson and C. Ochsenfeld, J. Chem. Phys. **147**, 144101 (2017).
[49] J. Kussmann, M. Beer, and C. Ochsenfeld, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **3**, 614 (2013).
[50] T. H. Thompson and C. Ochsenfeld, J. Chem. Phys. **150**, 044101 (2019).
[51] Y. Shao, Z. Gan, E. Epifanovsky, A. T. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kuś, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. Lee Woodcock III, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio, Jr., H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. Hanson-Heine, P. H. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. D. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt T. Stein, D. Stück, Y.-C. Su, A. J. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard III, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer III, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. V. Voorhis, J. M. Herbert, A. I. Krylov, P. M. Gill, and M. Head-Gordon Mol. Phys. **113**, 184 (2015).
[52] P. Jurečka, J. Šponer, J. Černý, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).

Published under license by AIP Publishing

[53]See http://www.begdb.com, where the molecular coordinates of a variety of benchmark sets are available for download, including the S22 set from which the 2-pyridoxine 2-aminopyridine molecule was taken.
[54]T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).
[55]R. A. Kendall, T. H. Dunning, and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).
[56]D. E. Woon and T. H. Dunning, J. Chem. Phys. **100**, 2975 (1994).
[57]P. Pulay, Chem. Phys. Lett. **73**, 393 (1980).
[58]P. Pulay, J. Comput. Chem. **3**, 556 (1982).
[59]S. A. Maurer, D. S. Lambrecht, D. Flaig, and C. Ochsenfeld, J. Chem. Phys. **136**, 144107 (2012).

Published under license by AIP Publishing

4.4 Publication IV

Highly efficient, linear scaling semi-numerical exact-exchange method for graphic processing units

H. Laqua, T. H. Thompson, J. Kussmann, and C. Ochsenfeld,
J. Chem. Theor. Comput. **16**, 1456-1468 (2020).

Abstract:

We present a highly efficient and asymptotically linear-scaling graphic processing unit accelerated seminumerical exact-exchange method (sn-LinK). We go beyond our previous central processing unit-based method (Laqua, H.; Kussmann, J.; Ochsenfeld, C. J. Chem. Theory Comput. 2018, **14**, 3451-3458) by employing our recently developed integral bounds (Thompson, T. H.; Ochsenfeld, C. J. Chem. Phys. 2019, **150**, 044101) and high-accuracy numerical integration grid (Laqua, H.; Kussmann, J.; Ochsenfeld, C. J. Chem. Phys. 2018, **149**, 204111). The accuracy is assessed for several established test sets, providing errors significantly below 1 mE_h for the smallest grid. Moreover, a comprehensive performance analysis for large molecules between 62 and 1347 atoms is provided, revealing the outstanding performance of our method, in particular, for large basis sets such as the polarized quadruple-zeta level with diffuse functions.

The following article is reprinted with permission from (H. Laqua, T. H. Thompson, J. Kussmann, and C. Ochsenfeld, *J. Chem. Theor. Comput.* **16**, 1456-1468 (2020).). Copyright (2020) American Chemical Society.

The article can be found online at: https://doi.org/10.1021/acs.jctc.9b00860

JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

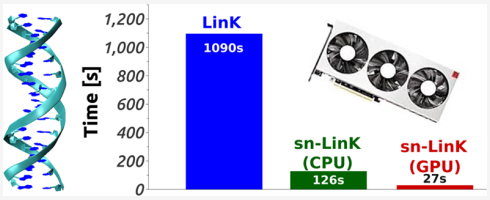
Highly Efficient, Linear-Scaling Seminumerical Exact-Exchange Method for Graphic Processing Units

Henryk Laqua, Travis H. Thompson, Jörg Kussmann, and Christian Ochsenfeld*

Cite This: https://dx.doi.org/10.1021/acs.jctc.9b00860 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We present a highly efficient and asymptotically linear-scaling graphic processing unit accelerated seminumerical exact-exchange method (sn-LinK). We go beyond our previous central processing unit-based method (Laqua, H.; Kussmann, J.; Ochsenfeld, C. *J. Chem. Theory Comput.* **2018**, *14,* 3451−3458) by employing our recently developed integral bounds (Thompson, T. H.; Ochsenfeld, C. *J. Chem. Phys.* **2019**, *150,* 044101) and high-accuracy numerical integration grid (Laqua, H.; Kussmann, J.; Ochsenfeld, C. *J. Chem. Phys.* **2018**, *149,* 204111). The accuracy is assessed for several established test sets, providing errors significantly below $1mE_h$ for the smallest grid. Moreover, a comprehensive performance analysis for large molecules between 62 and 1347 atoms is provided, revealing the outstanding performance of our method, in particular, for large basis sets such as the polarized quadruple-zeta level with diffuse functions.



1. INTRODUCTION

During the last 15 years, graphic processing units (GPUs) have gained increasing interest within the quantum chemistry community, focusing, in particular, on the evaluation of 4-center-2-electron (4c-2e) integrals, which represent the major bottleneck in most Hartree−Fock and Kohn−Sham calculations.[1−15] Since, for Hartree−Fock (HF) and hybrid density functional theory (DFT) calculations, the mandatory computation of exact (Fock-like) exchange matrices is particularly expensive, efficient and linear-scaling implementations have been developed since the late 1990s[16−21] including recent developments.[7,9] However, for larger molecules and particularly in combination with larger basis sets (i.e., triple-ζ or larger), resolution-of-the-identity (RI)[22,23] or seminumerical methods, that is, grid-based methods employing 3-center-1-electron (3c-1e) integrals,[24−46] are possibly more efficient due to their superior $O(N_{bas}^2)$ formal scaling compared to the formal $O(N_{bas}^4)$ scaling of the conventional 4c-2e integral-based methods.

As we demonstrated recently,[44] seminumerical exchange methods can, in contrast to the asymptotically $O(M^3)$ scaling RI-K method,[22,23] be implemented in an asymptotically linear-scaling fashion. This is an increasingly important property since modern computer hardware now allows for the routine calculation of multiple thousand atoms on conventional server nodes or workstations. In addition, seminumerical methods may directly be employed to compute the exact-exchange part of local hybrid functionals, which represent a very promising new class of functionals due to their higher variability and therefore more general applicability.[47−58] Indeed, the prospects of these new functionals have been the major motivation for many recently developed seminumerical methods.[38−40,42−46]

In this publication, we present a reformulation of our previous method[44] that allows for an efficient and highly performant GPU implementation. These changes include the use of our recently developed generalized integral bounds[59] and our improved molecular grids.[60−62] Not only were these new techniques necessary for a performant GPU implementation but they are also applied to our existing central processing unit (CPU) implementation, in this way, further improving its performance as well, especially if run on modern CPUs, which provide an ever-increasing support for single-instruction-multiple-data (SIMD) vector instructions. Particularly, the batch-wise integral selection, which we pioneered in our previous work[44] and refined in this work, is essential for a highly efficient and performant implementation on SIMD computer architectures, such as GPUs and modern CPUs. That is, our new method exploits the superior computing performance of SIMD computer architectures while maintaining the asymptotic linear-scaling behavior of our previous work.

The paper is organized as follows: we begin with a brief review of the theory underlying the seminumerical method in Section 2.1, followed by the description of our revised integral screening in Sections 2.2 and 2.3, and our newly developed prescreening method in Section 2.4. Subsequently, we provide an outline of our GPU implementation for the Compute Unified Device Architecture (CUDA)[63] and the Open Computing Language (OpenCL)[64,65] frameworks in Section

Received: August 28, 2019

Downloaded via LMU MUENCHEN on February 17, 2020 at 17:51:13 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

ACS Publications © XXXX American Chemical Society

3, focusing on the particular techniques employed to maximize performance. Finally, we assess our new accelerated semi-numerical exchange method, denoted as sn-LinK, in terms of the accuracy of the numerical integration in Section 5.1 and in terms of performance in Sections 5.2.1 to 5.2.5. For simplicity, we restrict the discussion within this paper to the computation of Fock exchange within the Hartree−Fock theory, noting that the application to global and local hybrid DFT is straightforward.

2. THEORY

2.1. Seminumerical Exchange Matrix. The exchange matrix is given in the atomic orbital (AO) basis as

$$K_{\mu\nu} = \sum_{\lambda\sigma} P_{\lambda\sigma}(\mu\sigma|\nu\lambda) \tag{1}$$

where μ, ν, λ, and σ represent AO basis function indices, and the 4-center-2-electron integral $(\mu\sigma \mid \nu\lambda)$ is defined as

$$(\mu\sigma|\nu\lambda) = \int\int \chi_\mu(\mathbf{r}_1)\chi_\sigma(\mathbf{r}_1)\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}\chi_\nu(\mathbf{r}_2)\chi_\lambda(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2 \tag{2}$$

Within the seminumerical ansatz, the integration over one of the coordinates ($\mathbf{r}_1$ or $\mathbf{r}_2$) is performed analytically, and the other one is performed numerically by employing discrete grid points with coordinates $\mathbf{r}_g$ and weights w_g. To preserve all the symmetries within the integral tensor, this decomposition is performed symmetrically over both coordinates $\mathbf{r}_1$ and $\mathbf{r}_2$, leading to

$$(\mu\sigma|\nu\lambda) \approx \frac{1}{2}\left[\sum_g w_g\chi_\mu(\mathbf{r}_g)\chi_\sigma(\mathbf{r}_g)\left(\int \frac{\chi_\nu(\mathbf{r})\chi_\lambda(\mathbf{r})}{|\mathbf{r}_g - \mathbf{r}|}d\mathbf{r}\right) + \sum_g w_g\left(\int \frac{\chi_\mu(\mathbf{r})\chi_\sigma(\mathbf{r})}{|\mathbf{r}_g - \mathbf{r}|}d\mathbf{r}\right)\chi_\nu(\mathbf{r}_g)\chi_\lambda(\mathbf{r}_g)\right] \tag{3}$$

Inserting eq 3 into the definition of the exchange matrix (eq 1) yields

$$K_{\mu\nu} \approx \frac{1}{2}\left[\sum_g w_g \sum_{\lambda\sigma}\chi_\mu(\mathbf{r}_g)\left(\int \frac{\chi_\nu(\mathbf{r})\chi_\lambda(\mathbf{r})}{|\mathbf{r} - \mathbf{r}_g|}d\mathbf{r}\right)P_{\lambda\sigma}\chi_\sigma(\mathbf{r}_g) + \text{transpose}\right] \tag{4}$$

where the transpose is the result of the symmetric integral decomposition in eq 3 in combination with the symmetry of the ground-state density matrix $P_{\mu\nu} = P_{\nu\mu}$.

The exchange matrix may thus be computed in three consecutive steps

$$F_{\lambda g} = \sum_\sigma \chi_\sigma(\mathbf{r}_g)P_{\lambda\sigma} \tag{5}$$

$$G_{\nu g} = \sum_\lambda w_g A_{\nu\lambda g}F_{\lambda g} \tag{6}$$

$$K_{\mu\nu} = \sum_g \chi_\mu(\mathbf{r}_g)G_{\nu g} \tag{7}$$

where $A_{\nu\lambda g}$ denotes 3-center-1-electron integrals of the form

$$A_{\nu\lambda g} = \int \frac{\chi_\nu(\mathbf{r})\chi_\lambda(\mathbf{r})}{|\mathbf{r}_g - \mathbf{r}|}\mathrm{d}\mathbf{r} \tag{8}$$

The integrals $A_{\nu\lambda g}$ are evaluated on-the-fly using optimized automatically generated Obara−Saika[66,67] recursions for the different l-quantum number combinations. The so-obtained exchange matrix K is subsequently symmetrized as

$$K^{\text{symm}}_{\mu\nu} = \frac{1}{2}(K_{\mu\nu} + K_{\nu\mu}) \tag{9}$$

to account for the transpose in eq 4.

2.2. Integral Screening. The integral screening of our previous method[44] targeted only the exchange energy, that is, significant integrals were selected solely based on their contribution to the exchange energy

$$\begin{aligned}\varepsilon^E_{\nu\lambda g} &= \left|w_g \sum_{\mu\sigma}\chi_\mu(\mathbf{r}_g)P_{\mu\nu}A_{\nu\lambda g}P_{\lambda\sigma}\chi_\sigma(\mathbf{r}_g)\right| \\ &= |w_g^{1/2}F_{\nu g}A_{\nu\lambda g}w_g^{1/2}F_{\lambda g}|\end{aligned} \tag{10}$$

This is a simple and symmetrical expression, which provides the tightest screening possible if one is only interested in energies.

However, during the self-consistent field (SCF) iterations, a high accuracy in the exchange potential matrix is also desirable, in particular if larger basis sets are employed. In analogy to the conventional (4c-2e integral-based) LinK method,[20] we revised our scheme to also include contributions to the exchange matrix K into the screening

$$\begin{aligned}\varepsilon^K_{\nu\lambda g} &= |w_g| \max\left(\sum_{\mu\sigma}|\chi_\mu(\mathbf{r}_g)||P_{\mu\nu}||A_{\nu\lambda g}||\chi_\sigma(\mathbf{r}_g)|, \right. \\ &\qquad \left. \sum_{\mu\sigma}|\chi_\mu(\mathbf{r}_g)||A_{\nu\lambda g}||P_{\lambda\sigma}||\chi_\sigma(\mathbf{r}_g)|\right) \\ &\le |w_g| \max(|F_{\nu g}|,|F_{\lambda g}|)|A_{\nu\lambda g}|\sum_\mu|\chi_\mu(\mathbf{r}_g)|\end{aligned} \tag{11}$$

In our new implementation, a 3c-1e integral $A_{\nu\lambda g}$ is labeled as significant if it is significant by either eq 10 or 11, that is

$$\varepsilon^E_{\nu\lambda g} \ge \vartheta_E \vee \varepsilon^K_{\nu\lambda g} \ge \vartheta_K \tag{12}$$

employing two different thresholds ϑ_E and ϑ_K for each criterion. Since, during the SCF, both the exchange matrix and the exchange energy are of interest, we employ both eqs 10 and 11 for the screening during the SCF, whereas for the final energy calculation (which is typically performed on a larger grid), we screen only for the energy (eq 10). A similar optimization was also described in ref 36.

In order for the screening to be efficient, not every integral is inspected individually; instead, a whole batch of spatially adjacent grid points is considered at once, which reduces the screening overhead to an insignificant amount (<5% of the total cost of the integral evaluation). For this purpose, the maximum contribution within a whole batch b of points has to be estimated, that is

https://dx.doi.org/10.1021/acs.jctc.9b00860
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

$$\varepsilon^{E}_{\nu\lambda b} = \max_{g\in b}(\varepsilon^{E}_{\nu\lambda g}) = \max_{g\in b}(|w_g^{1/2}F_{\nu g}|)\max_{g\in b}(|w_g^{1/2}F_{\lambda g}|)\max_{g\in b}(|A_{\nu\lambda g}|) \quad (13)$$

and

$$\varepsilon^{K}_{\nu\lambda b} = \max_{g\in b}(\varepsilon^{K}_{\nu\lambda g}) = \max(\max_{g\in b}(|w_g^{1/2}F_{\nu g}|), \max_{g\in b}(|w_g^{1/2}F_{\lambda g}|)) \max_{g\in b}\left(|w_g^{1/2}|\sum_{\mu}|\chi_{\mu}(\mathbf{r}_g)|\right)\max_{g\in b}(|A_{\nu\lambda g}|) \quad (14)$$

The necessary quantities $\max_{g\in b}(|w_g^{1/2}F_{\nu g}|)$ and $\max_{g\in b}\left(|w_g^{1/2}|\sum_{\mu}|\chi_{\mu}(\mathbf{r}_g)|\right)$ are trivially precomputed from F and $\chi_{\mu}(\mathbf{r}_g)$, and a rigorous upper bound for the integral $\max_{g\in b}(|A_{\nu\lambda g}|)$ is obtained as a special case of our recently developed partition bounds for many classes of electronic integrals.[59] This bound is briefly described below.

2.3. Integral Bounds for 3c-1e Integrals. In ref 59, two different rigorous bounds result for $|A_{\nu\lambda g}|$. A distance-dependent bound that captures the Coulomb decay of far away grid points can be formulated by calculating rigorous centers and extents for each shell pair. The resulting bounds are very tight but also necessarily batch-dependent.

A simpler bound that is independent of $\mathbf{r}_g$, and therefore also batch-independent, is given by

$$\max_{g\in b}(|A_{\nu\lambda g}|) \le \max_{\mathbf{r}_g\in\mathbb{R}^3}\left(\int\frac{|\chi_{\nu}(\mathbf{r})\chi_{\lambda}(\mathbf{r})|}{|\mathbf{r}_g-\mathbf{r}|}d\mathbf{r}\right) = \mathcal{V}_{\nu\lambda} \quad (15)$$

The batch independency of 15 allows to decrease the complexity of the screening algorithm because it only relies on precomputable shell-pair quantities. This is particularly useful for a high-performance GPU implementation, where algorithm complexity should be kept to a minimum. The equation necessary to compute $\mathcal{V}_{\nu\lambda}$ is given in Appendix A of ref 59. In short, we bound $\mathcal{V}_{\nu\lambda}$ by simpler integrals over spherically symmetric functions and use the fact that for any function S that is spherically symmetric with respect to a point $\mathbf{p}$, one can show that

$$\max_{\mathbf{r}_g\in\mathbb{R}^3}\left(\int\frac{|S(\mathbf{r})|}{|\mathbf{r}_g-\mathbf{r}|}d\mathbf{r}\right) = \int\frac{|S(\mathbf{r})|}{|\mathbf{p}-\mathbf{r}|}d\mathbf{r} \quad (16)$$

i.e., the maximum is always achieved at the spherical center.

In contrast to our previous screening scheme,[44] our new screening employing (eq 15) is completely rigorous and requires only a few multiplications for each grid batch and shell pair instead of a full integral evaluation, which better suits the parallel architecture of GPUs. In our sn-LinK method, this screening is performed in a hierarchical way, that is, a set of significant shell pairs is first selected for a large batch (typically around 10,000 to 20,000 grid points) on the CPU and subsequently a tighter selection for sub-batches of 64 grid points is made on-the-fly on the GPU.

This batch-wise integral selection is essential for both an efficient CPU and, to even greater extent, an efficient GPU implementation since it does not interfere with single-instruction-multiple-data (SIMD) vector instructions because identical branching within one sub-batch is guaranteed.

2.4. Prescreening. Having determined a tight screening for 3c-1e integrals (i.e., the shell pairs $\nu\lambda$ involved in eq 6), we now consider the screening of the set of indices μ/σ and ν/λ to also guarantee the asymptotic linear-scaling evaluation of eqs 5 and 7.

For each given batch of grid points, the set containing all the significant basis function indices μ is identical to the set containing the indices σ. These sets, which we refer to as $\{\mu\}$, are determined solely by the extent of the AO basis functions, that is, only functions χ_{μ} with a significant basis function value within a given batch are labeled significant. Due to the exponential decay of Gaussian-type AO basis functions, the size of the set $\{\mu\}$ is asymptotically constant for any given grid batch.

Analogously, the sets of significant basis function indices ν are also identical to the sets containing the indices λ. However, these sets (denoted as $\{\nu\}$) cannot be determined by the basis function extents since they couple indirectly via $\{\mu\}$ and the density matrix element $P_{\mu\nu}$ or, equivalently, by the extent of the exchange hole. Therefore, depending on the electronic structure of the system, a variable amount of basis function shells need to be considered in $\{\nu\}$.

For this preselection, we simply select all the significant shell pairs $\nu\lambda$ by eq 13 or 14 and subsequently select the set of all shells that contribute to at least one of these shell pairs since all other shells do not contribute to the exchange matrix at all because all of their contributions would be screened out by the integral screening later anyhow. This method thus introduces no additional error and is sufficient to ensure asymptotic linear scaling (and constant memory scaling) since only a constant amount of shell pairs $\nu\lambda$ are significant for each batch. That is, all batch-wise quantities are of asymptotically constant size, resulting in a constant workload per batch and therefore in an overall linear scaling since the amount of grid batches scales linearly with the system size.

One complication arises from the fact that the intermediate quantity $\boldsymbol{F}$ is required for the preselection, while the set $\{\nu\}$ needs to be known prior to the computation of $\boldsymbol{F}$, since $\{\nu\}$ is just the set of significant entries in $\boldsymbol{F}$. Therefore, an upper bound for the absolute value of $\boldsymbol{F}$ that can be computed at low cost prior to the computation of $\boldsymbol{F}$ is required. For this quantity, we choose the batch-wise maximum of $\boldsymbol{F}$, that is

$$\max_{g\in b}(|w_g^{1/2}\boldsymbol{F}_{\lambda g}|) = \max_{g\in b}\left(\left|\sum_{\sigma}P_{\lambda\sigma}w_g^{1/2}\chi_{\sigma}(\mathbf{r}_g)\right|\right) \le \sum_{\sigma}|P_{\lambda\sigma}|\max_{g\in b}(|w_g^{1/2}\chi_{\sigma}(\mathbf{r}_g)|) \quad (17)$$

Therefore, an upper bound for $\boldsymbol{F}$ can be obtained by only one matrix-vector multiplication of $|\boldsymbol{P}|$ with $\max_{g\in b}(|w_g^{1/2}\chi_{\sigma}(\mathbf{r}_g)|)$ for each batch.

3. IMPLEMENTATION

We implemented the above described sn-LinK method within our C++-based FermiONs++ program,[7,9] revising our CPU-based local hybrid implementation described in ref 44. Our implementation for the AMD GPUs is based on OpenCL,[64] whereas our NVIDIA GPU implementation employs CUDA[63]

https://dx.doi.org/10.1021/acs.jctc.9b00860
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

since we found OpenCL to be less performant on NVIDIA GPUs. All steps are performed exclusively with double precision (fp64) to obtain reliable results and to allow for tight convergence even with large basis sets.

In contrast to the analytical integral-direct method, which consists of only one compute-intensive step, the seminumerical implementation contains multiple bottlenecks, that is, the three steps of eqs 5−7 and the evaluation of the basis function values $\chi_\mu(\mathbf{r}_g)$. Therefore, all these steps have to be performed on the GPU to minimize bottlenecks from the CPU and the CPU-GPU data transfer. Additionally, we decided to use multiple concurrent streams of instructions on each GPU, which allows for the data transfer of one stream to be performed concurrently with GPU kernel execution of another stream, maximizing the utilization of the available hardware in this way.

3.1. GPU Implementation. The sn-LinK algorithm operates on grid batches of typically 10 000 to 20 000 grid points on GPUs and typically 512 points per batch on CPUs. We found 256 points per AMD compute unit (CU) or NVIDIA streaming multiprocessor (SM) to be optimal, totaling 15,360 points for the AMD Radeon VII GPU and 20,480 points for the NVIDIA GV100. In contrast to our previous implementation,[44] we adapted our Hilbert curve-based sub-batching scheme (see Section 3.3 of ref 44) to also generate the large grid batches. The main advantage of the new approach is the fixed size of grid points per batch, that is, every grid batch except the very last one contains exactly the same amount of points, which ensures optimal utilization of the parallel compute capabilities of the GPUs.

Algorithm 1 sn-LinK GPU implementation

```
for all batches b do                                          ▷ openMP parallel
    perform pre-screening (Section 2.4)                       ▷ CPU
end for
Allocate GPU memory
for all batches b do                  ▷ openMP/multi-GPU/multi-stream parallel
    copy grid data and significant shell-data to GPU
    compute basis function values χ_μ(r_g)                    ▷ GPU
    get batch-local density matrix P                          ▷ CPU
    copy batch-local P to GPU
    evaluate eq. (5) (BLAS-3)                                 ▷ GPU
    collect data for all shellpairs significant by eqs. (13) and (14)  ▷ CPU
    copy shell-pair-data to GPU
    evaluate eq. (6) (integral evaluation)                    ▷ GPU
    evaluate eq. (7) (BLAS-3)                                 ▷ GPU
    copy batch-local exchange matrix K to CPU
    add batch-local K to global K                             ▷ CPU
end for
```

In our GPU implementation (see Algorithm 1), we primarily parallelize over these large grid batches, employing multiple parallel host threads, each of which maps to one device stream, which we implemented using CUDA streams and OpenCL command queues. For maximum performance, we found two or three parallel streams per device to be optimal (see also Section 5.2.3), allowing for concurrent CPU execution, GPU execution, and CPU-GPU data transfer, maximizing hardware utilization in this way. This strategy requires the pre-allocation of GPU memory since allocation of device memory forces stream synchronization.

For a small system (up to ∼200 atoms), the evaluation of the 3c-1e integrals in eq 6 is by far the slowest step, amounting to over 90% of the computation time, whereas for larger systems, the matrix multiplications of eqs 5 and 7 become comparatively more expensive, for example, for the system over 1000 atoms, the integral evaluation amounts to less than 50% of the total computation time. In contrast to the Intel Xeon Phi implementation presented in ref 43, we therefore decided to implement all four compute-intensive steps, that is, the computation of the basis functions $\chi_\mu(\mathbf{r})$ and the evaluation of eqs 5−7 on the GPU, thereby also reducing the amount of CPU-GPU memory transfer.

To achieve asymptotic linear scaling of the implementation while still utilizing the high performance of dense matrix algebra routines provided by basic linear algebra subroutines (BLAS-3) libraries (i.e., Intel MKL for CPUs, cuBLAS for NVIDIA GPUs, and clBLAS for AMD GPUs), we employ dense batch-local submatrices of asymptotically constant size for $\boldsymbol{P}$ and $\boldsymbol{K}$, containing only entries for the significant basis functions within the current batch, determined by the preselection algorithm outlined in Section 2.4, thereby also guaranteeing asymptotically constant GPU memory requirements.

3.2. Implementation of the 3c-1e Integrals. The prescreening of Section 2.4 also provides an asymptotically constant-sized set of shell pairs for each batch, which is further refined on the CPU using the integral selection methods described in Sections 2.2 and 2.3. The shell-pair data is then copied to the GPU, where all the significant 3c-1e integrals $A_{\nu\lambda g}$ for the respective batch are subsequently computed and directly multiplied with $F_{\lambda g}$ to form $G_{\nu g}$ according to eq 6 (see Algorithm 2), performing on-the-fly integral screening on the sub-batch level.

The performance of GPUs relies heavily on single-instruction-mutliple-data (SIMD) vector operations, that is, 32 or 64 parallel threads are collected within one "warp" (NVIDIA) or "wavefront" (AMD), respectively. Since branching within a warp necessitates the evaluation of both branches, such warp-level branching has to be avoided for a highly performant code, which is particularly problematic if combined with integral screening. However, our sub-batch implementation of ref 44 provides spatially local sub-batches with exactly the same number of grid points. Therefore, we choose sub-batches of exactly 64 points, which perfectly maps to the warp/wavefront size of current GPUs. We thus perform the tightest level of integral screening (employing eqs 13 and 14) for 64 points at once, thereby minimizing the screening overhead and ensuring identical branching within each warp/wavefront.

Algorithm 2 Evaluation of eq. (6) (integral evaluation)

```
for all sign. shell-pairs νλ do                                  ▷ sequential
    for all sub-batches b̃ of 64 points do                        ▷ multi warp parallel
        if shell-pair νλ is significant by eqs. (13) and (14) for b̃ then
            for all grid points g ∈ b̃ do                         ▷ SIMD-parallel (within warp)
                for all primitive shell-pairs do                 ▷ sequential
                    compute primitive integrals (Boys integrals)
                    perform optimized Obara-Saika recursions → Cartesian integrals
                end for
                multiply Cartesian integrals with F_λg and add onto G_νg
            end for
        end if
    end for
end for
```

For our GPU implementation, we employ the same computer-optimized Obara−Saika[66,67] recursions as for the CPU code, that is, our CUDA, OpenCL, and CPU implementations share the same input file for the integral kernels. This reuse of the 3c-1e integral code simplifies the GPU implementation significantly, an important advantage compared to the analytical 4c-2e integral-based methods (see, e.g., refs7,11), where considerable modifications have to be made to obtain an efficient and performant code. Since the integral kernels are parallelized solely over the grid point within each batch, there is no need for communication

https://dx.doi.org/10.1021/acs.jctc.9b00860
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

between different threads. This also considerably simplifies the GPU implementation because neither shared memory nor explicit synchronization have to be utilized.

The optimized recurrence relations were generated by application of common sub-expression elimination (CSE) (implemented in the SymPy Python package[68]) to the unrolled Obara−Saika recursions for each specific l-quantum number combination. Moreover, we found that for most practical applications, Head-Gordon−Pople-like (HGP)[69] shifts on the contracted level do not provide speedups for 3c-1e integrals in practice since the recursions are only relevant for larger l-quantum number combinations, but basis functions with l-quantum numbers larger than 1 (d functions or higher) rarely contain more than one primitive Gaussian. Therefore, we decided not to perform any HGP-like contracted recursion steps.

To avoid the transformation between pure and Cartesian integrals, we perform the whole sn-LinK algorithm with Cartesian basis functions, that is, we initially transform the density matrix into the Cartesian basis, then perform the whole sn-LinK algorithm in the Cartesian basis, and finally transform the exchange matrix back to the pure basis. Analogously, we also multiply the non-axial normalization factors, which are needed to ensure normalization of the non-axial basis functions (e.g., d_{xy}), onto the initial density matrix and onto the final exchange matrix, thereby avoiding the necessity to multiply these factors within the integral code.

4. COMPUTATIONAL DETAILS

To provide a fair comparison between GPU and CPU codes, all possible optimization options were enabled for both the CPU and GPU integral codes. The CPU kernels are compiled with the Intel C++ compiler (ICPC) version 19.0.1[70] with the "-Ofast", "-march = native", options, to enable autovectorization of our integral code using the AVX2 instruction set extensions. We have also tested GCC[71] and Clang[72] but found that the Intel C++ compiler provides significantly better performance (up to a factor of two) due to better optimization heuristics, more aggressive autovectorization, and more advanced instruction reordering. The 3c-1e CPU integral kernels benefit particularly from these optimization because parallelization over the grid index is well suited for SIMD vectorization, whereas for the 4c-2e integral kernels, vectorization is hindered by the heterogeneity of the shell pairs (i.e., different amounts of primitive Gaussians), the branching associated with LinK,[20] and the need for more local storage.

The CUDA kernels were compiled with NVCC-10.0 (CUDA-10.0)[63] with "-O3" and "-use_fast_math" using GCC-7.1 as the host compiler. The OpenCL kernels were precompiled with amdgpu-pro-19.20[65] employing the "-O3", "-cl-mad-enable", "-cl-finite-math-only", and "-cl-no-signed-zeros" options. The CPU timings are performed on one server node with 2 Intel Xeon Silver 4216 CPUs comprising 32 cores at 2.1 GHz providing a performance of 1.075×10^{12} floating-point operations (FLOPs) per second (1.075 TFLOPs/s). The GPU timings are performed on the NVIDIA-GV100 GPU (8.33 TFLOPs/s) and the Radeon VII (3.36 TFLOPs/s). The geometries of the molecules[73] employed in this work are available online at http://www.cup.lmu.de/pc/ochsenfeld/download/.

Throughout this work, we employ our recently developed grids defined in the appendix of ref 62 and briefly summarized in Table 1. All presented timings are given for one full exchange matrix build employing a converged density matrix and the smaller (SCF) grid of the multigrids defined in Table 1, in this way, representing a typical SCF step without incremental Fock builds. Note that molecular grids typically contain about 10 to 30% less grid points than the atomic grids defined in Table 1 due to the erasure of grid points with zero weights, a consequence of our modification to Becke's molecular partitioning scheme[60] (see also discussion in ref 62).

Table 1. Specification of the Grids Employed in the Present Work Given as "n_{rad}/n_{ang} (Number of Points per C Atom)"[a]

grid	SCF grid	final grid
"gm3"	35/110 (2586)	50/302 (9564)
"gm4"	40/194 (5056)	55/434 (15526)
"gm5"	50/302 (9564)	60/590 (21330)

[a]Within the SCF, a coarser grid (denoted as SCF grid) was employed and a finer grid was used for the final energy calculation (denoted as final grid). Grids have been pruned, that is, less angular points are employed for the inner radial shells of each atom.

The timings of the conventional (4c-2e integral-based) code exclude the preLinK[7] preselection, since in the current version of our FermiONs++ program,[7,9] the two matrix multiplications within the preLinK algorithm are performed on the CPU using dense matrix algebra, adding a significant overhead for large systems. However, the sn-LinK timings comprise every step needed for exchange matrix formation, including the preselection.

For all sn-LinK calculations, we choose the screening thresholds $\vartheta_K = 1.0 \times 10^{-7}$ and $\vartheta_E = 1.0 \times 10^{-10}$ during the SCF and $\vartheta_E = 1.0 \times 10^{-11}$ for final energy calculation. These thresholds provide screening errors smaller than $1nE_h$ per basis function for all tested systems, which is consistent with our default threshold for the analytical 4c-2e integrals (10^{-10}). Although significantly looser thresholds could probably be used for most applications, we wanted to provide a very safe default in terms of numerical stability and encourage the user to fine tune these parameters for the specific system of interest to obtain even better performance than presented here.

5. RESULTS AND DISCUSSION

5.1. Accuracy of the Numerical Integration Grids. We begin the analysis of the sn-LinK method by investigating the errors caused by the numerical integration. In Table 2, we investigate the grid-induced errors in the Hartree−Fock energy and the indirectly induced errors in the MP2 energy, caused by the errors in the converged density matrix and serving as a measure for the accuracy of the density matrix. We employ the G2 test set[74] (atomization energies of small molecules), the S22x5 test set[75] (noncovalently bound small dimers), and the L7 test set[76] (7 noncovalently bound dimers with up to 101 atoms) in combination with the def2-TZVP basis set.[77]

Even for our smallest grid "gm3", all errors are significantly below $1mE_h$ and are therefore considered insignificant compared to typical errors from methods and basis sets. Moreover, these errors rapidly decrease with larger grids, and the "gm5" grid provides numerical accuracy up to a few μE_h. Interestingly, the Hartree−Fock errors agree well with the observation we made in ref 62 about the grid errors of the Perdew−Burke−Ernzerhof (PBE)[78] functional despite the use of a very different energy functional.

If only single-point energies are of interest, "gm3" should be the best choice for maximum efficiency, whereas if energy

https://dx.doi.org/10.1021/acs.jctc.9b00860
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 2. Grid-Induced Errors in the Absolute Hartree−Fock (HF) Energy and the Absolute MP2 Correlation Energy (G2 Test Set) or the Respective Interaction Energies (S22x5 and L7 Test Sets) Referenced to the Analytical (4c-2e Integral-Based) Method Employing the def2-TZVP Basis Set[a]

		HF			MP2		
test set	deviation	gm3	gm4	gm5	gm3	gm4	gm5
G2	MaxD	20.0	7.0	2.0	69.1	12.2	2.1
	MAD	2.3	0.7	0.2	6.0	1.2	0.1
S22 (0.9x)	MaxD	84.5	20.2	5.9	178.2	39.6	5.2
	MAD	18.8	4.5	1.3	31.4	9.6	1.9
S22 (1.0x)	MaxD	47.8	17.0	7.1	176.8	43.2	3.7
	MAD	15.1	3.7	1.3	31.9	9.9	1.5
S22 (1.2x)	MaxD	57.4	13.8	4.9	66.2	46.7	3.4
	MAD	16.6	4.6	1.2	32.5	9.7	1.1
S22 (1.5x)	MaxD	63.0	18.9	4.1	154.6	49.3	4.5
	MAD	12.7	3.8	0.9	34.8	10.5	1.2
S22 (2.0x)	MaxD	80.3	16.9	3.0	117.6	51.0	5.2
	MAD	13.6	3.7	0.8	30.8	10.9	1.2
L7	MaxD	165.3	42.4	21.3	489.9	119.1	25.2
	MAD	20.2	24.3	5.0	147.7	28.6	7.4

[a]The errors in the MP2 energy are only due to the errors in the converged density matrix. The seminumerical integration was only used for the exchange matrix formation within the SCF but not within the MP2 calculation.

derivatives (forces and vibrational frequencies) are investigated, we recommend to default to the finer "gm5" grid for the higher numerical stability. Note that much smaller grids have been recommended in the works of Neese et al.[36] and Friesner and co-workers.[27,28,33,79] However, we advise caution for the use of small grid especially when computing molecular properties and recommend to carefully test the influence of the grid on the specific quantity of interest prior to any application.

In contrast to our method, analytical corrections to the seminumerical exchange matrix are added within the approach of ref 36 and to even greater extent within the approach of refs,[27,28,33,79] that is, some selected 4c-2e integrals are computed analytically to reduce the grid error. In particular, Neese et al.[36] proposed to only employ one-center corrections, that is, all integrals where all four basis functions reside at the same atom are computed analytically. We tested this approach but found no improvement at all since our grids integrate every atom-centered function pair virtually exactly. The grid errors thus arise solely from non atom-centered function pairs, where the two different functions reside on different atoms and are therefore not considered within the one-center corrections.

5.2. Performance Analysis. In the following, the performance of sn-LinK is assessed in terms of asymptotic scaling behavior (Section 5.2.1), floating-point performance (Section 5.2.2), multistream GPU performance (Section 5.2.3), and multi-GPU performance (Section 5.2.4). Finally, we give a comparison with the 4c-2e-based preLinK method[7,20] on both CPUs and GPUs in Section 5.2.5.

5.2.1. Scaling with Respect to the System Size. Although the evaluation of eqs 5 to 7 formally scales as $O(M^3)$ with respect to the system size (more specifically $O(N_{grid}N_{bas}^2)$), exploitation of the locality of the Gaussian basis functions and of the locality of the exchange interaction for systems with nonzero HOMO−LUMO gaps should result in an asymptotic $O(M)$ scaling, if the screening techniques described in Sections 2.2 to 2.4 are employed. That is, sn-LinK is asymptotically linear scaling by construction. For most practical systems, however, the observed scaling lies somewhere between the formal $O(M^3)$ and the asymptotic $O(M)$ scaling.

In Figure 1, we investigate the scaling behavior for linear alkanes, separated into the 3c-1e integral part required for the evaluation of eq 6 and the matrix multiplication (BLAS-3) steps of eqs 5 and 7.

In all cases, almost linear scaling is reached for the largest fragments. Unsurprisingly, with larger and more diffuse basis sets, linear scaling is reached later (i.e., for larger fragments) since the selection schemes of Sections 2.2 and 2.4 exploit the locality of the Gaussian basis functions. Interestingly, the 3c-1e integral part reaches linear scaling faster than the BLAS-3 steps. This is a consequence of different screening techniques employed for these two steps, that is, the preselection scheme of Section 2.4 compared to the integral selection scheme of Section 2.4, where the latter is tighter (individual contributions are overestimated to a lesser extent).

Although linear alkane chains are a valuable model system to analyze the asymptotic scaling behavior, more globular systems are of interest for many practical applications. Therefore, a more detailed efficiency analysis of our sn-LinK method is given for adenine-thymine DNA fragments in Figure 2 and for spherical water clusters in Figure 3.

Here, all the observations discussed above for linear alkanes are still valid. That is, the integrals reach linear scaling faster than the BLAS-3 steps, and the asymptotically linear scaling is reached later for larger basis sets. Indeed, the linear-scaling onset for def2-TZVP is so late that even the largest fragment of $(DNA)_{16}$ still scales quadratically. Such a late onset of linear scaling has also been observed by ref 36. In contrast to our previous work,[44] sn-LinK (present work) selects significant shells and shell pairs according to their contributions to the exchange potential matrix instead of the exchange energy. This results in a later onset of linear scaling but provides better SCF convergence, particularly for larger basis sets.

Moreover, due to the heterogeneity of GPU computing, the total execution time within sn-LinK also contains a considerable amount of noncompute steps, for example, CPU-GPU data transfer and memory management, as illustrated for $(DNA)_{16}$/TZVP in Figure 4. The performance impact of these other steps can, however, be significantly reduced by employing multiple streams per GPU since the different steps do not compete for the same computational resources (see also Section 5.2.3). Moreover, the high cost of these other steps necessitates the use of rather large grid batches and prohibits the use of block-sparse matrix multiplications to accelerate the BLAS-3 steps since the management steps would dominate the computation time otherwise. The larger grid batches also contribute to a later onset of linear scaling within the BLAS-3 steps.

In summary, although sn-LinK scales linearly by construction, perfect $O(M)$ scaling is only archived for the largest systems and smaller basis sets. This is the expected behavior since the selection schemes within sn-LinK exploit the locality of the basis functions and of the electronic structure.

5.2.2. FLOP Utilization of the 3c-1e Integral Kernels. Since the 3c-1e integrals still represent the most time-consuming step in the seminumerical exchange build, we put significant effort into its optimization. In particular, the batch-wise integral screening described in Section 2.2 allows for SIMD parallelization resulting in comparatively high utilization of the

https://dx.doi.org/10.1021/acs.jctc.9b00860
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

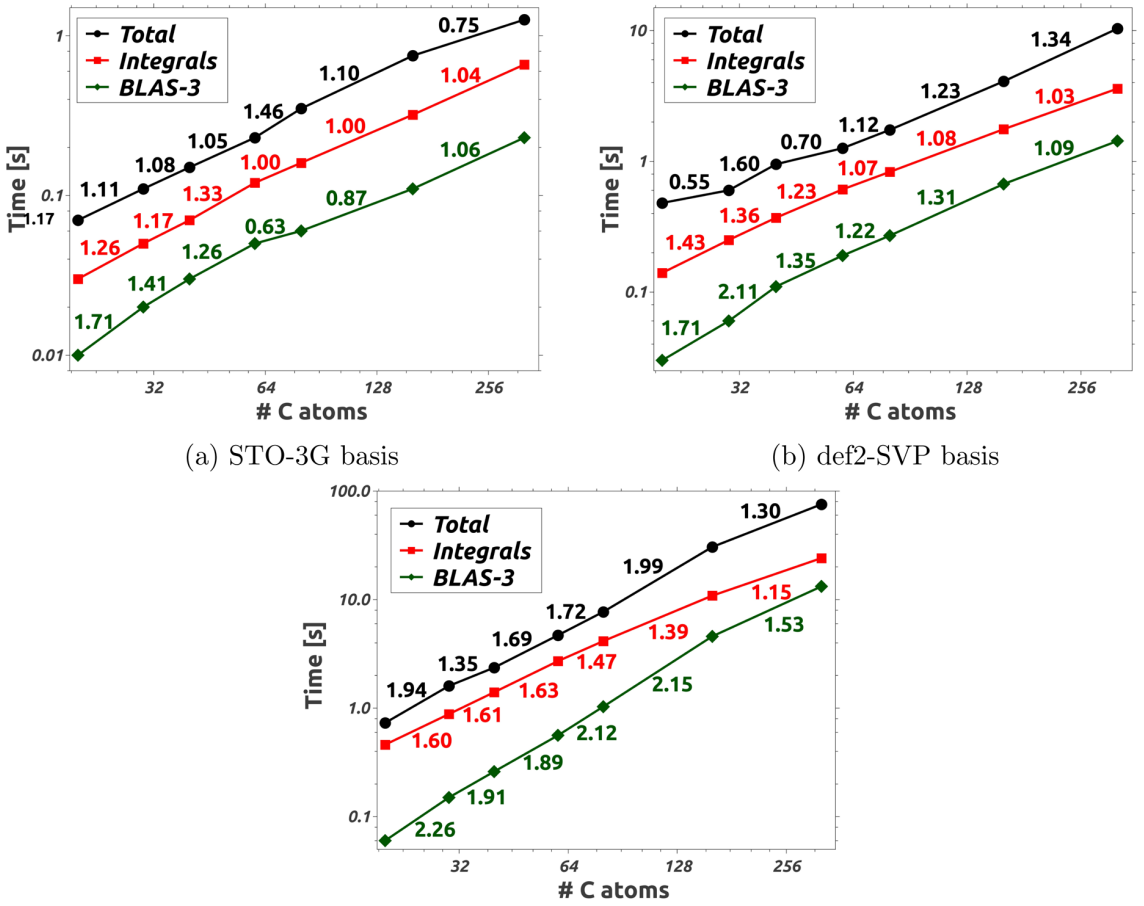


Figure 1. (a−c) Total program execution time and individual execution times for 3c-1e integrals (eq 6) and for the two BLAS-3 steps (eqs 5 and 7) within one exchange build for linear alkanes on one NVIDIA GV100 GPU with one CUDA stream given as a double logarithmic plot. The colored numbers correspond to the scaling with respect to the preceding fragment.

theoretical floating-point performance, as presented in Table 3. An outline how the FLOP counts were obtained is contained within the Supporting Information.

To provide some context, the best dense linear algebra libraries achieve about 70 to 80% FLOP utilization, the GAMESS program package[80,81] was reported[13] to provide 0.51 to 1.4 GFLOPs/s (6 to 17.5% utilization) on CPUs, and GPU implementations for 4c-2e integrals were reported to provide 10 to 30 GFLOPs/s (13 to 39% utilization)[13] in double precision and up to 80 GFLOPs/s (6% utilization)[6] in single precision. Although the theoretical FLOP performance of processors grows exponentially due to ongoing developments in microarchitectures, it becomes increasingly difficult to utilize their full potential since other bottlenecks (cache, memory latency, and bandwidth) dominate in many cases. In this context, the FLOP performance of our 3c-1e integral kernels (230 to 330 GFLOPs/s; 22 to 32% utilization on CPUs and up to 1040 GFLOPs/s; 11 to 16% utilization on GPUs) is very promising.

5.2.3. Multiple Streams on One GPU. In all of the above performance analysis, only one stream per GPU was utilized to time the different steps separately. However, employing more than one stream per GPU should provide some additional speedup since CPU workloads, GPU workloads, and CPU-GPU data transfer allocate different resources and can therefore be performed concurrently. That is, one stream can, for example, transfer data to the GPU, while another stream performs GPU calculations at the same time, thus optimizing the total device utilization (see also discussion in Section 3). The performance gains of this optimization are presented in Table 4.

Compared to the single-streamed evaluation, speedups of up to 50% can be achieved with multiple streams, where the majority of this speedup is already achieved with two streams per GPU. However, the memory use of each GPU scales proportionally with the amount of employed streams, and we therefore decided to employ three streams per GPU as a sensible compromise between performance and GPU memory usage.

5.2.4. Multi-GPU Scaling. Since many high-performance-computing (HPC) servers or workstation are available with up to 16 GPUs per node, the parallel scaling with an increasing amount of GPUs is also of high interest, particularly if employing comparatively inexpensive GPUs like the AMD Radeon VII. We therefore present the multi-GPU scaling of our sn-LinK code in Table 5, activating one, two, or four AMD Radeon VII GPUs.

https://dx.doi.org/10.1021/acs.jctc.9b00860
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

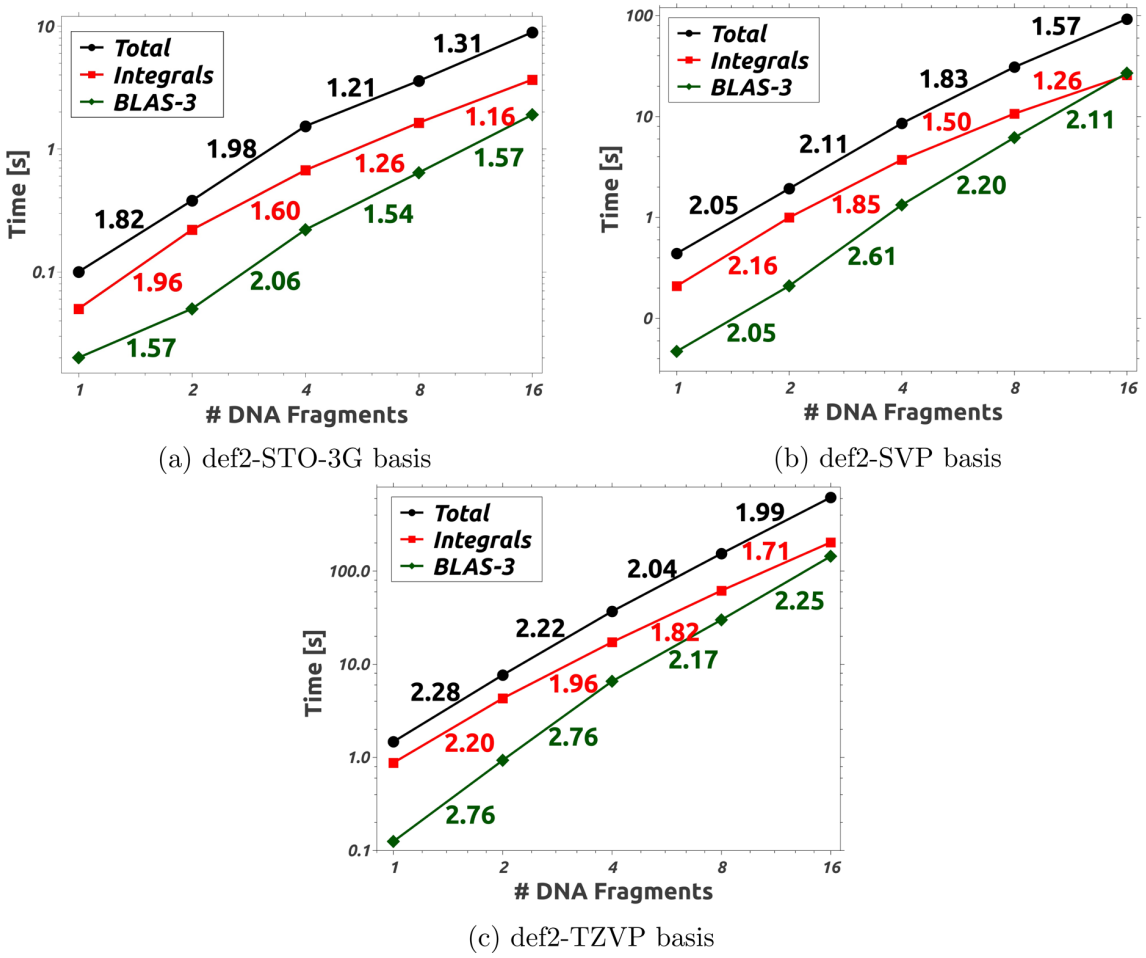


Figure 2. (a–c) Total program execution time and individual execution times for 3c-1e integrals (eq 6) and for the two BLAS-3 steps (eqs. 5 and 7) within one exchange build for adenine-thymine DNA fragments on one NVIDIA GV100 GPU with one CUDA stream given as a double logarithmic plot. The colored numbers correspond to the scaling with respect to the preceding fragment.

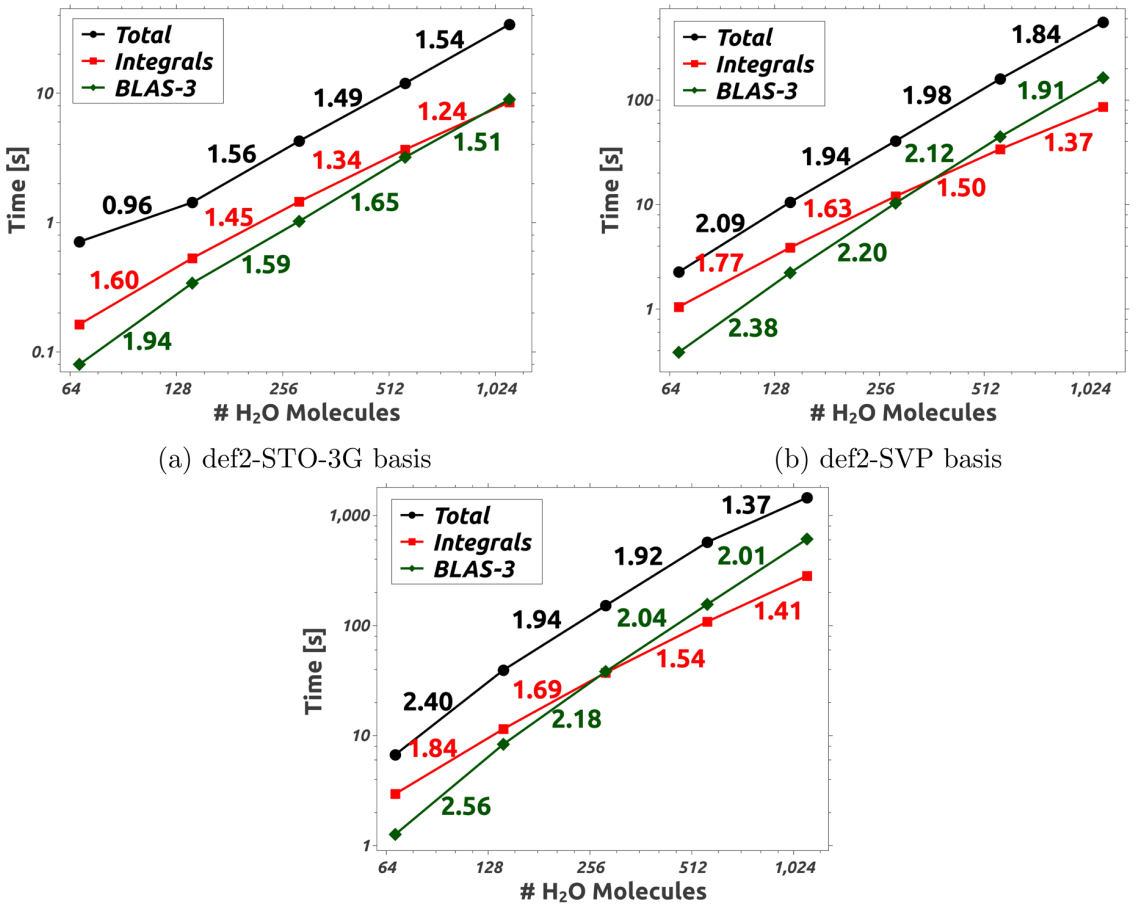
We indeed observe very favorable parallel scaling (3.5× to 4.4× speedups for 4 GPUs), particularly considering that the available CPU resources and memory bandwidth need to be shared between four devices. The over 100% parallel efficiency for $(DNA)_4$/TZVP/"gm5" is a consequence of a particular fortunate workload distribution, that is, all 12 streams finished at very similar times. Such timing fluctuations are typical within such a highly parallel setup because the number of work batches per stream is very small, for example, for $(DNA)_4$/"gm3", there are only 40 grid batches in total, which needs to be split between a total of 12 streams if all four GPUs are utilized.

5.2.5. Comparison with PreLinK. In Table 6, we compare the CPU and GPU performance of the 3c-1e integral-based sn-LinK method of the present work with the analytical (4c-2e integral-based) preLinK method[7] employing 32 CPU cores, 4 Radeon VII GPUs, or 1 NVIDIA GV100 GPU. In this comparison, preLinK typifies all other 4c-2e integral-based methods for Fock exchange, as implemented in most quantum chemistry programs, and allows for a consistent comparison within the same program on both CPUs and GPUs.

The sn-LinK method outperforms the analytical method in most tested applications on CPUs and, to even greater extent, on GPUs. The performance gains from sn-LinK compared to the analytical method are most significant for larger systems and larger basis sets (e.g., factor 17 (14.7 s vs 252 s) for $(DNA)_4$/def2-TZVP/"gm3") due to the superior basis set scaling ($O(N_{bas}^2)$) of sn-LinK compared to preLinK ($O(N_{bas}^4)$). Moreover, the seminumerical code provides better CPU → GPU speedups (up to a factor of 9.5 on four AMD Radeon VII GPUs and a factor of 5.5 on one NVIDIA GV100 GPU) than the analytical code (up to 4.6 on four Radeon VII and 3.8 on one GV100). The better speedups are a direct consequence of the reduced local storage requirements of the 3c-1e integral code compared to the 4c-2e code, resulting in a significantly better utilization of the GPU's floating-point compute units.

In summary, the sn-LinK methods transfer particularly well to GPUs and therefore enable the routine computation of large molecules containing hundreds of atoms and large basis sets. This represents a substantial improvement over existing seminumerical methods, for example, for the fullerene C_{240}/cc-PVTZ, our sn-LinK method is close to 100 times faster than the seminumerical Intel Xeon Phi-based implementation of ref 43 (30.5 s vs 2970 s). In addition, our sn-LinK method allows for routine calculation for hundreds of atoms and augmented quadruple-ζ basis sets, (e.g., one exchange build for $(DNA)_4$/def2-QZVPPD/"gm3" takes only 257 s), which is of particular interest in combination with post-Hartree–Fock correlation

https://dx.doi.org/10.1021/acs.jctc.9b00860
J. Chem. Theory Comput. XXXX, XXX, XXX–XXX



(a) def2-STO-3G basis

(b) def2-SVP basis

(c) def2-TZVP basis

Figure 3. (a–c) Total program execution time and individual execution times for 3c-1e integrals (eq 6) and for the two BLAS-3 steps (eqs. 5 and 7) within one exchange build for spherical water clusters on one NVIDIA GV100 GPU with one CUDA stream given as a double logarithmic plot. The colored numbers correspond to the scaling with respect to the preceding fragment.

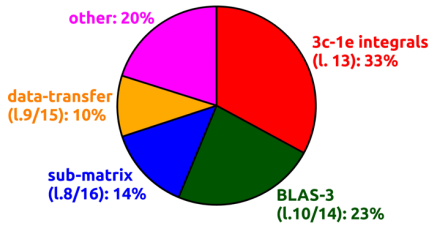


Figure 4. Breakdown of the total execution time of one exchange build for $(DNA)_{16}$/TZVP into different lines of Algorithm 1.

Table 3. Number of Floating-Point Operations Necessary for the Evaluation of All 3c-1e Integrals (with Batch-Wise Integral Selection Activated) for One Exchange Build for $(DNA)_4$ and Floating-Point Performance of the Integral Code Given as GFLOPs/s (Utilization of the Theoretical FLOP Performance in Parentheses)

basis	#GFLOPs	CPU	GV100	R. VII
STO-3G	700	330 (30.7%)	1040 (15.5%)	426 (12.8%)
def2-SVP	2780	239 (22.2%)	750 (11.2%)	429 (12.9%)
def2-TZVP	16,100	234 (21.8%)	932 (13.9%)	470 (14.1%)
def2-QZVPPD	199,000	257 (23.9%)	797 (11.9%)	

methods, which typically require large basis set for accurate results.

Furthermore, calculations employing basis functions with very high angular momentum (e.g., *g*-functions) are very challenging for 4c-2e-based GPU implementations since the high complexity of the high-*l*-quantum number kernels (e.g., our (*gg*|*gg*) kernel contains over 100,000 lines of code) can lead to numerical instabilities of our present GPU code. The extent of this problem depends on the specific GPU in use and has also been reported by other groups.[6] Our sn-LinK method, however, does not suffer from these issues because the 3c-1e integrals are much simpler to evaluate.

6. CONCLUSIONS AND OUTLOOK

Within the present work, we described a new, highly efficient seminumerical exchange method, denoted as sn-LinK, and outlined its implementation for graphic processing units. After validating the accuracy of the numerical integration, we compared the performance of this new method with our conventional (4c-2e integral-based) preLinK method[7] and found outstanding performance improvements, especially for larger basis sets. Moreover, we showed that the sn-LinK algorithm benefits particularly well from GPU acceleration due to the lower local storage requirements of the 3c-1e integral

https://dx.doi.org/10.1021/acs.jctc.9b00860
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 4. Time in Seconds for One Exchange Build Using One to Four Streams on One GPU Employing the "gm3" Grid

		GV100				R. VII			
system	basis streams	1	2	3	4	1	2	3	4
$(DNA)_1$	def2-SVP	0.36	0.29	0.26	0.27	0.85	0.68	0.79	0.98
$(DNA)_1$	def2-TZVP	1.36	1.11	0.98	1.10	3.01	2.73	2.38	2.85
$(DNA)_4$	def2-SVP	7.4	5.5	5.3	5.0	15.1	11.3	10.9	10.5
$(DNA)_4$	def2-TZVP	33.8	27.4	25.4	25.2	64.3	52.9	50.7	49.8

Table 5. Multi-GPU Scaling Employing One, Two, and Four Radeon VII GPUs within One Node[a]

system	basis	grid	1 GPU	2 GPUs	4 GPUs
$(DNA)_4$	def2-SVP	"gm3"	10.9	5.4 (2.0)	3.0 (3.6)
$(DNA)_4$	def2-TZVP	"gm3"	50.7	26.0 (1.9)	14.7 (3.5)
$(DNA)_4$	def2-SVP	"gm5"	27.3	13.9 (1.9)	6.3 (4.4)
$(DNA)_4$	def2-TZVP	"gm5"	143.3	73.2 (2.0)	37.9 (3.8)

[a]Timings are given in seconds for one exchange matrix build, employing a converged density matrix using the smaller (SCF) grid from Table 1. Speedup compared to 1 GPU in parentheses.

kernels compared to the 4c-2e kernels that are required within conventional implementations. Furthermore, we could verify the asymptotic linear-scaling behavior of our implementation for linear alkanes, DNA fragments, and spherical water clusters for small basis sets. For the larger def2-TZVP basis sets, the onset for linear scaling is so late that it was only observed for large linear alkanes and water clusters.

Although the focus of the present work was solely on single-point calculations, seminumerical methods are particularly efficient for computing molecular forces since no integral derivatives need to be evaluated.[29,36] Moreover, the extension of the sn-LinK algorithm to local hybrid functionals is straightforward in principle, however, requiring quite some additional implementation effort to merge the CPU-based DFT code with the GPU-based sn-LinK code. Thus, our, herein, presented sn-LinK algorithm also facilitates future developments of local hybrid functionals, which used to be restrained by their high computational cost. These two extensions are currently under development and will be discussed in future work.

Finally, we want to emphasize the applicability of the seminumerical/pseudospectral method to other molecular properties[31,32,82,83] and post-Hartree−Fock correlation methods[84] as well as the conceptional similarities to the tensor hypercontraction (THC) framework.[85]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.9b00860.

Evaluation of 3c-1e integral FLOP counts (PDF)

Table 6. Timings in Seconds for CPU (32 Cores/64 Threads@2.10 GHz) and GPU Codes Run on either Four AMD Radeon VII (4× R. VII) or One NVIDIA GV100 (GV100) Using sn-LinK (Denoted as "sn-LinK@gm3"/"sn-LinK@gm5") as Compared to the preLinK Method of Ref 7 (Denoted as "preLinK")[a]

system	basis	method	#BFs	CPU	4× R. VII	GV100
$(DNA)_1$	def2-SVP	"sn-LinK@gm3"	660	0.8	0.3 (2.4)	0.4 (2.0)
$(DNA)_1$	def2-SVP	"sn-LinK@gm5"	660	2.5	0.6 (4.2)	0.8 (3.2)
$(DNA)_1$	def2-SVP	"preLinK"	660	1.7	1.0 (1.8)	1.2 (1.5)
$(DNA)_1$	def2-TZVP	"sn-LinK@gm3"	1422	3.6	1.0 (3.7)	1.2 (3.0)
$(DNA)_1$	def2-TZVP	"sn-LinK@gm5"	1422	11.0	2.4 (4.7)	3.1 (3.6)
$(DNA)_1$	def2-TZVP	"preLinK"	1422	29.6	10.4 (2.8)	15.6 (1.9)
$(DNA)_1$	def2-QZVPPD	"sn-LinK@gm3"	3815	30.0	8.1 (3.7)	10.5 (2.9)
$(DNA)_1$	def2-QZVPPD	"sn-LinK@gm5"	3815	99.6	21.5 (4.6)	32.5 (3.1)
$(DNA)_1$	def2-QZVPPD	"preLinK"	3815	2035	–	–
$(DNA)_4$	def2-SVP	"sn-LinK@gm3"	2904	17.6	3.0 (5.9)	5.8 (3.1)
$(DNA)_4$	def2-SVP	"sn-LinK@gm5"	2904	59.3	6.3 (9.5)	14.4 (4.1)
$(DNA)_4$	def2-SVP	"preLinK"	2904	54.2	18.5 (2.9)	23.2 (2.3)
$(DNA)_4$	def2-TZVP	"sn-LinK@gm3"	6336	139.5	14.7 (9.5)	27.2 (5.1)
$(DNA)_4$	def2-TZVP	"sn-LinK@gm5"	6336	316.3	37.9 (8.3)	75.3 (4.2)
$(DNA)_4$	def2-TZVP	"preLinK"	6336	1038.8	252.2 (4.1)	419.9 (2.5)
$(DNA)_4$	def2-QZVPPD	"sn-LinK@gm3"	16,574	1334	257.4 (5.2)	307.8 (4.3)
$(DNA)_4$	def2-QZVPPD	"sn-LinK@gm5"	16,574	5119	814.6 (6.3)	924.2 (5.5)
$(DNA)_4$	def2-QZVPPD	"preLinK"	16,574	101,250	–	–
C_{240}	cc-pVDZ	"sn-LinK@gm3"	3600	49.9	6.6 (7.6)	12.2 (4.1)
C_{240}	cc-pVDZ	"sn-LinK@gm5"	3600	168.8	19.4 (8.7)	38.6 (4.4)
C_{240}	cc-pVDZ	"preLinK"	3600	411.4	162.1 (2.5)	215.6 (1.9)
C_{240}	cc-pVTZ	"sn-LinK@gm3"	8400	207.0	30.5 (6.8)	55.0 (3.8)
C_{240}	cc-pVTZ	"sn-LinK@gm5"	8400	678.6	86.8 (7.8)	170.4 (4.0)
C_{240}	cc-pVTZ	"preLinK"	8400	6294	1365 (4.6)	1660 (3.8)

[a]Timings are given for one exchange matrix build, employing a converged density matrix using the smaller (SCF) grid from Table 1. For context, the number of Cartesian basis functions (#BFs) is given for each system, and the CPU → GPU speedups are given in parentheses.

https://dx.doi.org/10.1021/acs.jctc.9b00860
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

AUTHOR INFORMATION

Corresponding Author

Christian Ochsenfeld − *Department of Chemistry, Chair of Theoretical Chemistry, University of Munich (LMU) D-81377 München, Germany;* orcid.org/0000-0002-4189-6558; Email: christian.ochsenfeld@uni-muenchen.de

Authors

Henryk Laqua − *Department of Chemistry, Chair of Theoretical Chemistry, University of Munich (LMU) D-81377 München, Germany*

Travis H. Thompson − *Department of Chemistry, Chair of Theoretical Chemistry, University of Munich (LMU) D-81377 München, Germany*

Jörg Kussmann − *Department of Chemistry, Chair of Theoretical Chemistry, University of Munich (LMU) D-81377 München, Germany;* orcid.org/0000-0002-4724-8551

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.9b00860

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This project was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) (SFB 1309-325871075). In addition, the authors acknowledge the financial support by the DFG cluster of excellence (EXC2111) "Munich Center for Quantum Science and Technology", MCQST. C.O. further acknowledges financial support as Max-Planck-Fellow at the MPI-FKF Stuttgart. We dedicate this work to Dr. Klaus Römer on the occasion of his 80th birthday in February 2020. We are grateful to him for supporting our Faculty at LMU Munich for many years.

REFERENCES

(1) Ufimtsev, I. S.; Martínez, T. J. Quantum Chemistry on Graphical Processing Units. 1. Strategies for Two-Electron Integral Evaluation. *J. Chem. Theory Comput.* **2008**, *4*, 222−231.

(2) Yasuda, K. Accelerating Density Functional Calculations with Graphics Processing Unit. *J. Chem. Theory Comput.* **2008**, *4*, 1230−1236.

(3) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5*, 2619−2628.

(4) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 2. Direct Self-Consistent-Field Implementation. *J. Chem. Theory Comput.* **2009**, *5*, 1004−1015.

(5) Luehr, N.; Ufimtsev, I. S.; Martínez, T. J. Dynamic Precision for Electron Repulsion Integral Evaluation on Graphical Processing Units (GPUs). *J. Chem. Theory Comput.* **2011**, *7*, 949−954.

(6) Titov, A. V.; Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. Generating Efficient Quantum Chemistry Codes for Novel Architectures. *J. Chem. Theory Comput.* **2013**, *9*, 213−221.

(7) Kussmann, J.; Ochsenfeld, C. Pre-selective screening for matrix elements in linear-scaling exact exchange calculations. *J. Chem. Phys.* **2013**, *138*, 134114.

(8) Maurer, S. A.; Kussmann, J.; Ochsenfeld, C. Communication: A reduced scaling J-engine based reformulation of SOS-MP2 using graphics processing units. *J. Chem. Phys.* **2014**, *141*, No. 051106.

(9) Kussmann, J.; Ochsenfeld, C. Preselective Screening for Linear-Scaling Exact Exchange-Gradient Calculations for Graphics Processing Units and General Strong-Scaling Massively Parallel Calculations. *J. Chem. Theory Comput.* **2015**, *11*, 918−922.

(10) Kussmann, J.; Ochsenfeld, C. Employing OpenCL to Accelerate Ab Initio Calculations on Graphics Processing Units. *J. Chem. Theory Comput.* **2017**, *13*, 2712−2716.

(11) Kussmann, J.; Ochsenfeld, C. Hybrid CPU/GPU Integral Engine for Strong-Scaling Ab Initio Methods. *J. Chem. Theory Comput.* **2017**, *13*, 3153−3159.

(12) Beuerle, M.; Kussmann, J.; Ochsenfeld, C. Screening methods for linear-scaling short-range hybrid calculations on CPU and GPU architectures. *J. Chem. Phys.* **2017**, *146*, 144108.

(13) Asadchev, A.; Allada, V.; Felder, J.; Bode, B. M.; Gordon, M. S.; Windus, T. L. Uncontracted Rys Quadrature Implementation of up to G Functions on Graphical Processing Units. *J. Chem. Theory Comput.* **2010**, *6*, 696−704.

(14) Asadchev, A.; Gordon, M. S. New Multithreaded Hybrid CPU/GPU Approach to Hartree-Fock. *J. Chem. Theory Comput.* **2012**, *8*, 4166−4176.

(15) Miao, Y.; Merz, K. M., Jr. Acceleration of Electron Repulsion Integral Evaluation on Graphics Processing Units via Use of Recurrence Relations. *J. Chem. Theory Comput.* **2013**, *9*, 965−976.

(16) Schwegler, E.; Challacombe, M. Linear scaling computation of the Hartree-Fock exchange matrix. *J. Chem. Phys.* **1996**, *105*, 2726−2734.

(17) Burant, J. C.; Scuseria, G. E.; Frisch, M. J. A linear scaling method for Hartree-Fock exchange calculations of large molecules. *J. Chem. Phys.* **1996**, *105*, 8969−8972.

(18) Challacombe, M.; Schwegler, E. Linear scaling computation of the Fock matrix. *J. Chem. Phys.* **1997**, *106*, 5526−5536.

(19) Schwegler, E.; Challacombe, M.; Head-Gordon, M. Linear scaling computation of the Fock matrix. II. Rigorous bounds on exchange integrals and incremental Fock build. *J. Chem. Phys.* **1997**, *106*, 9708−9717.

(20) Ochsenfeld, C.; White, C. A.; Head-Gordon, M. Linear and sublinear scaling formation of Hartree-Fock-type exchange matrices. *J. Chem. Phys.* **1998**, *109*, 1663−1669.

(21) Ochsenfeld, C. Linear scaling exchange gradients for Hartree-Fock and hybrid density functional theory. *Chem. Phys. Lett.* **2000**, *327*, 216−223.

(22) Früchtl, H. A.; Kendall, R. A.; Harrison, R. J.; Dyall, K. G. An implementation of RI-SCF on parallel computers. *Int. J. Quantum Chem.* **1997**, *64*, 63−69.

(23) Weigend, F. A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285−4291.

(24) Friesner, R. A. Solution of self-consistent field electronic structure equations by a pseudospectral method. *Chem. Phys. Lett.* **1985**, *116*, 39−43.

(25) Friesner, R. A. Solution of the Hartree-Fock equations by a pseudospectral method: application to diatomic molecules. *J. Chem. Phys.* **1986**, *85*, 1462−1468.

(26) Friesner, R. A. Solution of the Hartree-Fock equations for polyatomic molecules by a pseudospectral method. *J. Chem. Phys.* **1987**, *86*, 3522−3531.

(27) Friesner, R. A. An automatic grid generation scheme for pseudospectral self-consistent field calculations on polyatomic molecules. *J. Phys. Chem.* **1988**, *92*, 3091−3096.

(28) Ringnalda, M. N.; Belhadj, M.; Friesner, R. A. Pseudospectral Hartree-Fock theory: applications and algorithmic improvements. *J. Chem. Phys.* **1990**, *93*, 3397−3407.

(29) Won, Y.; Lee, J. G.; Ringnalda, M. N.; Friesner, R. A. Pseudospectral Hartree-Fock gradient calculations. *J. Chem. Phys.* **1991**, *94*, 8152−8157.

(30) Greeley, B. H.; Russo, T. V.; Mainz, D. T.; Friesner, R. A.; Langlois, J.-M.; Goddard, W. A., III; Donnelly, R. E., Jr.; Ringnalda, M. N. New pseudospectral algorithms for electronic structure calculations: length scale separation and analytical two-electron integral corrections. *J. Chem. Phys.* **1994**, *101*, 4028−4041.

(31) Cao, Y.; Friesner, R. A. Molecular (hyper)polarizabilities computed by pseudospectral methods. *J. Chem. Phys.* **2005**, *122*, 104102.

(32) Ko, C.; Malick, D. K.; Braden, D. A.; Friesner, R. A.; Martínez, T. J. Pseudospectral time-dependent density functional theory. *J. Chem. Phys.* **2008**, *128*, 104103.
(33) Bochevarov, A. D.; Harder, E.; Hughes, T. F.; Greenwood, J. R.; Braden, D. A.; Philipp, D. M.; Rinaldo, D.; Halls, M. D.; Zhang, J.; Friesner, R. A. Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *Int. J. Quantum Chem.* **2013**, *113*, 2110−2142.
(34) Zhang, J.; Weisman, A. L.; Saitta, P.; Friesner, R. A. Efficient simulation of large materials clusters using the jaguar quantum chemistry program: Parallelization and wavefunction initialization. *Int. J. Quantum Chem.* **2016**, *116*, 357−368.
(35) Cao, Y.; Hughes, T.; Giesen, D.; Halls, M. D.; Goldberg, A.; Vadicherla, T. R.; Sastry, M.; Patel, B.; Sherman, W.; Weisman, A. L.; Friesner, R. A. Highly efficient implementation of pseudospectral time-dependent density-functional theory for the calculation of excitation energies of large molecules. *J. Comput. Chem.* **2016**, *37*, 1425−1441.
(36) Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Efficient, approximate and parallel Hartree-Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree-Fock exchange. *Chem. Phys.* **2009**, *356*, 98−109.
(37) Plessow, P.; Weigend, F. Seminumerical calculation of the Hartree-Fock exchange matrix: Application to two-component procedures and efficient evaluation of local hybrid density functionals. *J. Comput. Chem.* **2012**, *33*, 810−816.
(38) Bahmann, H.; Kaupp, M. Efficient Self-Consistent Implementation of Local Hybrid Functionals. *J. Chem. Theory Comput.* **2015**, *11*, 1540−1548.
(39) Maier, T. M.; Bahmann, H.; Kaupp, M. Efficient Seminumerical Implementation of Global and Local Hybrid Functionals for Time-Dependent Density Functional Theory. *J. Chem. Theory Comput.* **2015**, *11*, 4226−4237.
(40) Klawohn, S.; Bahmann, H.; Kaupp, M. Implementation of Molecular Gradients for Local Hybrid Density Functionals Using Seminumerical Integration Techniques. *J. Chem. Theory Comput.* **2016**, *12*, 4254−4262.
(41) Liu, F.; Furlani, T.; Kong, J. Optimal Path Search for Recurrence Relation in Cartesian Gaussian Integrals. *J. Phys. Chem. A* **2016**, *120*, 10264−10272.
(42) Liu, F.; Kong, J. Efficient Computation of Exchange Energy Density with Gaussian Basis Functions. *J. Chem. Theory Comput.* **2017**, *13*, 2571−2580.
(43) Liu, F.; Kong, J. An efficient implementation of semi-numerical computation of the Hartree-Fock exchange on the Intel Phi processor. *Chem. Phys. Lett.* **2018**, *703*, 106−111.
(44) Laqua, H.; Kussmann, J.; Ochsenfeld, C. Efficient and Linear-Scaling Seminumerical Method for Local Hybrid Density Functionals. *J. Chem. Theory Comput.* **2018**, *14*, 3451−3458.
(45) Maier, T. M.; Ikabata, Y.; Nakai, H. Efficient Semi-Numerical Implementation of Relativistic Exact Exchange within the Infinite-Order Two-Component Method Using a Modified Chain-of-Spheres Method. *J. Chem. Theory Comput.* **2019**, *15*, 4745−4763.
(46) Grotjahn, R.; Furche, F.; Kaupp, M. Development and Implementation of Excited-State Gradients for Local Hybrid Functionals. *J. Chem. Theory Comput.* **2019**, 5508.
(47) Becke, A. D. A real-space model of nondynamical correlation. *J. Chem. Phys.* **2003**, *119*, 2972−2977.
(48) Becke, A. D. Real-space post-Hartree-Fock correlation models. *J. Chem. Phys.* **2005**, *122*, 064101.
(49) Becke, A. D. Density functionals for static, dynamical, and strong correlation. *J. Chem. Phys.* **2013**, *138*, 074109.
(50) Becke, A. D. Communication: Calibration of a strong-correlation density functional on transition-metal atoms. *J. Chem. Phys.* **2013**, *138*, 161101.
(51) Johnson, E. R. A density functional for strong correlation in atoms. *J. Chem. Phys.* **2013**, *139*, 074110.
(52) Kong, J.; Proynov, E. Density Functional Model for Nondynamic and Strong Correlation. *J. Chem. Theory Comput.* **2016**, *12*, 133−143.
(53) Perdew, J. P.; Staroverov, V. N.; Tao, J.; Scuseria, G. E. Density functional with full exact exchange, balanced nonlocality of correlation, and constraint satisfaction. *Phys. Rev. A* **2008**, *78*, No. 052513.
(54) Jiménez-Hoyos, C. A.; Janesko, B. G.; Scuseria, G. E.; Staroverov, V. N.; Perdew, J. P. Assessment of a density functional with full exact exchange and balanced non-locality of correlation. *Mol. Phys.* **2009**, *107*, 1077−1088.
(55) Bahmann, H.; Rodenberg, A.; Arbuznikov, A. V.; Kaupp, M. A thermochemically competitive local hybrid functional without gradient corrections. *J. Chem. Phys.* **2007**, *126*, 011103−011103/4.
(56) Arbuznikov, A. V.; Bahmann, H.; Kaupp, M. Local Hybrid Functionals with an Explicit Dependence on Spin Polarization. *J. Phys. Chem. A* **2009**, *113*, 11898−11906.
(57) Arbuznikov, A. V.; Kaupp, M. Advances in local hybrid exchange-correlation functionals: from thermochemistry to magnetic-resonance parameters and hyperpolarizabilities. *Int. J. Quantum Chem.* **2011**, *111*, 2625−2638.
(58) Theilacker, K.; Arbuznikov, A. V.; Kaupp, M. Gauge effects in local hybrid functionals evaluated for weak interactions and the GMTKN30 test set. *Mol. Phys.* **2016**, *114*, 1118−1127.
(59) Thompson, T. H.; Ochsenfeld, C. Integral partition bounds for fast and effective screening of general one-, two-, and many-electron integrals. *J. Chem. Phys.* **2019**, *150*, No. 044101.
(60) Becke, A. D. A multicenter numerical integration scheme for polyatomic molecules. *J. Chem. Phys.* **1988**, *88*, 2547−2553.
(61) Treutler, O.; Ahlrichs, R. Efficient molecular numerical integration schemes. *J. Chem. Phys.* **1995**, *102*, 346−354.
(62) Laqua, H.; Kussmann, J.; Ochsenfeld, C. An improved molecular partitioning scheme for numerical quadratures in density functional theory. *J. Chem. Phys.* **2018**, *149*, 204111.
(63) *CUDA Toolkit 10.0*, see https://developer.nvidia.com/cuda-10.0-download-archive.
(64) *OpenCL-2.1*, see: https://www.khronos.org/opencl.
(65) *AMDGPU-Pro Driver 19.20*, see: https://www.amd.com.
(66) Obara, S.; Saika, A. Efficient recursive computation of molecular integrals over Cartesian Gaussian functions. *J. Chem. Phys.* **1986**, *84*, 3963−3974.
(67) Obara, S.; Saika, A. General recurrence formulas for molecular integrals over Cartesian Gaussian functions. *J. Chem. Phys.* **1988**, *89*, 1540−1559.
(68) *SymPy version 1.1.1*, see https://www.sympy.org.
(69) Head-Gordon, M.; Pople, J. A. A method for two-electron Gaussian integral and integral derivative evaluation using recurrence relations. *J. Chem. Phys.* **1988**, *89*, 5777−5786.
(70) *Intel C++ Compiler version 19.0.2.187*, see https://software.intel.com/c-compilers.
(71) *GNU compiler collection*, see http:/gcc.gnu.org.
(72) *Clang C++ Compiler version 4.0.0*, see https://clang.llvm.org.
(73) Maurer, S. A.; Lambrecht, D. S.; Flaig, D.; Ochsenfeld, C. Distance-dependent Schwarz-based integral estimates for two-electron integrals: Reliable tightness vs. rigorous upper bounds. *J. Chem. Phys.* **2012**, *136*, 144107.
(74) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation. *J. Chem. Phys.* **1997**, *106*, 1063−1079.
(75) Gráfová, L.; Pitoňák, M.; Řezáč, J.; Hobza, P. Comparative Study of Selected Wave Function and Density Functional Methods for Noncovalent Interaction Energy Calculations Using the Extended S22 Data Set. *J. Chem. Theory Comput.* **2010**, *6*, 2365−2376.
(76) Sedlak, R.; Janowski, T.; Pitoňák, M.; Řezáč, J.; Pulay, P.; Hobza, P. Accuracy of Quantum Chemical Methods for Large Noncovalent Complexes. *J. Chem. Theory Comput.* **2013**, *9*, 3364−3374.

https://dx.doi.org/10.1021/acs.jctc.9b00860
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(77) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.
(78) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(79) Murphy, R. B.; Cao, Y.; Beachy, M. D.; Ringnalda, M. N.; Friesner, R. A. Efficient pseudospectral methods for density functional calculations. *J. Chem. Phys.* **2000**, *112*, 10131−10141.
(80) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; et al. General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993**, *14*, 1347−1363.
(81) Gordon, M. S.; Schmidt, M. W. Advances in electronic structure theory: GAMESS a decade later. *Theory Appl. Comput. Chem.: First Forty Years* **2005**, 1167−1189.
(82) Friesner, R. A.; Bentley, J. A.; Menou, M.; Leforestier, C. Adiabatic-pseudospectral methods for multidimensional vibrational potentials. *J. Chem. Phys.* **1993**, *99*, 324−335.
(83) Cao, Y.; Beachy, M. D.; Braden, D. A.; Morrill, L.; Ringnalda, M. N.; Friesner, R. A. Nuclear-magnetic-resonance shielding constants calculated by pseudospectral methods. *J. Chem. Phys.* **2005**, *122*, 224116.
(84) Friesner, R. A.; Murphy, R. B.; Beachy, M. D.; Ringnalda, M. N.; Pollard, W. T.; Dunietz, B. D.; Cao, Y. Correlated ab Initio Electronic Structure Calculations for Large Molecules. *J. Phys. Chem. A* **1999**, *103*, 1913−1928.
(85) Parrish, R. M.; Hohenstein, E. G.; Schunck, N. F.; Sherrill, C. D.; Martínez, T. J. Exact tensor hypercontraction: a universal technique for the resolution of matrix elements of local finite-range N-body potentials in many-body quantum problems. *Phys. Rev. Lett.* **2013**, *111*, 132505.

https://dx.doi.org/10.1021/acs.jctc.9b00860
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Chapter 5

Conclusions and Outlook

The integral bounds developed in this thesis are applicable to a wide range of quantum chemical theories and can be used in combination with the best approximations currently available for reducing the computational time scaling of numerical implementations. The integral partition bounds can be used to reduce the complexity of explicitly correlated methods, which offer the best way to get very close to complete basis set results by direct treatment of the short-range cusp behavior of the wave function. Due to the near universal form of correlation cusps, very few parameters are needed to describe them accurately, and instead, the high dimensional integrals that arise are the main computational challenge. At the same time, the very short-ranged nature of the cusps means that the integrals within modern explicitly correlated methods exhibit a level of sparsity that particularly lend them to accelerated calculation through integral screening. This work details several working equations for the F12 correction to the MP2 energy, all of which can be calculated with asymptotic linear scaling cost through the application of the integral partition bounds if the electronic structure of the system is sufficiently local. The most promising formulations are those that use three-dimensional numerical quadrature to factorize two-, three-, and four-electron integrals, since this reduces the formal scaling of the method in addition to drastic reductions in memory requirements. At the same time, such formulations lend themselves to high performance, highly parallel computing architectures such as graphical processing units (GPUs). In this work, the integral partition bounds have been shown to be quite effective for carrying out semi-numerical exact exchange matrix calculations on GPUs.

The bounds and formulas developed here for the MP2-F12 method will be similarly effective for accelerating other explicitly correlated methods. In areas such as the explicitly correlated random phase approximation (RPA) [185–187], explicitly correlated coupled cluster theories [62–71, 188, 189], explicitly correlated multi-reference calculations [190–194], and highly accurate explicitly correlated methods such as Hylleraas-CI [31–35], the same or similar expressions are needed.

The Schwarz-type bound developed for the complex basis function method in non-Hermitian quantum chemistry solves a scaling problem that will become increasingly important as the method matures and its reach extends to molecular resonances of large

systems. The techniques used for the formulation of this bound should also be important when extending other types of integral bounds, such as the integral partition bounds developed in this work, to the complex-symmetric, non-Hermitian integral tensors that arise in complex basis function methods. Such developments would be especially useful for explicitly correlated versions of complex basis function methods, which would have considerable merit considering the large basis sets that are often required [174]. Numerical quadrature combined with screening based on the integral partition bounds could also deliver similarly large performance boosts for complex basis function methods, and this is an opportunity for further research.

Bibliography

[1] E. Schrödinger, *Phys. Rev.* **28** (1926), 1049.

[2] M. Born and R. Oppenheimer, *Ann. Phys.* **389** (1927), 457.

[3] P. Hohenberg and W. Kohn, *Phys. Rev.* **136** (1964), B864.

[4] W. Kohn and L.J. Sham, *Phys. Rev.* **140** (1965), A1133.

[5] C. Hättig, W. Klopper, A. Köhn and D.P. Tew, *Chem. Rev.* **112** (2011), 4.

[6] L. Kong, F.A. Bischoff and E.F. Valeev, *Chem. Rev.* **112** (2011), 75.

[7] W. Kutzelnigg, *Theoret. Chim. Acta* **68** (1985), 445.

[8] W. Kutzelnigg and W. Klopper, *J. Chem. Phys.* **94** (1991), 1985.

[9] A. Szabo and N. Ostlund: *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory.* Dover Books on Chemistry. Dover Publications, 1996.

[10] N. Moiseyev, *Phys. Rep.* **302** (1998), 212.

[11] T. Helgaker, J. Olsen and P. Jørgensen: *Molecular Electronic-Structure Theory.* Wiley-Blackwell, 2013.

[12] C.C.J. Roothaan, *Rev. Mod. Phys.* **23** (1951), 69.

[13] P.O. Löwdin, *J. Chem. Phys.* **18** (1950), 365.

[14] E.F. Valeev, *Chem. Phys. Lett.* **395** (2004), 190.

[15] E.A. Hylleraas, *Z. Phys.* **65** (1930), 209.

[16] D.P. Carroll, H.J. Silverstone and R.M. Metzger, *J. Chem. Phys.* **71** (1979), 4142.

[17] C.F. Bunge, *Theor. Chim. Acta* **16** (1970), 126.

[18] A.W. Weiss, *Phys. Rev.* **122** (1961), 1826.

[19] C. Schwartz, *Phys. Rev.* **126** (1962), 1015.

[20] W. Lakin, *J. Chem. Phys.* **43** (1965), 2954.

[21] T. Kato, *Commun. Pure Appl. Math.* **10** (1957), 151.

[22] R.T. Pack and W.B. Brown, *J. Chem. Phys.* **45** (1966), 556.

[23] W. Kutzelnigg and J.D. Morgan, *J. Chem. Phys.* **96** (1992), 4484.

[24] T. Helgaker, W. Klopper, H. Koch and J. Noga, *J. Chem. Phys.* **106** (1997), 9639.

[25] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen and A.K. Wilson, *Chem. Phys. Lett.* **286** (1998), 243.

[26] W. Klopper, C.C.M. Samson, G. Tarczay and A.G. Császár, *J. Comput. Chem.* **22** (2001), 1306.

[27] T.H. Dunning, *J. Chem. Phys.* **90** (1989).

[28] J.C. Slater, *Phys. Rev.* **31** (1928), 333.

[29] E.A. Hylleraas, *Z. Physik* **54** (1929), 347.

[30] T. Helgaker and W. Klopper, *Theor. Chem. Acc.* **103** (2000), 180.

[31] J.S. Sims and S.A. Hagstrom, *Phys. Rev. A* **4** (1971), 908.

[32] J.S. Sims and S.A. Hagstrom, *J. Phys. B:At. Mol. Opt. Phys.* **37** (2004), 1519.

[33] J.S. Sims and S.A. Hagstrom, *J. Phys. B:At. Mol. Opt. Phys.* **40** (2007), 1575.

[34] J.S. Sims and S.A. Hagstrom, *J. Phys. B:At. Mol. Opt. Phys.* **48** (2015), 175003.

[35] F.E. Harris. In *Novel Electronic Structure Theory: General Innovations and Strongly Correlated Systems*, edited by P. E. Hoggan, volume 76 of *Adv. Quantum Chem.* Academic Press (2018), pages 187 – 210.

[36] D.C. Clary and N.C. Handy, *Phys. Rev. A* **14** (1976), 1607.

[37] S.F. Boys and M.V. Wilkes, *Proc. R. Soc. A* **258** (1960), 402.

[38] K. Singer and M.V. Wilkes, *Proc. R. Soc. A* **258** (1960), 412.

[39] W. Cencek and J. Rychlewski, *J. Chem. Phys.* **98** (1993), 1252.

[40] S. Bubin, M. Pavanello, W.C. Tung, K.L. Sharkey and L. Adamowicz, *Chem. Rev.* **113** (2013), 36.

[41] J. Mitroy, S. Bubin, W. Horiuchi, Y. Suzuki, L. Adamowicz, W. Cencek, K. Szalewicz, J. Komasa, D. Blume and K. Varga, *Rev. Mod. Phys.* **85** (2013), 693.

[42] S.F. Boys and N.C. Handy, *Proc. R. Soc. A* **310** (1969), 43.

[43] S. Ten-no, *Chem. Phys. Lett.* **330** (2000), 169.

[44] O. Hino, Y. Tanimura and S. Ten-no, *J. Chem. Phys.* **115** (2001), 7865.

[45] O. Hino, Y. Tanimura and S. Ten-no, *Chem. Phys. Lett.* **353** (2002), 317.

[46] T. Yanai and T. Shiozaki, *J. Chem. Phys.* **136** (2012), 084107.

[47] S. Ten-no, *J. Chem. Phys.* **121** (2004), 117.

[48] R.A. Bachorz, F.A. Bischoff, A. Glöß, C. Hättig, S. Höfener, W. Klopper and D.P. Tew, *J. Comput. Chem.* **32** (2011), 2492.

[49] A. Preiskorn and B. Żurawski, *Int. J. Quantum Chem.* **27** (1985), 641.

[50] A. Largo-Cabrerizo and E. Clementi, *Int. J. Quantum Chem.* **8** (1987), 1191.

[51] Y. Lu and Z. Huang, *Int. J. Quantum Chem.* **38** (1990), 447.

[52] D. Frye, A. Preiskorn and E. Clementi, *J. Comput. Chem.* **12** (1991), 560.

[53] S. Ten-no, *Chem. Phys. Lett.* **398** (2004), 56.

[54] S. Höfener, F.A. Bischoff, A. Glöß and W. Klopper, *Phys. Chem. Chem. Phys.* **10** (2008), 3390.

[55] S. Ten-no, *Chem. Phys. Lett.* **330** (2000), 175.

[56] A. Komornicki and H.F. King, *J. Chem. Phys.* **134** (2011), 244115.

[57] S. Reine, T. Helgaker and R. Lindh, *Wiley Interdiscip. Rev.-Comput. Mol. Sci.* **2** (2012), 290.

[58] G.M.J. Barca, P.F. Loos and P.M.W. Gill, *J. Chem. Theory Comput.* **12** (2016), 1735.

[59] G.M.J. Barca and P.F. Loos, *J. Chem. Phys.* **147** (2017), 024103.

[60] G.M. Barca and P.F. Loos. In *Novel Electronic Structure Theory: General Innovations and Strongly Correlated Systems*, edited by P. E. Hoggan, volume 76 of *Advances in Quantum Chemistry*. Academic Press (2018), pages 147–165.

[61] G.M.J. Barca: *Single-Determinant Theory of Electronic Excited States and Many-Electron Integrals for Explicitly Correlated Methods*. Research School of Chemistry, The Australian National University, PhD Thesis, 2017.

[62] J. Noga and W. Kutzelnigg, *J. Chem. Phys.* **101** (1994), 7738.

[63] T. Shiozaki, M. Kamiya, S. Hirata and E.F. Valeev, *J. Chem. Phys.* **129** (2008), 071101.

[64] T. Shiozaki, M. Kamiya, S. Hirata and E.F. Valeev, *Phys. Chem. Chem. Phys.* **10** (2008), 3358.

[65] D.P. Tew, W. Klopper and C. Hättig, *Chem. Phys. Lett.* **452** (2008), 326 .

[66] H. Fliegl, W. Klopper and C. Hättig, *J. Chem. Phys.* **122** (2005), 084107.

[67] H. Fliegl, C. Hättig and W. Klopper, *Int. J. Quantum Chem.* **106** (2006), 2306.

[68] D.P. Tew, W. Klopper, C. Neiss and C. Hättig, *Phys. Chem. Chem. Phys.* **9** (2007), 1921.

[69] T.B. Adler, G. Knizia and H.J. Werner, *J. Chem. Phys.* **127** (2007), 221106.

[70] G. Knizia, T.B. Adler and H.J. Werner, *J. Chem. Phys.* **130** (2009), 054104.

[71] E.F. Valeev, *Phys. Chem. Chem. Phys.* **10** (2008), 106.

[72] J.L. Whitten, *J. Chem. Phys.* **58** (1973), 4496.

[73] E. Baerends, D. Ellis and P. Ros, *Chem. Phys.* **2** (1973), 41.

[74] B.I. Dunlap, J.W.D. Connolly and J.R. Sabin, *J. Chem. Phys.* **71** (1979), 4993.

[75] C. Van Alsenoy, *J. Comput. Chem.* **9** (1988), 620.

[76] O. Vahtras, J. Almlöf and M.W. Feyereisen, *Chem. Phys. Lett.* **213** (1993), 514.

[77] H.F. Schurkus, A. Luenser and C. Ochsenfeld, *J. Chem. Phys.* **146** (2017), 211106.

[78] H.J. Werner, F.R. Manby and P.J. Knowles, *J. Chem. Phys.* **118** (2003), 8149.

[79] H.F. Schurkus and C. Ochsenfeld, *J. Chem. Phys.* **144** (2016), 031101.

[80] A. Luenser, H.F. Schurkus and C. Ochsenfeld, *J. Chem. Theory Comput.* **13** (2017), 1647.

[81] A.D. Becke, *J. Chem. Phys.* **88** (1988), 2547.

[82] C.W. Murray, N.C. Handy and G.J. Laming, *Mol. Phys.* **78** (1993), 997.

[83] O. Treutler and R. Ahlrichs, *J. Chem. Phys.* (1995), 346.

[84] R. Stratmann, G.E. Scuseria and M.J. Frisch, *Chem. Phys. Lett.* **257** (1996), 213.

[85] P.M.W. Gill and S.H. Chien, *J. Comput. Chem.* **24** (2003), 732.

[86] H. Laqua, J. Kussmann and C. Ochsenfeld, *J. Chem. Phys.* **149** (2018), 204111.

[87] S.F. Boys, *Proc. R. Soc. London Ser. A* **200** (1950), 542.

[88] H. Taketa, S. Huzinaga and K. O-ohata, *J. Phys. Soc. Jpn.* **21** (1966), 2313.

[89] L.E. McMurchie and E.R. Davidson, *J. Comput. Phys.* **26** (1978), 218.

[90] J.A. Pople and W.J. Hehre, *J. Comput. Phys.* **27** (1978), 161.

[91] H.F. King and M. Dupuis, *J. Comput. Phys.* **21** (1976), 144.

[92] M. Dupuis, J. Rys and H.F. King, *J. Chem. Phys.* **65** (1976), 111.

[93] J. Rys, M. Dupuis and H.F. King, *J. Comput. Chem.* **4** (1983), 154.

[94] D. Hegarty and G. van der Velde, *Int. J. Quantum Chem.* **23** (1983), 1135.

[95] S. Obara and A. Saika, *J. Chem. Phys.* **84** (1986), 3963.

[96] M. Head-Gordon and J.A. Pople, *J. Chem. Phys.* **89** (1988), 5777.

[97] P.M.W. Gill, M. Head-Gordon and J.A. Pople, *Int. J. Quantum Chem.* **36** (1989), 269.

[98] P.M.W. Gill and J.A. Pople, *Int. J. Quantum Chem.* **40** (1991), 753.

[99] P.M.W. Gill, B.G. Johnson and J.A. Pople, *Int. J. Quantum Chem.* **40** (1991), 745.

[100] K. Ishimura and S. Nagase, *Theor. Chem. Acc.* **120** (2008), 185.

[101] A. Asadchev, V. Allada, J. Felder, B.M. Bode, M.S. Gordon and T.L. Windus, *J. Chem. Theory Comput.* **6** (2010), 696.

[102] J. Almlöf, K. Faegri Jr. and K. Korsell, *J. Comput. Chem.* **3** (1982), 385.

[103] H.P. Lüthi and J. Almlöf, *Chem. Phys. Lett.* **135** (1987), 357.

[104] D. Cremer and J. Gauss, *J. Comput. Chem.* **7** (1986), 274.

[105] M. Häser and R. Ahlrichs, *J. Comput. Chem.* **10** (1989), 104.

[106] L. Hörmander: *The Analysis of Linear Partial Differential Operators. I.* Springer-Verlag, Berlin, 2003.

[107] S.A. Maurer, D.S. Lambrecht, J. Kussmann and C. Ochsenfeld, *J. Chem. Phys.* **138** (2013), 014101.

[108] D.S. Hollman, H.F. Schaefer and E.F. Valeev, *J. Chem. Phys.* **142** (2015), 154106.

[109] D.S. Lambrecht and C. Ochsenfeld, *J. Chem. Phys.* **123** (2005), 184101.

[110] D.S. Lambrecht, B. Doser and C. Ochsenfeld, *J. Chem. Phys.* **123** (2005), 184102.

[111] S.A. Maurer, D.S. Lambrecht, D. Flaig and C. Ochsenfeld, *J. Chem. Phys.* **136** (2012), 144107.

[112] C.A. White, B.G. Johnson, P.M. Gill and M. Head-Gordon, *Chem. Phys. Lett.* **230** (1994), 8.

[113] C.A. White, B.G. Johnson, P.M. Gill and M. Head-Gordon, *Chem. Phys. Lett.* **253** (1996), 268.

[114] M. Beuerle, J. Kussmann and C. Ochsenfeld, *J. Chem. Phys.* **146** (2017), 144108.

[115] F. Aquilante, T.B. Pedersen, A. Sánchez de Merás and H. Koch, *J. Chem. Phys.* **125** (2006), 174101.

[116] J.M. Millam and G.E. Scuseria, *J. Chem. Phys.* **106** (1997), 5569.

[117] S. Schweizer, J. Kussmann, B. Doser and C. Ochsenfeld, *J. Comput. Chem.* **29** (2008), 1004.

[118] R.A. Friesner, *Chem. Phys. Lett.* **116** (1985), 39.

[119] R.A. Friesner, *J. Chem. Phys.* **85** (1986), 1462.

[120] R.A. Friesner, *J. Chem. Phys.* **86** (1987), 3522.

[121] F. Neese, F. Wennmohs, A. Hansen and U. Becker, *Chem. Phys.* **356** (2009), 98.

[122] P. Plessow and F. Weigend, *J. Comput. Chem.* **33** (2012), 810.

[123] H. Bahmann and M. Kaupp, *J. Chem. Theory Comput.* **11** (2015), 1540.

[124] T.M. Maier, H. Bahmann and M. Kaupp, *J. Chem. Theory Comput.* **11** (2015), 4226.

[125] S. Klawohn, H. Bahmann and M. Kaupp, *J. Chem. Theory Comput.* **12** (2016), 4254.

[126] F. Liu and J. Kong, *J. Chem. Theory Comput.* **13** (2017), 2571.

[127] H. Laqua, J. Kussmann and C. Ochsenfeld, *J. Chem. Theory Comput.* **14** (2018), 3451.

[128] C. Ochsenfeld, C.A. White and M. Head-Gordon, *J. Chem. Phys.* **109** (1998), 1663.

[129] C. Ochsenfeld, *Chem. Phys. Lett.* **327** (2000), 216.

[130] J. Kussmann and C. Ochsenfeld, *J. Chem. Phys.* **138** (2013), 134114.

[131] W. Klopper and C.C.M. Samson, *J. Chem. Phys.* **116** (2002), 6397.

[132] S. Kedžuch, M. Milko and J. Noga, *Int. J. Quantum Chem.* **105** (2005), 929.

[133] H.J. Werner, T.B. Adler and F.R. Manby, *J. Chem. Phys.* **126** (2007), 164102.

[134] J. Almlöf, *Chem. Phys. Lett.* **181** (1991), 319.

[135] M. Häser and J. Almlöf, *J. Chem. Phys.* **96** (1992), 489.

[136] M. Häser, *Theoret. Chim. Acta* **87** (1993), 147.

[137] S.A. Maurer, L. Clin and C. Ochsenfeld, *J. Chem. Phys.* **140** (2014), 224112.

[138] D. Bokhan, S. Bernadotte and S. Ten-no, *Chem. Phys. Lett.* **469** (2009), 214.

[139] T.C. Jagau, K.B. Bravaya and A.I. Krylov, *Annu. Rev. Phys. Chem.* **68** (2017), 525.

[140] E.J. Heller, *Acc. Chem. Res.* **14** (1981), 368.

[141] D. Neuhauser, *J. Chem. Phys.* **95** (1991), 4927.

[142] F. Grossmann, *Chem. Phy. Lett.* **262** (1996), 470.

[143] A.U. Hazi and H.S. Taylor, *Phys. Rev. A* **1** (1970), 1109.

[144] V.A. Mandelshtam, T.R. Ravuri and H.S. Taylor, *J. Chem. Phys.* **101** (1994), 8792.

[145] V.A. Mandelshtam and H.S. Taylor, *J. Chem. Phys.* **102** (1995), 7390.

[146] P.W. Langhoff. *Stieltjes-Tchebycheff Moment-Theory Approach to Molecular Photoionization Studies*. Springer US, Boston, MA, 1979. pages 183–224.

[147] L.S. Cederbaum, W. Domcke and J. Schirmer, *Phys. Rev. A* **22** (1980), 206.

[148] P. Norman and A. Dreuw, *Chem. Rev.* **118** (2018), 7208.

[149] I. Shimamura. In *Advances in Quantum Chemistry*, edited by C. A. Nicolaides, E. Brändas and J. R. Sabin, volume 63 of *Advances in Quantum Chemistry*. Academic Press (2012), pages 165–245.

[150] U.V. Riss and H.D. Meyer, *J. Phys. B* **26** (1993), 4503.

[151] R. Santra and L.S. Cederbaum, *Phys. Rep.* **368** (2002), 1.

[152] J. Muga, J. Palao, B. Navarro and I. Egusquiza, *Phys. Rep.* **395** (2004), 357.

[153] N. Moiseyev and J.O. Hirschfelder, *J. Chem. Phys.* **88** (1988), 1063.

[154] P.O. Löwdin. volume 19 of *Adv. Quantum Chem.* Academic Press (1988), pages 87–138.

[155] K.B. Bravaya, D. Zuev, E. Epifanovsky and A.I. Krylov, *J. Chem. Phys.* **138** (2013), 124106.

[156] E. Balslev and J.M. Combes, *Commun. Math. Phys.* **22** (1971), 280.

[157] J. Aguilar and J.M. Combes, *Commun. Math. Phys.* **22** (1971), 269.

[158] B. Simon, *Commun. Math. Phys.* **27** (1972), 1.

[159] I.W. Herbst and B. Simon, *Phys. Rev. Lett.* **41** (1978), 67.

[160] I.W. Herbst, *Commun. Math. Phys.* **64** (1979), 279.

[161] I.W. Herbst and B. Simon, *Commun. Math. Phys.* **80** (1981), 181.

[162] B. Simon, *Phy. Lett. A* **71** (1979), 211.

[163] C.W. McCurdy and T.N. Rescigno, *Phys. Rev. Lett.* **41** (1978), 1364.

[164] T.N. Rescigno, A.E. Orel and C.W. McCurdy, *J. Chem. Phys.* **73** (1980), 6347.

[165] C.W. McCurdy, T.N. Rescigno, E.R. Davidson and J.G. Lauderdale, *J. Chem. Phys.* **73** (1980), 3268.

[166] M. Mishra, Y. Öhrn and P. Froelich, *Phy. Lett. A* **84** (1981), 4.

[167] C.W. McCurdy and R.C. Mowrey, *Phys. Rev. A* **25** (1982), 2529.

[168] J.G. Lauderdale, C.W. McCurdy and A.U. Hazi, *J. Chem. Phys.* **79** (1983), 2200.

[169] M. Honigmann, R.J. Buenker and H.P. Liebermann, *J. Chem. Phys.* **125** (2006), 234304.

[170] M. Honigmann, R.J. Buenker and H.P. Liebermann, *J. Chem. Phys.* **131** (2009), 034303.

[171] A.F. White, C.W. McCurdy and M. Head-Gordon, *J. Chem. Phys.* **143** (2015), 074103.

[172] A.F. White, M. Head-Gordon and C.W. McCurdy, *J. Chem. Phys.* **142** (2015), 054103.

[173] A.F. White, E. Epifanovsky, C.W. McCurdy and M. Head-Gordon, *J. Chem. Phys.* **146** (2017), 234107.

[174] T.C. Jagau, *J. Chem. Phys.* **145** (2016), 204115.

[175] T.C. Jagau, *J. Chem. Phys.* **148** (2018), 204102.

[176] M. Hernández Vera and T.C. Jagau, *J. Chem. Phys.* **151** (2019), 111101.

[177] B. Boudaïffa, P. Cloutier, D. Hunting, M.A. Huels and L. Sanche, *Science* **287** (2000), 1658.

[178] J. Simons, *Acc. Chem. Res.* **39** (2006), 772.

[179] J.D. Gorfinkiel and S. Ptasinska, *J. Phys. B: At. Mol. Opt. Phys.* **50** (2017), 182001.

[180] J. Kohanoff, M. McAllister, G.A. Tribello and B. Gu, *J. Phys.: Condens. Matter* **29** (2017), 383001.

[181] P. Christopher, H. Xin, A. Marimuthu and S. Linic, *Nat. Mater.* **11** (2012), 1044.

[182] S. Mukherjee, F. Libisch, N. Large, O. Neumann, L.V. Brown, J. Cheng, J.B. Lassiter, E.A. Carter, P. Nordlander and N.J. Halas, *Nano Lett.* **13** (2013), 240.

[183] I. Newton, T. Leseur and F. Jacquier: *Philosophiae naturalis principia mathematica.* Number v. 1-2 in Philosophiae naturalis principia mathematica. G. Brookman, 1833.

[184] E.H. Lieb and R. Seiringer: *The Stability of Matter in Quantum Mechanics.* Cambridge University Press, 2009.

[185] A.S. Hehn and W. Klopper, *J. Chem. Phys.* **138** (2013), 181104.

[186] A.S. Hehn, D.P. Tew and W. Klopper, *J. Chem. Phys.* **142** (2015), 194106.

[187] A.S. Hehn, C. Holzer and W. Klopper, *Chem. Phys.* **479** (2016), 160.

[188] A. Köhn, *J. Chem. Phys.* **130** (2009), 131101.

[189] A. Köhn, *J. Chem. Phys.* **133** (2010), 174118.

[190] R.J. Gdanitz, *Chem. Phys. Lett.* **210** (1993), 253.

[191] R.J. Gdanitz, *Chem. Phys. Lett.* **283** (1998), 253.

[192] S. Ten-no, *Chem. Phys. Lett.* **447** (2007), 175.

[193] T. Shiozaki and H.J. Werner, *J. Chem. Phys.* **133** (2010), 141103.

[194] T. Shiozaki, G. Knizia and H.J. Werner, *J. Chem. Phys.* **134** (2011), 034113.