

---

## CHAPTER 1

# Linear-Scaling Methods in Quantum Chemistry

Christian Ochsenfeld, Jörg Kussmann, and  
Daniel S. Lambrecht

*Institut für Physikalische und Theoretische Chemie, Universität  
Tübingen, D-72076 Tübingen, Germany*

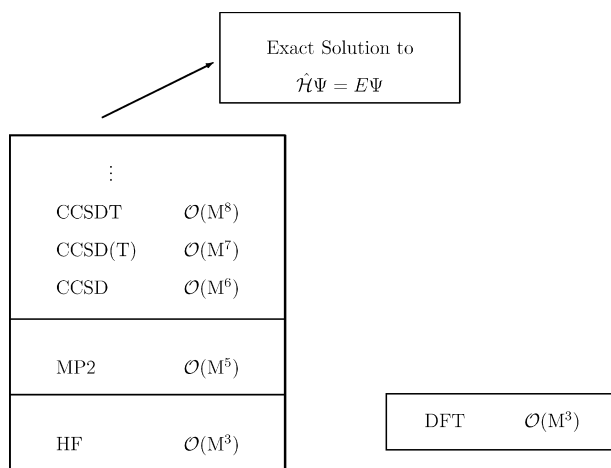
---

---

### INTRODUCTION

With the introduction of the Schrödinger equation in 1926,<sup>1</sup> it was in principle clear how to describe a molecular system and its properties exactly in a nonrelativistic sense. However, for most molecular systems of chemical interest, the analytic solution of the Schrödinger equation is not possible. Therefore, since 1926, a multitude of hierarchical approximations (some of which are displayed in Figure 1) have been devised that allow for a systematic approach to the exact solution of the Schrödinger equation. Although the Schrödinger equation as the fundamental equation in electronic structure theory is already quite old, the field of quantum chemistry is still fairly young and fast moving, and much can be expected in the future for developing and applying quantum chemical methods for the treatment of molecular systems.

The importance of the systematic hierarchy for solving the Schrödinger equation cannot be overemphasized, because it allows one, in principle, to systematically approach the exact result for a molecular property of interest. The simplest approach in this hierarchy is the Hartree–Fock (HF) method, which describes electron–electron interactions within a mean-field approach.<sup>2–4</sup> The electron–correlation effects neglected in this approach can be described by the so-called

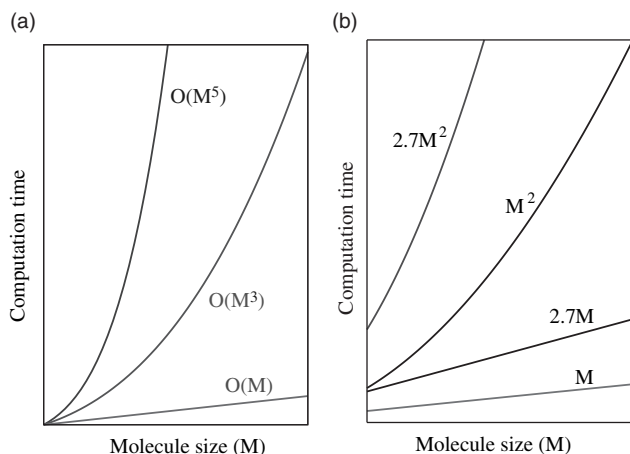


**Figure 1** The hierarchy of ab initio methods: A selection of common approximations for solving the electronic Schrödinger equation is displayed. In addition, the asymptotic scaling order ( $\mathcal{O}()$ ) with respect to molecular size  $M$  is listed.

post-HF methods, with prominent examples such as perturbation theory (e.g., MP2: Møller–Plesset second-order perturbation theory<sup>5</sup>) or the coupled-cluster (CC) expansion (see e.g. Ref. 6 for a review; CCSD: CC singles doubles; CCSD(T): CCSD with perturbative triples; or CCSDT: CC singles doubles triples). In this way, the hierarchy of ab initio methods allows for reliable “measurements” and for estimating the error bars of simpler approximations. In Figure 1, we also list density functional theory (DFT),<sup>7–9</sup> although it does not provide (at least in its current form) a systematic way of improving upon the result. Despite this deficiency in its current form, DFT has pragmatically proven to be highly useful for the description of many molecular systems, while offering a good compromise between accuracy and computational cost. Therefore, DFT has become a standard tool of modern quantum chemistry.

The main difficulty associated with the hierarchy of quantum chemical methods is the strong increase of the computational effort with molecular size ( $M$ ) (compare Figures 1 and 2), especially when approaching the exact solution. Even the simplest approach, the HF method,<sup>4</sup> scales conventionally as  $\mathcal{O}(M^3)$ , where  $\mathcal{O}()$  denotes the order of the asymptotic scaling behavior. This means that when choosing another molecule to study that is 10 times larger than the current molecule, the computational effort is increased by a factor of 1000. The increase becomes even more dramatic if the electron correlation effects neglected in the HF approach are either accounted for by, e.g., MP2 or CCSD, for which the scaling behavior is  $\mathcal{O}(M^5)$  or  $\mathcal{O}(M^6)$ , respectively. The  $\mathcal{O}(M^6)$  scaling entails an increase of the computational effort by a factor of 1, 000, 000 for a 10-fold larger system.

At this stage it is worthwhile to spend some time to clarify the scaling behavior. The focus of this chapter is on methods whose efforts increase only linearly



**Figure 2** The computation time behaves approximately as: computation time =  $a \cdot M^n$ . Here,  $M^n$  is called the scaling behavior, and  $a$  is the prefactor. The graphs provide a schematical comparison of computation times for (a) different scaling behaviors and (b) different prefactors ( $a = 1$  and  $a = 2.7$ ).

with molecular size,  $M$  (defined by, e.g., the number of atoms), while the atom-centered basis set framework is retained. Within the same atomic-orbital (AO) basis, the total number of basis functions ( $N$ ) scales similarly with the molecular size, so that the scaling behavior can be described as well by the number of AOs. However, increasing the number of basis functions for a specific molecule would typically not lead to a linear-scaling behavior. The size of the atom-centered basis simply defines the prefactor of the calculation (i.e., the constant factor with which the scaling behavior is multiplied; see Figure 2). In the current tutorial, we therefore mainly employ the molecule size  $M$  for describing the scaling property.

To illustrate how prohibitive even an  $\mathcal{O}(M^3)$  scaling would be for the calculation of large molecules, we can think about “Moore’s law.”<sup>10</sup> It is an empirical observation proposed in 1965, leading to the statement that computer speed roughly doubles every 1.5 years and that has been, as a rule of thumb, astonishingly valid over the last decades. The factor of 1000 for a 10-fold larger molecule can be described as roughly  $2^{10}$ , which would—with Moore’s assumption—correspond to 15 years of computer development required, whereas an  $\mathcal{O}(M^6)$  scaling would lead to even 30 years. In other words, one would need to wait 15 years for the computers to evolve to perform an HF calculation for a 10-fold larger molecule within the same time frame. This is clearly not an option for any enthusiastic researcher attempting to grasp deeper insights into molecular processes in chemistry, biochemistry, or even biology.

Therefore, the aim of this didactical review is to provide some insights into reducing the scaling behavior of quantum chemical methods so that they scale linearly with molecular size. In this way, any increase in computer speed translates directly into an increase of the treatable molecular size with

respect to time requirements. The focus of this review is on presenting some basic ideas of these linear-scaling methods, without giving a complete overview of the many different approaches introduced in the literature. For basic aspects of quantum mechanics and quantum chemical methods, the reader is referred to the textbook literature such as, e.g., Refs. 4 and 11–13.

In this chapter, we describe mainly linear-scaling self-consistent field (SCF) methods such as HF and DFT, which are closely related in the way energies, energy gradients, and molecular properties are computed. With these linear-scaling methods, molecular systems with more than 1000 atoms can nowadays be computed on simple one-processor workstations. In addition, we provide a brief outlook concerning electron-correlation methods and what might be expected in the future for reducing their scaling behavior while preserving rigorous error bounds. The review is structured as follows:

- After a brief introduction to some basics of SCF theories, we describe in the following four sections how Fock-type matrices can be built in a linear-scaling fashion, which is one of the key issues in SCF theories.
- The reduction of the scaling for forming Fock-type matrices leads then to the necessity for avoiding the second rate-determining step in SCF energy computations, the cubically scaling diagonalization step.
- With the described methods, the linear-scaling calculation of SCF energies becomes possible. However, for characterizing stationary points on potential energy surfaces, the calculation of energy gradients is crucial, which is described in the succeeding section.
- To obtain a link to experimental studies, the computation of response properties is often very important. Examples include vibrational frequencies or nuclear magnetic resonance (NMR) chemical shifts, for which the response of the one-particle density matrix with respect to a perturbation needs to be computed. Therefore, we describe ways to reduce the strong increase of the computational effort with molecular size.
- Finally, we provide in the last section a brief outlook on the long-range behavior of electron correlation effects for the example of MP2 theory and show how significant contributions to the correlation energy can be preselected, so that the scaling behavior can be reduced to linear.

---

### SOME BASICS OF SCF THEORY

The simplest approximation used to solve the time-independent Schrödinger equation

$$\hat{H}\Psi = E\Psi \quad [1]$$

within the commonly used Born–Oppenheimer approach<sup>14,15</sup> of clamped nuclei and the electronic Hamiltonian

$$\begin{aligned}\hat{H}_{el} &= -\frac{1}{2}\sum_i \nabla_i^2 - \sum_i \sum_A \frac{Z_A}{r_{iA}} + \sum_i \sum_{j>i} \frac{1}{r_{ij}} \\ &= \sum_i \hat{h}_i + \sum_i \sum_{j>i} \frac{1}{r_{ij}}\end{aligned}\quad [2]$$

is the expansion of the wave function in a Slater determinant<sup>16</sup> as an antisymmetrized product of one-particle functions  $\varphi_i$  (spin orbitals):

$$\begin{aligned}\Psi(\mathbf{r}_1\mathbf{r}_2 \cdots \mathbf{r}_N) &= |\varphi_1\varphi_2 \cdots \varphi_N\rangle \\ &= \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_1(\mathbf{r}_1) & \varphi_2(\mathbf{r}_1) & \cdots & \varphi_N(\mathbf{r}_1) \\ \varphi_1(\mathbf{r}_2) & \varphi_2(\mathbf{r}_2) & \cdots & \varphi_N(\mathbf{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_1(\mathbf{r}_N) & \varphi_2(\mathbf{r}_N) & \cdots & \varphi_N(\mathbf{r}_N) \end{vmatrix}\end{aligned}\quad [3]$$

With this expansion for the wave function, the expectation value using the electronic Hamiltonian (Eq. [2]) can be calculated using the Slater–Condon rules.<sup>4</sup> The result is (in Dirac notation):

$$E_{\text{HF}} = \sum_i \langle \varphi_i | \hat{h} | \varphi_i \rangle + \frac{1}{2} \sum_i \sum_j \langle \varphi_i \varphi_j | | \varphi_i \varphi_j \rangle \quad [4]$$

Minimizing the HF expectation value (Eq. [4]) with respect to orbital rotations while imposing orthonormality constraints leads to the well-known HF equation:<sup>2–4</sup>

$$\hat{F}\varphi_i = \varepsilon_i \varphi_i \quad [5]$$

with  $\hat{F}$  as the Fock operator and  $\varepsilon_i$  as the orbital energy. To algebraize this equation and allow for a suitable solution on computers, it is necessary to expand the one-particle functions in a set of fixed basis functions  $\chi_\mu$  (typically contracted Gaussian basis functions are used in quantum chemistry):

$$\varphi_i = \sum_\mu C_{\mu i} \chi_\mu \quad [6]$$

leading to the Roothaan–Hall equations<sup>17,18</sup>

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\varepsilon} \quad [7]$$

where  $\mathbf{F}$  is the Fock matrix,  $\mathbf{S}$  is the overlap,  $\mathbf{C}$  is the coefficient matrix of the molecular orbitals (MOs), and  $\boldsymbol{\varepsilon}$  is the diagonal matrix of the molecular-orbital energies. The Fock matrix of a closed-shell molecule is built by contracting the one-particle density matrix

$$P_{\mu\nu} = \sum_i^{N_{\text{occ}}} C_{\mu i} C_{\nu i}^* \quad [8]$$

with the four-center two-electron integrals and adding the one-electron part  $h_{\mu\nu}$ :

$$F_{\mu\nu} = h_{\mu\nu} + \sum_{\lambda\sigma} P_{\lambda\sigma} [2(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu)] \quad [9]$$

We use the Mulliken notation for two-electron integrals over (real-valued) Gaussian atomic basis functions in the following:

$$(\mu\nu|\lambda\sigma) = \int \chi_{\mu}(\mathbf{r}_1)\chi_{\nu}(\mathbf{r}_1) \frac{1}{r_{12}} \chi_{\lambda}(\mathbf{r}_2)\chi_{\sigma}(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad [10]$$

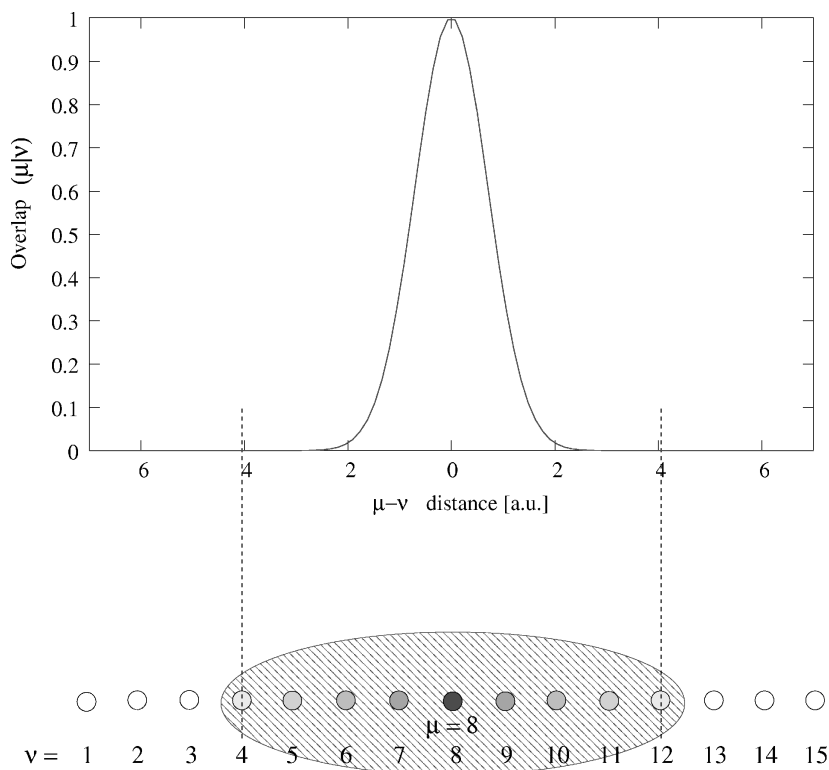
Because the Fock matrix depends on the one-particle density matrix  $\mathbf{P}$  constructed conventionally using the MO coefficient matrix  $\mathbf{C}$  as the solution of the pseudo-eigenvalue problem (Eq. [7]), the SCF equation needs to be solved iteratively. The same holds for Kohn–Sham density functional theory (KS–DFT)<sup>8,9</sup> where the exchange part in the Fock matrix (Eq. [9]) is at least partly replaced by a so-called exchange–correlation functional term. For both HF and DFT, Eq. [7] needs to be solved self-consistently, and accordingly, these methods are denoted as SCF methods.

Two rate-determining steps occur in the iterative SCF procedure. The first is the formation of the Fock matrix, and the second is the solution of the pseudo-eigenvalue problem. The latter step is conventionally done as a diagonalization to solve the generalized eigenvalue problem (Eq. [7]), and thus, the computational effort of conventional SCF scales cubically with system size [ $\mathcal{O}(M^3)$ ].

The construction of the Fock matrix scales formally with  $M^4$  (or more precisely with  $N^4$ ; see discussion above) due to the two-electron integrals being four-index quantities. However, the asymptotic scaling of the number of two-electron integrals reduces to  $\mathcal{O}(M^2)$  for larger molecular systems. This can be understood by considering the following example: The charge distribution of electron 1 in a two-electron integral (Eq. [10]) is described by the product of basis functions  $\chi_{\mu} \cdot \chi_{\nu}$ . If we consider a selected basis function  $\chi_{\mu}$ , then only basis functions  $\chi_{\nu}$  that are “close” to the center of  $\chi_{\mu}$  will form non-vanishing charge distributions. This is because the Gaussian basis functions

decay exponentially with distance. Therefore, the number of basis functions  $\chi_v$  overlapping with the function  $\chi_\mu$  will asymptotically (for large molecules) remain constant with increasing molecular size (in a way one can imagine a “sphere” around the selected basis function as shown in Figure 3). Overall there are  $\mathcal{O}(M)$  basis-function pairs describing each of the two electrons, so that a total of  $\mathcal{O}(M^2)$  two-electron integrals results:

$$\int \underbrace{\chi_\mu(\mathbf{r}_1)\chi_\nu(\mathbf{r}_1)}_{\mathcal{O}(M)} \frac{1}{r_{12}} \underbrace{\chi_\lambda(\mathbf{r}_2)\chi_\sigma(\mathbf{r}_2)}_{\mathcal{O}(M)} d\mathbf{r}_1 d\mathbf{r}_2 \quad [11]$$



**Figure 3** Illustration of the basis functions-pair domain behavior. For a given basis function (or shell)  $\mu$ , only those  $\nu$ 's must be considered whose overlap integrals  $S_{\mu\nu} = (\mu|\nu)$  exceed a certain threshold. In the upper graph the values of the overlap integral is depicted as a function of the  $\mu - \nu$  distance. In the lower graph we consider a linear chain of 15 Gaussian functions (circles) at selected points in space. For  $\mu = 8$  (center atom) and a numerical threshold of  $10^{-7}$ , only the shell pairs closer than  $4.01 \text{ a.u.} = 2.12 \text{ \AA}$  ( $\nu = 4-12$ ) are numerically significant (shaded area); all other  $\nu$ 's (1-3, 13-15) are numerically insignificant for the formation of the selected charge distribution  $\Omega_{\mu\nu}$  and may be neglected. (Chosen distance is  $1 \text{ a.u.}$  Each Gaussian is an  $s$  function of unit exponent.)

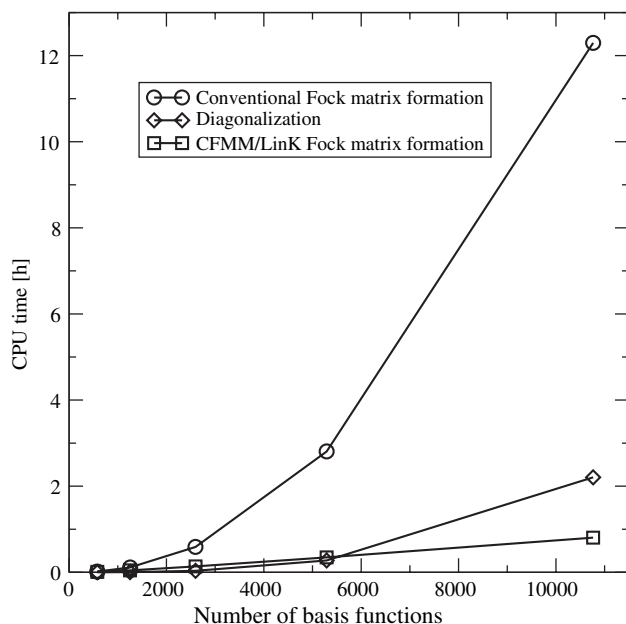
Although the diagonalization of the Fock matrix scales cubically with system size as compared with the quadratic scaling for the formation of the Fock matrix, its prefactor is rather small as shown schematically in Figure 4. Therefore, the diagonalization dominates only for large molecules and/or for fast Fock formation methods as described later.

---

## DIRECT SCF METHODS AND TWO-ELECTRON INTEGRAL SCREENING

In “nondirect” SCF methods, all two-electron integrals are calculated once, stored on disk, and later reused in the subsequent SCF iterations. Because the number of integrals scales formally as  $M^4$ , storing and retrieving two-electron integrals is an extremely expensive step as far as disk space and input/output (I/O) time are concerned. For large molecules the required disk space (and calculation time, see discussion below) easily exceeds all available capacities.

Almlöf et al.<sup>19</sup> observed in a seminal paper that recomputing integrals whenever needed, rather than storing them to disk, could not only be



**Figure 4** Typical timing behavior of the quadratic Fock matrix formation versus the cubically scaling diagonalization step (small prefactor) in SCF energy calculations. The timings for a conventional Fock matrix formation, the linear-scaling CFMM/LinK schemes (as explained later in this review), and Fock matrix diagonalization for a series of DNA molecules  $(A-T)_n$ ,  $n = 1 - 16$  are depicted. Integral threshold is  $10^{-6}$ , basis set 6-31G\*.

competitive to the methods used so far, but also even surpass them as far as computational and storage efficiency are concerned.

Direct schemes have two advantages: First, storage requirements are greatly decreased, and second, calculations for large molecules can actually be made much faster than for nondirect methods, because within the SCF iterations, information on the locality of the molecular system (via the one-particle density matrix; see discussion below) can be exploited. In this way, the introduction of the “direct” SCF approach constitutes an important step toward the applicability of quantum chemical methods to large molecules.

The formation of Fock-type matrices can be schematically divided into two steps:

- Selection of numerically significant integrals: integral screening.
- Calculation of integrals and formation of final matrices.

In the following section, we focus first on the selection of numerically significant integrals, and later we discuss the different contractions of the two-electron integrals.

### Schwarz Integral Estimates

Although the asymptotic  $\mathcal{O}(M^2)$  scaling of the four-center two-electron integrals had been known at least since 1973,<sup>20</sup> it was only in 1989 in the seminal work of Häser and Ahlrichs<sup>21</sup> that an efficient and widely accepted way of rigorously preselecting the numerically significant two-electron integrals was introduced:

$$|(\mu\nu|\lambda\sigma)| \leq |(\mu\nu|\mu\nu)|^{\frac{1}{2}} \cdot |(\lambda\sigma|\lambda\sigma)|^{\frac{1}{2}} = Q_{\mu\nu}Q_{\lambda\sigma} \quad [12]$$

This so-called Schwarz integral screening provides a rigorous upper bound to the four-index integrals, while requiring just the computation of simple two-index quantities. In this way, small four-index integrals below a chosen threshold can be neglected and the formal  $M^4$  scaling associated with the formation of the Fock matrix in HF theory is reduced to  $\mathcal{O}(M^2)$ . As we will see shortly, this was a breakthrough for direct SCF methods<sup>19,21,22</sup> and increased the applicability of SCF methods dramatically.

In contrast to computing the two-electron integrals before an SCF calculation, the key feature of the direct SCF method is the recalculation of two-electron integrals in each SCF iteration. Although this recalculation has the disadvantage of computing the integrals multiple times for building the Coulomb and exchange parts (denoted as  $J$  and  $K$ ) of the Hamiltonian,

$$J_{\mu\nu} - K_{\mu\nu} = \sum_{\lambda\sigma} P_{\lambda\sigma} [2(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu)] \quad [13]$$

it not only avoids the bottleneck of storing the huge number of four-center two-electron integrals, but also it becomes possible to screen the two-electron integrals in combination with the corresponding one-particle density matrix  $\mathbf{P}$  available in each iteration. For example, when calculating the Coulomb part ( $J_{\mu\nu}$ ), integrals are neglected if their contributions to the Coulomb matrix are below a selected threshold of  $10^{-\vartheta}$ :

$$\text{neglect } (\mu\nu|\lambda\sigma), \quad \text{if } |P_{\lambda\sigma}| \cdot Q_{\mu\nu}Q_{\lambda\sigma} \leq 10^{-\vartheta} \quad [14]$$

An analogous screening criterion may be formulated for the exchange matrix  $K_{\mu\nu}$ .

It has to be pointed out that in nondirect SCF, the density matrix is *not* available for screening, because all integrals are calculated prior to the SCF run. Equation [12] has to be employed for screening instead of Eq. [14]. This procedure has a severe disadvantage: Although an integral  $(\mu\nu|\lambda\sigma)$  itself may be large, its contribution  $P_{\lambda\sigma}(\mu\nu|\lambda\sigma)$  to the Coulomb (or, analogously, the exchange) matrix and finally the total energy may be negligible, because the density matrix elements  $P_{\lambda\sigma}$  are often small. Integral screening for nondirect SCF is, therefore, much less efficient than for direct SCF, because a large number of integrals whose contribution to the final result is negligible cannot be discarded due to the missing coupling with the density matrix.

A further improvement on integral screening can be achieved by employing difference densities (c.f. Ref. 19 and 21). The Fock matrices of iterations  $n$  and  $n - 1$  are given by

$$\begin{aligned} \mathbf{F}^{(n)} &= \mathbf{h} + \mathbf{P}^{(n)} \cdot \mathbf{\Pi} \\ \mathbf{F}^{(n-1)} &= \mathbf{h} + \mathbf{P}^{(n-1)} \cdot \mathbf{\Pi} \end{aligned} \quad [15]$$

with  $\mathbf{\Pi}$  as the antisymmetrized two-electron integrals. Instead of constructing the full Fock matrix in each iteration, a recursive scheme as in the following equation may be used:

$$\mathbf{F}^{(n)} = \mathbf{F}^{(n-1)} + \Delta\mathbf{P}^{(n)} \cdot \mathbf{\Pi} \quad [16]$$

with the difference density  $\Delta\mathbf{P}^{(n)}$  for the  $n$ th iteration defined as

$$\Delta\mathbf{P}^{(n)} = \mathbf{P}^{(n)} - \mathbf{P}^{(n-1)} \quad [17]$$

Within this scheme, the number of two-electron integrals needed for the Fock matrix updates  $\Delta\mathbf{P}^{(n)} \cdot \mathbf{\Pi}$  in each iteration may be screened by replacing  $P_{\lambda\sigma}$  with  $\Delta P_{\lambda\sigma}$  in Eq. [14] for the Coulomb part

$$\text{neglect } (\mu\nu|\lambda\sigma), \quad \text{if } \left| \Delta P_{\lambda\sigma}^{(n)} \right| \cdot Q_{\mu\nu}Q_{\lambda\sigma} \leq 10^{-\vartheta} \quad [18]$$

and in an analogous fashion for the exchange part. As the SCF calculation approaches convergence, the change  $\Delta\mathbf{P}^{(n)}$  in the density matrix becomes smaller and smaller and finally approaches zero (within numerical accuracy). The number of two-electron integrals surviving the screening of Eq. [18] is therefore significantly smaller than without difference density screening (Eq. [14]). For an improved algorithm by Häser and Ahlrichs, where the norm of the difference densities  $\Delta\mathbf{P}^{(n)}$  is minimized further, the reader is referred to Ref. 21.

### Multipole-Based Integral Estimates (MBIE)

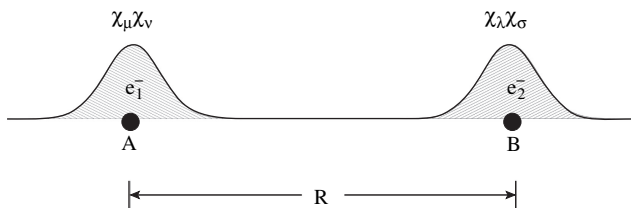
The Schwarz estimates introduced by Häser and Ahlrichs<sup>21</sup> are now used in almost every quantum chemical code for two-electron integral screening. However, they are not optimal in a certain sense: They do not describe the  $1/R$  decay behavior between the charge distributions of the two-electron integrals, as we will explain shortly.

Consider a two-electron repulsion integral (ERI).

$$(\mu\nu|\lambda\sigma) \equiv (A|B) = \int \Omega_A(\mathbf{r}_1) \frac{1}{r_{12}} \Omega_B(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad [19]$$

which consists of the two charge distributions  $\Omega_A$  and  $\Omega_B$  describing the spatial distribution of electrons 1 and 2, respectively (see Figure 5). Here,  $A$  and  $B$  are collective indices for the “bra” and “ket” basis functions, i.e.,  $A = \mu\nu$  and  $B = \lambda\sigma$ . The  $\Omega_A(\mathbf{r}_1)$  and  $\Omega_B(\mathbf{r}_2)$  are Gaussian distributions built as products of two Gaussians  $\Omega_A(\mathbf{r}_1) = \chi_\mu(\mathbf{r}_1) \cdot \chi_\nu(\mathbf{r}_1)$  and  $\Omega_B(\mathbf{r}_2) = \chi_\lambda(\mathbf{r}_2) \cdot \chi_\sigma(\mathbf{r}_2)$ , respectively. The integral describes the Coulomb repulsion between electrons  $e_1^-$  and  $e_2^-$ , whose spatial distribution is represented by  $\Omega_A$  and  $\Omega_B$ , respectively, as illustrated in Figure 5. As stated by Coulomb’s law, the repulsion energy between two charges is proportional to  $1/R$ , where  $R$  is the distance between the two particles. Similarly, for the two-electron integral, one finds (see Ref. 13) that

$$(\mu\nu|\lambda\sigma) \sim \frac{S_{\mu\nu}S_{\lambda\sigma}}{R} \quad [20]$$



**Figure 5** The spatial distributions of electrons  $e_1^-$  and  $e_2^-$  are described by the orbital products  $\Omega_A = \chi_\mu\chi_\nu$  and  $\Omega_B = \chi_\lambda\chi_\sigma$  centered about A and B, respectively. The distance between both centers is denoted as  $R$ .

for sufficiently large separations. We will refer to this in the following as exponential and  $1/R$ -coupling as denoted by

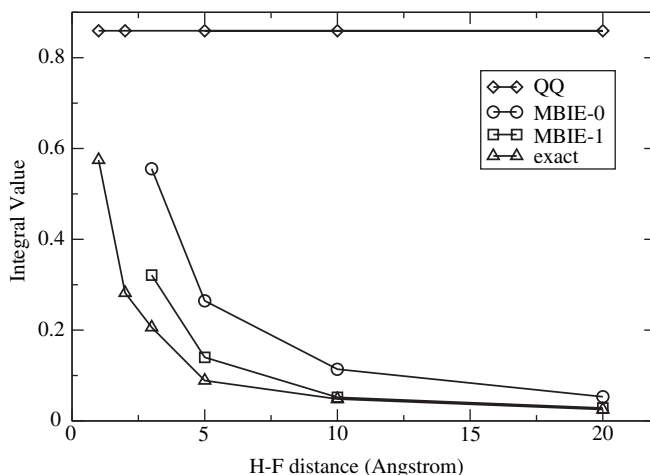
$$\begin{aligned} \text{exponential coupling : } (\mu\nu|\lambda\sigma) &\sim S_{\mu\nu}S_{\lambda\sigma} \sim e^{-a_{\mu\nu}\cdot R_{\mu\nu}^2}e^{-a_{\lambda\sigma}\cdot R_{\lambda\sigma}^2} \\ 1/R - \text{coupling : } (\mu\nu|\lambda\sigma) &\sim \frac{1}{R} \end{aligned} \quad [21]$$

where  $R_{\mu\nu}$  ( $R_{\lambda\sigma}$ ) is the distance between basis function centers  $\chi_\mu$  and  $\chi_\nu$  ( $\chi_\lambda$  and  $\chi_\sigma$ ).  $a_{\mu\nu}$  and  $a_{\lambda\sigma}$  are some constants irrelevant for the following discussion.

Although the Schwarz integral estimates account correctly for the exponential coupling of  $\mu\nu$  and  $\lambda\sigma$ , the  $1/R$  decay when increasing the distance  $R$  between the charge distributions is entirely missing. This is illustrated in Figure 6 where both the exact behavior of a two-electron integral and the Schwarz estimate (abbreviated as QQ) behavior are shown. The  $1/R$  distance decay becomes not only important for the treatment of large molecules in SCF theories, but also in electron correlation methods, where the decay behavior is at least  $1/R^4$ .<sup>23,24</sup> We will return to the latter issue in our outlook on electron correlation methods later in this review.

Almlöf pointed out in 1972<sup>25</sup> that the missing  $1/R$  dependence in the Schwarz screening might be approximated by the following equation via overlap integrals ( $S_A$  and  $S_B$ ):

$$(\mu\nu|\lambda\sigma) \equiv (A|B) \approx \frac{S_A S_B}{R_{AB}} \quad [22]$$



**Figure 6** Comparison of integral estimates MBIE-0, MBIE-1, QQ (Schwarz), and exact  $1/R$ -dependence of two-electron repulsion integrals in a hydrogen-fluoride dimer for integral  $(d_{zz}d_{zz}|p_zp_z)$  with minimum exponents on bra and ket side, respectively, of  $\zeta_{min} = 8.000000 \cdot 10^{-1}$  and  $\zeta_{min} = 6.401217 \cdot 10^{-1}$  using a 6-31G\*\* basis.

However, Eq. [22] does not represent a rigorous upper bound to the two-electron integral. Almlöf<sup>26</sup> as well as Häser and Ahlrichs<sup>21</sup> noted later that nonrigorous bounds for integrals cannot be used in screening as efficiently as rigorous bounds, because the error is uncontrollable. To achieve sufficient accuracy with nonrigorous integral bounds, the thresholds would need to be lowered to an extent that renders them virtually useless for practical applications.

Recently, new multipole-based integral estimates (MBIE) have been introduced by Lambrecht and Ochsenfeld.<sup>23</sup> They are simple, rigorous, and tight upper bounds to the two-electron integrals, and at the same time, they account for the  $1/R$  decay behavior. Because these estimates can be applied generally in quantum chemistry and are expected to be particularly important in view of electron-correlation theories for larger molecules, we briefly outline the main ideas of this MBIE method. For a discussion of the latter in the context of electron correlation, see also the last section of this tutorial.

For a two-electron integral with well-separated charge distributions (we will define this in more detail in the section on multipole expansions of two-electron integrals), it is possible to expand the  $\frac{1}{r_{12}}$  operator in a multipole series as<sup>13,27-29</sup>

$$(\mu\nu|\lambda\sigma) = \frac{MM^{(0)}}{R} + \frac{MM^{(1)}}{R^2} + \frac{MM^{(2)}}{R^3} + \dots \quad [23]$$

where the  $MM^{(n)}$  denote  $n$ -th order multipole terms. For example,  $MM^{(0)}$  describes the monopole-monopole (overlap-overlap) interaction,  $MM^{(1)}$  stands for dipole-monopole terms and  $MM^{(2)}$  contains the quadrupole-monopole and dipole-dipole interactions. This series intrinsically contains the  $1/R$ -dependence for which we aim.

With the definition of “absolute spherical multipoles” of order  $n$ ,  $\mathcal{M}^{(n)}$  as the absolute value of the radial part of spherical multipoles

$$\mathcal{M}_A^{(n)} \equiv \int |\Omega_A(r)r^n| r^2 dr \quad [24]$$

and collecting all the  $n$ th order terms over the absolute multipole integrals by  $\mathcal{MM}^{(n)}$ , we obtain an upper bound to the two-electron integral:

$$|(\mu\nu|\lambda\sigma)| \leq \frac{\mathcal{MM}^{(0)}}{R} + \frac{\mathcal{MM}^{(1)}}{R^2} + \frac{\mathcal{MM}^{(2)}}{R^3} + \mathcal{O}(R^{-4}) \quad [25]$$

Here  $\mathcal{MM}^{(n)}$  stands for expressions involving *absolute* multipole integrals of order  $n$ . Although this expansion represents a rigorous upper bound to the two-electron integral, it is of no practical use in this form, because the series involves, in principle, an infinite (or, at least, a high) number of terms. Discarding the higher order terms would, of course, not lead to a rigorous upper bound.

The key feature of the MBIE method is to replace the higher order terms by lower order ones, while preserving the rigorous upper bound. This is not trivial, but analytical expressions can be derived that relate higher order multipoles to lower order terms.<sup>23</sup> The key idea is illustrated in the following equation:

$$|(\mu\nu|\lambda\sigma)| \leq \left| \frac{\mathcal{M}\mathcal{M}_{\mu\nu\lambda\sigma}^{(0)}}{R'} \sum_{n=0}^{\infty} \left(\frac{1}{R'}\right)^n \right| \quad [26]$$

Here, all multipoles with  $n \geq 1$  are related to the monopole (overlap) term  $\mathcal{M}\mathcal{M}^{(0)}$  by virtue of the analytically derived modified distance  $R'$  (to be described later). This replacement greatly simplifies the form of the series; summing up the geometric series, we obtain the estimate to zeroth order (MBIE-0):

$$|(\mu\nu|\lambda\sigma)| \leq \left| \frac{\mathcal{M}_{\mu\nu}^{(0)}\mathcal{M}_{\lambda\sigma}^{(0)}}{R' - 1} \right| \quad [27]$$

Note that this integral bound contains the  $1/R$ -coupling through the modified distance  $R'$ .

The crucial point of MBIE is that  $R'$  must be changed analytically such that the MBIE expression is a rigorous upper bound. After a tedious derivation,<sup>23</sup> it was found that

$$R' \equiv R - R_{A+B} = R_{AB} - R_A - R_B \quad [28]$$

and

$$R_A \geq K \cdot \left(\frac{1}{2\zeta}\right)^{\frac{n+1}{2n}}, \text{ with } K = \frac{(n+1+l)^{\frac{(n+1+l)}{2n}}}{l^{l/2n}} \cdot \left(\frac{1}{e}\right)^{\frac{n+1}{2n}} \quad [29]$$

guarantees that MBIE-0 is indeed a rigorous upper bound. In the previous equation,  $n$  is the multipole order up to which MBIE is valid.  $l$  and  $\zeta$  denote the total angular momentum and the orbital exponents of the Gaussian basis function product, respectively.

In the foregoing outline, all terms with  $n \geq 1$  were related back to monopoles ( $n_0 = 0$ ). In a similar fashion, we can relate higher order terms back to dipoles (MBIE-1), quadrupoles (MBIE-2), etc. ( $n_0 = 1, 2, \dots$ ). For example, the MBIE-1 criterion, where all higher order terms are related back to expressions over dipoles, has the following form:

$$|(\mu\nu|\lambda\sigma)| \leq \left| \frac{\mathcal{M}_{\mu\nu}^{(0)}\mathcal{M}_{\lambda\sigma}^{(0)}}{R} \right| + \left| \frac{\mathcal{M}_{\mu\nu}^{(1)}\mathcal{M}_{\lambda\sigma}^{(0)} + \mathcal{M}_{\mu\nu}^{(0)}\mathcal{M}_{\lambda\sigma}^{(1)}}{R'^2 - R'} \right| \quad [30]$$

It is important to note that independent of the order, MBIE always guarantees upper bounds to the two-electron integral.<sup>23</sup> The efficiency of the MBIE integral estimates is illustrated in Figure 6, which shows that in contrast to the Schwarz estimates (QQ), MBIE accounts for the  $1/R$  decay behavior of two-electron integrals. For SCF methods, we have found MBIE-1 to be a sufficiently good screening criterion. It overestimates the true number of significant two-electron integrals by just a few percent, while the screening overhead is negligible. The presentation of actual timings with the MBIE screening will be deferred to a later section, once we have introduced the linear-scaling methods for forming the Fock matrix.

As mentioned, it is clear that the MBIE estimates require the validity of the multipole expansion for two-electron integrals, similar to the requirements for the fast multipole methods (these multipole expansions will be presented in detail in the next section). For the near-field part of the integrals, i.e., for charge distributions that are so close that the multipole expansion is not applicable, MBIE cannot be used. Here one can resort to, for example, the Schwarz bounds.

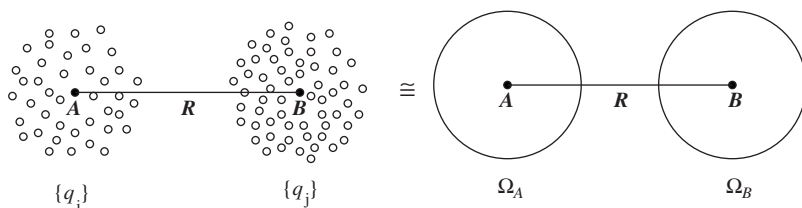
MBIE<sup>23</sup> is the first rigorous integral screening criterion that takes both the exponential and the  $1/R$ -coupling into account. We also point out that the Schwarz bound significantly overestimates the true integral value, if the “bra” and “ket” basis-function exponents are very different, as has been discussed in the work of Gill et al.<sup>30</sup> In contrast, MBIE does not suffer from such a problem.

Because the computation, handling, and contraction of two-electron integrals is central to many quantum chemical methods, it is clear that MBIE can be widely applied. Therefore, by introducing the  $1/R$  dependence in the two-electron integral estimates within MBIE, we not only gain performance for the treatment of large molecules in SCF theories, but MBIE becomes the first screening criterion that allows for the rigorous preselection of contributions to the computation of electron correlation effects in AO-based theories. In these a coupling of at least  $1/R^4$  is observed, which finally leads to linear scaling for electron correlation methods, as we will outline in our outlook on electron correlation methods in a later section of this review.

---

## CALCULATION OF INTEGRALS VIA MULTIPOLE EXPANSION

As we have seen, the number of Coulomb integrals scales as  $\mathcal{O}(M^2)$  for sufficiently large molecules. To overcome this potential bottleneck, the naive pair-wise summation over electron–electron interactions has to be circumvented. We will see that the multipole expansion of the two-electron integrals can be used, thus allowing us to achieve an overall  $\mathcal{O}(M)$  scaling for calculating the Coulomb matrix.



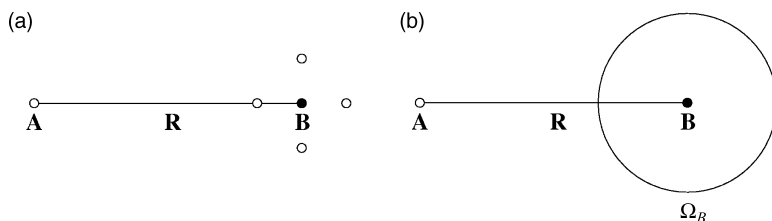
**Figure 7** The interaction of several point charges  $\{q_i\}$  and  $\{q_j\}$  (small open circles) can be approximated as the net interaction term between two charge distributions  $\Omega_A$  and  $\Omega_B$  (large circles).

One advantage of using the multipole expansion in tackling the Coulomb problem is that instead of treating individual pair-wise interactions between point charges, one can collect them into charge distributions and use the total net interaction between these distributions as illustrated in Figure 7. Combined with a clever tree algorithm, the multipole expansion can be used to avoid the quadratic step of summing over pair-wise interactions to obtain an  $\mathcal{O}(M)$  scaling behavior. Another advantage is the separation of “bra” and “ket” quantities, making it possible to precalculate some auxiliary entities before integral calculation itself starts, thus reducing the scaling prefactor.

In the next section, we consider as an introductory example the replacement of individual interactions with effective interactions by using a multipole series. After gaining some basic understanding of the multipole expansion, we derive in detail the spherical multipole expansion, which is one of the most prominent types of multipole expansions used in the calculation of molecular integrals. Once our mathematical tools are derived, we explain an algorithm that scales linearly with the number of interacting particles, namely the fast multipole method (FMM), but that is only suitable for point charges. Then we consider continuous (Gaussian) charge distributions by introducing a generalization of FMM, the continuous fast multipole method (CFMM) in the next section. We complete our tour through multipole methods with a brief overview of other approaches that make use of multipole expansions and tree codes to speed up the calculation of two-electron integrals.

## A First Example

Before deriving and discussing the multipole expansion in detail, let us first have a glimpse at its usefulness by means of a simple example. Imagine we want to calculate the Coulomb interaction energy between a point charge  $q_1$  and a set of point charges  $\{q_2, q_3, q_4, q_5\}$  (all of unit charge) as depicted in Figure 8(a).



**Figure 8** Cartoon illustrating the usefulness of the multipole expansion. (a) Naive approach: The interaction between a point charge (open circle) located at **A** and a set of four point charges distributed around **B** is calculated by summing over each individual pair-wise interaction term. (b) Multipole expansion: The four individual charges have been replaced by their net effect, where they behave like a single new (more complicated) charge distribution  $\Omega_B$ . The separation between **A** and **B** is  $R = 10 \text{ a.u.}$  The point charges of  $\Omega_B$  have a distance of  $r = 1 \text{ a.u.}$  from **B**.

The simplest way to calculate the interaction energy  $U_{1B}$  between  $q_1$  and all other charges is by summing over all four pair-wise interaction terms. Using the geometry described in Figure 8(a), this yields the following result:

$$U_{1B} = q_1 \cdot \sum_{i=2}^5 \phi_i(\mathbf{r}_1) = q_1 \cdot \sum_{i=2}^5 \frac{q_i}{r_{1i}} = 0.4010 \text{ a.u.} \quad [31]$$

Here  $\phi_i(\mathbf{r}_1)$  denotes the electrostatic potential generated by charge  $i$  as it is felt at the location of  $q_1$ . This calculation of the repulsion energy is neither difficult nor time-consuming, but it must not be forgotten that the total number of interaction terms in this naive approach scales like  $\mathcal{O}(M^2)$ : If we want to evaluate all interaction energies between two sets of  $M$  particles  $\{q_{1a}, q_{2a}, \dots, q_{Ma}\}$  and  $\{q_{1b}, q_{2b}, \dots, q_{Mb}\}$ , we end up with a number of pair-wise interaction terms on the order of  $M^2$ . As the number of interacting particles grows, the number of interaction terms soon becomes intractable. For example, a DNA molecule with 1052 atoms would require the calculation of  $1,359,959 \times (1,359,959 - 1)/2 = 924,743,560,861$  pair interactions (with a 6-31G\* basis set and an integral screening threshold of  $10^{-6}$ ), presenting the researcher with the enormous number of almost one trillion interaction terms! For large molecules described by several thousand basis functions, we must therefore avoid the naive quadratic loop over pairs of interacting particles.

This is where the strength of the multipole expansion comes into play. As we will derive later, the potential  $\phi$  arising from an *arbitrary* charge can be expanded in terms of monopole ( $q$ ), dipole ( $\mathbf{D}$ ), quadrupole ( $\mathbf{Q}$ ), and higher order multipole interactions:

$$\begin{aligned} \phi(\mathbf{r}) &= \phi^{(0)}(\mathbf{r}) + \phi^{(1)}(\mathbf{r}) + \phi^{(2)}(\mathbf{r}) + \dots \\ &= \frac{q}{r} + \frac{\mathbf{D} \cdot \hat{\mathbf{r}}}{r^2} + \frac{1}{2} \cdot \frac{\hat{\mathbf{r}} \cdot \mathbf{Q} \cdot \hat{\mathbf{r}}}{r^3} + \dots \end{aligned} \quad [32]$$

where  $\hat{r}$  denotes the unit vector in the direction of  $\mathbf{r}$  and  $r = |\mathbf{r}|$  is the length of  $\mathbf{r}$ . Even if a charge has a complicated structure, for example, is composed of several point charges like in our examples (Figures 7 and 8), its potential can always be expanded in the form of Eq. [32], where the multipoles are those of the *composite* charge distribution  $\Omega$ . Instead of looking at the field generated by the individual point charges, we can therefore consider their *net effect* [Figure 8(b)].

The expansion can be truncated after a finite number of, say,  $L$  terms. For the spherical multipole expansion,  $L = 15\text{--}21$  is known to provide accuracies on the order of  $10^{-7}$  *a.u.* and better in the total energy.<sup>31</sup> Instead of taking into account all  $M$  field terms of the individual charges  $q_i$ , the *total field* is then given by the  $L$  terms of the multipole expansion of the total charge distribution  $\Omega$  as

$$\phi^{\text{total}}(\mathbf{r}) = \sum_i^M \phi_i(\mathbf{r}) \rightarrow \phi_{\text{MP}}^{\text{net}}(\mathbf{r}) \approx \sum_{n=0}^L \phi^{(n)}(\mathbf{r}) \quad [33]$$

In the same manner, the total interaction energy of a point charge  $q_i$  with all other charges  $q_j$  is replaced with the net interaction

$$U^{\text{total}} = q_i \sum_j \frac{q_j}{r_{ij}} \rightarrow U_{\text{MP}}^{\text{net}} \approx q_i \sum_{n=0}^L \phi^{(n)}(\mathbf{r}_i) \quad [34]$$

Thus, instead of having  $\mathcal{O}(M)$  interaction terms for each point charge  $q_i$ , we end up with a sum over the different orders of the multipoles of the charge distribution(s)  $\Omega$ . Because  $L$  is constant for a certain level of accuracy, we can thus calculate the interaction energy between  $q_i$  and one composite charge  $\Omega$  with complexity that is independent of the number of particles belonging to  $\Omega$ .

To see how this works, let us return to our previous example [Figure 8(a,b)], where we combine the four point particles  $\{q_2, q_3, q_4, q_5\}$  into *one* new charge distribution  $\Omega_B$  using the multipole series. Index  $B$  denotes the center of the new charge distribution and also the center of the multipole expansion. The new ‘‘particle’’  $\Omega_B$  certainly has a more complicated structure than its constituent point-charge particles; the composite charge distribution has not only a charge (monopole), but in general also a dipole, quadrupole, and higher moments. We now calculate these lower order multipole moments to evaluate the interaction energy.

The *monopole* is simply the total charge of the constituting point charges:

$$q_B = \sum_{j=2}^5 q_j = 4.0000 \quad [35]$$

The interaction energy between  $q_1$  and  $\Omega_B$  is, therefore, to zeroth order:

$$U_{1B}^{(0)} = \frac{q_1 q_B}{R} = 0.4000 \text{ a.u.} \quad [36]$$

For the next-highest term, the *dipole interaction*, we obtain exactly zero because of symmetry reasons:

$$\begin{aligned} \mathbf{d}_B &= \sum_{j \in B} q_j \mathbf{r}_j = \mathbf{0} \\ U_{1B}^{(1)} &= 0.0000 \text{ a.u.} \end{aligned} \quad [37]$$

The quadrupole moment tensor of  $\Omega_B$  is given as

$$\mathbf{Q}_B = \sum_i \begin{pmatrix} 3x_i^2 - r_i^2 & x_i y_i & x_i z_i \\ y_i x_i & 3y_i^2 - r_i^2 & y_i z_i \\ z_i x_i & z_i y_i & 3z_i^2 - r_i^2 \end{pmatrix} = \begin{pmatrix} 2.0000 & 0.0000 & 0.0000 \\ 0.0000 & -4.0000 & 0.0000 \\ 0.0000 & 0.0000 & 2.0000 \end{pmatrix} \quad [38]$$

and the second-order *quadrupole* interaction term gives the result

$$U_{1B}^{(2)} = \frac{1}{2} \cdot \frac{\hat{\mathbf{r}}_1 \cdot \mathbf{Q}_B \cdot \hat{\mathbf{r}}_1}{R^3} = 0.0010 \text{ a.u.} \quad [39]$$

We truncate our expansion here because a distribution of four point charges can have, at most, a quadrupole moment. For high-accuracy calculations on more complicated distributions, the expansion must be carried out to higher orders.

Putting everything together, we can approximate the interaction energy by

$$U_{1B} \approx U_{1B}^{(0)} + U_{1B}^{(1)} + U_{1B}^{(2)} = 0.4010 \text{ a.u.} \quad [40]$$

Note that for this example the exact result of Eq. [31] is reproduced.

Instead of treating each interaction term of  $q_1$  with  $\{q_2, \dots, q_5\}$  explicitly, we approximated the total net influence of all charges using the multipole expansion and ended up with  $L$  (instead of  $M$ ) interaction terms. The computational workload is thus of order  $\mathcal{O}(L)$  instead of  $\mathcal{O}(M)$  when evaluating the total interaction energy between a single point charge and a set of other point charges. It is clear that if the number of point charges in  $\Omega_B$  is large, using the multipole series with only  $L$  net field terms may lead to significant savings in CPU time.

## Derivation of the Multipole Expansion

The multipole expansion may be carried out in several coordinate systems, which may be chosen depending on the symmetry properties of the problem under investigation. Spherical polar and Cartesian coordinates are used most commonly when calculating two-electron integrals. We outline here the derivation for the spherical series. The interested reader may find more detailed discussions, for example, in the books of Eyring, Walter and Kimball<sup>27</sup> or Morse and Feshbach.<sup>32,33</sup> A discussion of the multipole expansion in the framework of atomic and molecular interactions and potentials may be found in the article of Williams in this series of reviews (Ref. 34) or the book by Hirschfelder, Curtiss and Bird.<sup>28</sup>

Before deriving the multipole series, let us start with some nomenclature. We want to evaluate the electron repulsion integral

$$(\mu\nu|\lambda\sigma) \equiv (A|B) = \int \frac{\Omega_A(\mathbf{r}_1)\Omega_B(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 \quad [41]$$

Here,  $A$  and  $B$  are collective indices for the “bra” and “ket” basis functions; i.e.,  $A = \mu\nu$  and  $B = \lambda\sigma$ .  $\Omega_A(\mathbf{r}_1)$  and  $\Omega_B(\mathbf{r}_2)$  are Gaussian distributions, which are products of two Gaussians  $\Omega_A(\mathbf{r}_1) = \chi_\mu(\mathbf{r}_1) \cdot \chi_\nu(\mathbf{r}_1)$  and  $\Omega_B(\mathbf{r}_2) = \chi_\lambda(\mathbf{r}_2) \cdot \chi_\sigma(\mathbf{r}_2)$ . The centers of the Gaussian distributions are  $\mathbf{A}$  and  $\mathbf{B}$ , respectively. For our task it is handy to express the electronic coordinates by their position relative to the centers  $\mathbf{A}$  and  $\mathbf{B}$  as

$$\begin{aligned} \mathbf{r}_1 &= \mathbf{r}_{1A} + \mathbf{A} \\ \mathbf{r}_2 &= \mathbf{r}_{2B} + \mathbf{B} \end{aligned} \quad [42]$$

and introduce the vector  $\Delta\mathbf{r}_{12}$  as

$$\begin{aligned} \Delta\mathbf{r}_{12} &= \mathbf{r}_{1A} - \mathbf{r}_{2B} \\ \Delta r_{12} &= |\Delta\mathbf{r}_{12}| \end{aligned} \quad [43]$$

The separation between the centers is given by

$$\begin{aligned} \mathbf{R} &= \mathbf{B} - \mathbf{A} \\ R &= |\mathbf{R}| \end{aligned} \quad [44]$$

Our objective is to find a series expansion of the interelectronic distance  $r_{12}$ , which facilitates the separation into an angular and a radial part and which decouples the coordinates of electrons 1 and 2. We follow here the derivation of Eyring et al.<sup>27</sup>

With the definitions introduced in the previous paragraph, the interelectronic separation can be expressed as

$$\begin{aligned} r_{12} &= |\mathbf{r}_{1A} - \mathbf{r}_{2B} + \mathbf{R}| \\ &= \sqrt{\Delta r_{12}^2 + R^2 - 2 \cos \theta \cdot \Delta r_{12} \cdot R} \end{aligned} \quad [45]$$

Here  $\theta$  is the angle subtended between the vectors  $\Delta \mathbf{r}_{12}$  and  $\mathbf{R}$ . We denote the larger radial part of the two vectors with  $r_>$  and the smaller with  $r_<$  as

$$\begin{aligned} r_> &= \begin{cases} R & R > \Delta r_{12} \\ \Delta r_{12} & R < \Delta r_{12} \end{cases} \\ r_< &= \begin{cases} \Delta r_{12} & R > \Delta r_{12} \\ R & R < \Delta r_{12} \end{cases} \end{aligned} \quad [46]$$

and introduce the fraction  $x$  as

$$x = \frac{r_<}{r_>} \quad [47]$$

The interelectronic distance may now be expressed in terms of  $x$  and  $r_>$  (containing all radial dependence) and the angle  $\theta$  (containing all angular dependence):

$$r_{12} = r_> \sqrt{1 + x^2 - 2x \cos \theta} \quad [48]$$

We are now ready to separate the angular and the radial parts of  $r_{12}$ . To this end the angular part of the Coulomb term is expanded in Legendre polynomials  $P_n(\cos \theta)$ , which form a complete and orthonormal set of eigenfunctions on the interval  $\cos \theta \in ]-1, 1[$ , and the radial part is expressed through coefficients  $a_n(x)$  that will be determined shortly:

$$\begin{aligned} r_{12}^{-1} &= \frac{1}{r_>} \frac{1}{\sqrt{1 + x^2 - 2x \cos \theta}} \\ &= \frac{1}{r_>} \sum_{n=0}^{\infty} a_n(x) P_n(\cos \theta) \end{aligned} \quad [49]$$

The coefficients are determined by exploiting the orthonormality of the Legendre polynomials:

$$\int_{-1}^1 P_n(\cos \theta) P_m(\cos \theta) \sin \theta d\theta = \frac{2}{2n+1} \delta_{nm} \quad [50]$$

Squaring the left- and right-hand side terms of Eq. [49], multiplying with the surface element of the unit sphere ( $\sin\theta d\theta$ ) and integrating, we obtain

$$\int \frac{1}{r_{>}^2} \frac{\sin\theta d\theta}{1+x^2-2x\cos\theta} = \int \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} a_n(x)a_m(x)P_n(\cos\theta)P_m(\cos\theta)\sin\theta d\theta$$

$$= \sum_{n=0}^{\infty} \frac{2}{2n+1} a_n^2(x) \quad [51]$$

The left-hand side integral may be calculated easily and yields a logarithmic function of  $x$ . Taylor-expanding this gives

$$\int \frac{1}{r_{>}^2} \frac{\sin\theta d\theta}{1+x^2-2x\cos\theta} = \frac{1}{x} \ln \frac{1+x}{1-x} = \sum_{n=0}^{\infty} \frac{2}{2n+1} x^{2n} \quad [52]$$

Comparing the right-hand side terms of Eqs. [51] and [52], we find that  $a_n(x) = x^n$  and our expansion reads as

$$r_{12}^{-1} = \frac{1}{r_{>}} \sum_{n=0}^{\infty} x^n P_n(\cos\theta) \quad [53]$$

This expansion facilitates the separation of radial and angular parts of the electron–electron and nucleus–nucleus distance vectors.

But we are not yet done with our derivation, because one important goal has not yet been achieved, namely the separation of the electronic coordinates  $\mathbf{r}_1$  and  $\mathbf{r}_2$ . These are still coupled through the angular parts in the Legendre polynomials and inverse powers of  $\Delta r_{12}$  occurring in  $x^n$ . Without decoupling the electron coordinates, we cannot precalculate terms of the series that depend on electrons 1 and 2 independently, and accordingly, we cannot circumvent a quadratic loop over the electronic coordinates, which would spoil our aim of an  $\mathcal{O}(M)$  method.

Invoking the addition theorem of the spherical harmonic functions  $Y_n^m(\theta, \phi)$  as expressed in

$$P_n(\cos\theta) = \sum_{m=-n}^n \frac{4\pi}{2n+1} Y_n^m(\theta_{12}, \phi_{12}) Y_n^{m*}(\theta_{AB}, \phi_{AB}) \quad [54]$$

the angular parts of the electronic and nuclear coordinates can be decoupled and the multipole series becomes

$$r_{12}^{-1} = \frac{1}{r_{>}} \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{4\pi}{2n+1} \frac{r_{\leq}^n}{r_{>}^n} Y_n^m(\theta_{12}, \phi_{12}) Y_n^{m*}(\theta_{AB}, \phi_{AB})$$

$$= \begin{cases} \frac{1}{R} \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{4\pi}{2n+1} \left(\frac{\Delta r_{12}}{R}\right)^n Y_n^m(\theta_{12}, \phi_{12}) Y_n^{m*}(\theta_{AB}, \phi_{AB}) & \text{for } \Delta r_{12} < R \\ \frac{1}{\Delta r_{12}} \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{4\pi}{2n+1} \left(\frac{R}{\Delta r_{12}}\right)^n Y_n^m(\theta_{12}, \phi_{12}) Y_n^{m*}(\theta_{AB}, \phi_{AB}) & \text{for } \Delta r_{12} > R \end{cases} \quad [55]$$

Here  $\Delta r_{12}, \theta_{12}, \phi_{12}$  are the components of  $\Delta \mathbf{r}_{12}$  in spherical coordinates and those of  $\mathbf{R} = \mathbf{R}_{AB}$  are denoted accordingly.

So far this expansion is obviously convergent and holds exactly for both cases  $\Delta r_{12} < R$  and  $R < \Delta r_{12}$ , because  $|P_n(\cos \theta)| \leq 1$ . (For a more detailed discussion the mathematically inclined reader is referred to Refs. 32,33, and 35.)

Clearly the upper branch of the series complies with our goal: It decouples the radial parts of the interelectronic coordinates, because only positive powers of  $\Delta r_{12}$  occur. This, in the end (see next section), facilitates the factorization of the resulting integrals into parts depending only on  $\mathbf{r}_{1A}$  and  $\mathbf{r}_{2B}$ . The integrals are also easy to compute using some standard algorithm for molecular integral evaluation.<sup>36-41</sup>

The lower branch of the expansion, however, does not allow factorization of the integrals and is difficult to calculate. For that reason, when employing the multipole expansion of the two-electron integral, one usually assumes that  $\Delta r_{12} < R$ . That is to say, the charge distributions  $A$  and  $B$  are required to be *nonoverlapping*. With this presumption, only the upper part of the series remains and the multipole expansion of the Coulomb operator for nonoverlapping distributions is

$$r_{12}^{-1} = \frac{1}{R} \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{4\pi}{2n+1} \left( \frac{\Delta r_{12}}{R} \right)^n Y_n^m(\theta_{12}, \phi_{12}) Y_n^{m*}(\theta_{AB}, \phi_{AB}) \quad [56]$$

The presumption  $\Delta r_{12} < R$  deserves some comment. The cautious reader will have noticed that when dealing with continuous charge distributions like Gaussian products, there are always regions of integration in which the charge distributions overlap to some extent such that  $\Delta r_{12} > R$ . Strictly speaking, one would always have to include both branches of the expansion when dealing with charge distributions extending over whole space in order to obtain a convergent and exact series representation of the two-electron integral. In the literature, this is sometimes described by the notion of “asymptotic convergence” of the multipole expansion,<sup>13</sup> which means that the series only converges exactly if the overlap tends to zero or the separation  $R$  between the “bra” and “ket” charge distributions  $\Omega_A$  and  $\Omega_B$  goes to infinity. For an in-depth investigation of this mathematically involved topic, the reader is referred to the original literature, cf. Ref. 42 and references therein.

In practice, however, this does not pose serious problems, because one can derive useful estimates of the error introduced by dropping the second term of the series. One often defines the extent of a Gaussian distribution  $\Omega_A$  as

$$R_A = \frac{\operatorname{erfc}^{-1}(10^{-\vartheta})}{\sqrt{\zeta_A}} \quad [57]$$

where  $10^{-\vartheta}$  denotes the desired accuracy. For Gaussian distributions of  $s$  angular momentum being separated by

$$R_{A+B} > R_A + R_B \quad [58]$$

one can then show that employing the multipole expansion (Eq. [56]) for calculating the  $[ss|ss]$  integral leads to an error on the order of  $10^{-\vartheta}$  times the size of the integral:

$$\begin{aligned} [A|B]^{(\text{exact})} &= [A|B]^{(\text{MP})} + \text{error} \\ \text{error} &\leq 10^{-\vartheta} \cdot [A|B]^{(\text{MP})} \end{aligned} \quad [59]$$

For example, to calculate an  $[ss|ss]$  integral with exponents  $\zeta_A = \zeta_B = 1$  to an accuracy of  $10^{-7}$  using the multipole expansion, centers  $A$  and  $B$  have to be separated by  $R_A + R_B = 4.0\text{\AA}$ . Similar expressions can be derived for higher angular momenta, but the expression for  $s$  Gaussians is usually sufficiently accurate. Together with judicious convergence criteria, the multipole expansion for ERIs produces results that are numerically exact for all practical purposes.

### *Spherical Multipole Expansion for Two-Electron Integrals*

The spherical multipole expansion as derived can be cast into a different form that achieves higher efficiency for computer implementations and finally decouples the angular parts of the electron coordinates.

Notice that when carrying out the multipole summation in the above formulation each term has to be multiplied with a normalization constant  $\frac{4\pi}{2n+1}$  and an inverse power of  $R$ . We can introduce new angular functions replacing the spherical harmonics, which already include the normalization constants, and we can precompute them prior to the evaluation of the two-electron integrals.

To remove the constant fraction in front of each term, the solid harmonics in Racah's normalization are used:

$$C_{nm}(\theta, \phi) = \sqrt{\frac{4\pi}{2n+1}} Y_n^m(\theta, \phi) \quad [60]$$

These are employed to define the scaled regular and irregular solid harmonics of

$$\begin{aligned} R_{nm}(\mathbf{r}) &= \frac{1}{\sqrt{(n-m)!(n+m)!}} r^n C_{nm}(\theta, \phi) \\ I_{nm}(\mathbf{r}) &= \sqrt{(n-m)!(n+m)!} r^{-(n+1)} C_{nm}(\theta, \phi) \end{aligned} \quad [61]$$

In terms of these functions, the one-center multipole expansion of the Coulomb operator reads as

$$r_{12}^{-1} = \sum_{n=0}^{\infty} \sum_{m=-n}^n R_{nm}(\Delta\mathbf{r}_{12}) I_{nm}^*(\mathbf{R}) \quad [62]$$

Using the addition theorem for regular solid harmonics

$$R_{nm}(\mathbf{r} + \mathbf{s}) = \sum_{k=0}^n \sum_{l=-k}^k R_{n-k,m-l}(\mathbf{r}) R_{kl}(\mathbf{s}) \quad [63]$$

we can finally separate the electronic coordinates, and exploiting the behavior of  $R_{kl}(\mathbf{s})$  in going from  $\mathbf{s}$  to  $-\mathbf{s}$ , and rewriting the summation indices, our final multipole expansion of  $r_{12}^{-1}$  with electron coordinates 1 and 2 separated reads:

$$r_{12}^{-1} = \sum_{n=0}^{\infty} \sum_{m=-n}^n \sum_{k=0}^{\infty} \sum_{l=-k}^k (-1)^k R_{nm}(\mathbf{r}_{1A}) I_{n+k,m+l}^*(\mathbf{R}) R_{kl}(\mathbf{r}_{2B}) \quad [64]$$

Until here we concentrated on the multipole expansion of the Coulomb operator. Now we obtain the series for the two-electron integrals. Inserting the expansion into the two-electron integral and absorbing the  $(-1)^k$  prefactor in the interaction matrix  $T$ , we arrive at an efficient multipole expansion of the two-electron integrals:

$$\begin{aligned} [A|B] &= \sum_{n=0}^{\infty} \sum_{m=-n}^n \sum_{k=0}^{\infty} \sum_{l=-k}^k q_{nm}^A(\mathbf{A}) T_{nm,kl}(\mathbf{R}) q_{kl}^B(\mathbf{B}) \\ q_{nm}^A(\mathbf{A}) &= \int \Omega_A(\mathbf{r}) R_{nm}(\mathbf{r}_A) d\mathbf{r} \\ q_{kl}^B(\mathbf{B}) &= \int \Omega_B(\mathbf{r}) R_{kl}(\mathbf{r}_B) d\mathbf{r} \\ T_{nm,kl}(\mathbf{R}) &= (-1)^k I_{n+k,m+l}^*(\mathbf{R}) \end{aligned} \quad [65]$$

Here the  $q$ 's are spherical multipole moments (monopole, dipole, quadrupole, etc.) of charge distributions  $\Omega_A$  and  $\Omega_B$ , respectively. Note that we used square brackets to denote an uncontracted two-electron integral. A suitable generalization for contracted integrals is described in the next section.

Collecting all multipole moments of center  $A$  into a vector  $\mathbf{q}^A(\mathbf{A})$ , those of  $B$  into  $\mathbf{q}^B(\mathbf{B})$ , and arranging the elements of the interaction tensor in matrix form  $\mathbf{T}(\mathbf{R})$ , the multipole expansion can also be formulated in matrix notation as

$$[A|B] = \mathbf{q}^A(\mathbf{A}) \cdot \mathbf{T}(\mathbf{R}) \cdot \mathbf{q}^B(\mathbf{B}) \quad [66]$$

In the following, we will often drop the arguments, because it is clear on which variables the terms depend.

We notice that because each multipole vector has  $\mathcal{O}(L^2)$  components [ $n = 0, \dots, L$ ; for each  $n$ , there are  $2L + 1$   $m$ -components, thus a total of  $L(2L + 1)$ ], the total cost of evaluating a single integral using the multipole expansion has  $\mathcal{O}(L^4)$  complexity. The spherical multipole integrals may be

calculated by means of some well-known recursive algorithms (cf. Refs. 13, 36–41, 43, and 44) in  $O(L^2M)$  work. For the interaction tensor, efficient recursion algorithms also exist (cf. Ref. 13).

The two-electron integrals are for our purposes real entities, so it is clear that using complex terms (solid harmonics) in the multipole expansion is unnecessary and only makes the computer implementation slower and more difficult. A reformulation in terms of real multipole integrals and interaction matrix elements is possible by splitting each term into a real and a complex part and dealing with them separately. After some algebra, one can see that the imaginary part drops out and one obtains the multipole expansion in terms of real-valued multipole moments and interaction matrix elements. The real-valued multipole expansion may be cast into exactly the same form as that of the complex series. As both formulations are formally very similar, we do not introduce the real formulation here but instead refer the interested reader to the literature, e.g., Ref. 13.

### *The Multipole Translation Operator*

So far we have only derived the multipole expansion for primitive Gaussian distributions. As pointed out in the introductory example, one of the main strengths of the multipole expansion is that it can be used to treat the interactions of several primitive charge distributions simultaneously by combining them into one single, albeit more complicated, distribution. It will turn out to be useful to translate the centers of multipole expansions to different points in space; e.g., if  $q(\mathbf{A})$  is an expansion about  $\mathbf{A}$ , we must find a way to convert it to a series about  $\mathbf{A} - \mathbf{t}$ , where  $\mathbf{t}$  is the translation vector.

We first consider a simple case: the transition from primitive to contracted Gaussian distributions. To this end, the multipole expressions for primitive charge distributions  $\Omega_{\alpha\beta} \equiv \Omega_a$  and  $\Omega_{\gamma\delta} \equiv \Omega_b$  have to be contracted with the contraction coefficients  $k_{\alpha\beta}^{\mu\nu}$  and  $k_{\gamma\delta}^{\lambda\sigma}$  as

$$(\mu\nu|\lambda\sigma) = (A|B) = \sum_{\alpha\beta} \sum_{\gamma\delta} k_{\alpha\beta}^{\mu\nu} k_{\gamma\delta}^{\lambda\sigma} \underbrace{\mathbf{q}^{\alpha\beta}(\mathbf{a}_{\alpha\beta}) \cdot \mathbf{T}^{\alpha\beta\gamma\delta}(\mathbf{R}_{\alpha\beta\gamma\delta}) \cdot \mathbf{q}^{\gamma\delta}(\mathbf{b}_{\gamma\delta})}_{=[\alpha\beta|\gamma\delta]} \quad [67]$$

Here,  $\mathbf{R}_{\alpha\beta\gamma\delta}$  is the distance vector between the primitive centers  $\mathbf{a}_{\alpha\beta}$  and  $\mathbf{b}_{\gamma\delta}$ . Using contracted multipole integrals instead of primitive integrals, we obtain the following expansion for contracted integrals:

$$\begin{aligned} (A|B) &= \mathbf{q}^{\mu\nu}(\mathbf{A}) \cdot \mathbf{T}(\mathbf{R}) \cdot \mathbf{q}^{\lambda\sigma}(\mathbf{B}) \\ q_{nm}^{\mu\nu}(\mathbf{A}) &= \sum_{\alpha\beta} k_{\alpha\beta}^{\mu\nu} \int \Omega_{\alpha\beta}(\mathbf{r}) R_{nm}(\mathbf{r}_a) d\mathbf{r} \\ q_{nm}^{\lambda\sigma}(\mathbf{B}) &= \sum_{\gamma\delta} k_{\gamma\delta}^{\lambda\sigma} \int \Omega_{\gamma\delta}(\mathbf{r}) R_{nm}(\mathbf{r}_b) d\mathbf{r} \end{aligned} \quad [68]$$

Note that the primitive expansion is centered at points  $\mathbf{a}_{\alpha\beta}$  and  $\mathbf{b}_{\gamma\delta}$ , which are, in general, different from the centers  $\mathbf{A}$  and  $\mathbf{B}$  of the contracted charge distributions. By contracting we have in fact carried out a translation from the primitive to the contracted centers. What we now must find out is how to translate expansions to an *arbitrary* center.

Recalling the addition theorem for regular solid harmonics (Eq. [63]), we see that a multipole expansion centered at  $\mathbf{a}$  can be translated by a vector  $\mathbf{t}$  to a new center  $\mathbf{A} = \mathbf{a} - \mathbf{t}$  as

$$\begin{aligned}
 q_{nm}^A(\mathbf{A}) &= \int \Omega_a(\mathbf{r}) R_{nm}(\mathbf{r}_a - \mathbf{t}) d\mathbf{r} \\
 &= \sum_{k=0}^n \sum_{l=-k}^k \underbrace{R_{n-k,m-l}(-\mathbf{t})}_{=W_{nm,kl}(\mathbf{t})} \int \underbrace{\Omega_a(\mathbf{r}) R_{kl}(\mathbf{r}_a)}_{=q_{kl}^a(\mathbf{a})} d\mathbf{r}
 \end{aligned} \tag{69}$$

We can therefore translate an expansion by simply multiplying with the “translation operator”  $W_{nm,kl}(\mathbf{t})$  and summing over the angular momenta.

Translating an expansion from a center  $\mathbf{a}$  by  $\mathbf{t}_a$  to a new center  $\mathbf{A}$  may be written in matrix form as

$$\mathbf{q}^A(\mathbf{A}) = \mathbf{W}(\mathbf{t}_a) \cdot \mathbf{q}^a(\mathbf{a}) \tag{70}$$

with a similar expression for the translation from  $\mathbf{b}$  by  $\mathbf{t}_b$  to  $\mathbf{B}$ . With this the multipole expansion reads as

$$(A|B) = \sum_{a \in A} \sum_{b \in B} [\mathbf{W}(\mathbf{t}_a) \mathbf{q}^a(\mathbf{a})] \cdot \mathbf{T}(\mathbf{R}) \cdot [\mathbf{W}(\mathbf{t}_b) \mathbf{q}^b(\mathbf{b})] \tag{71}$$

where the summation runs over all  $a$  and  $b$  centers that need to be shifted to the new centers  $A$  and  $B$ , respectively.

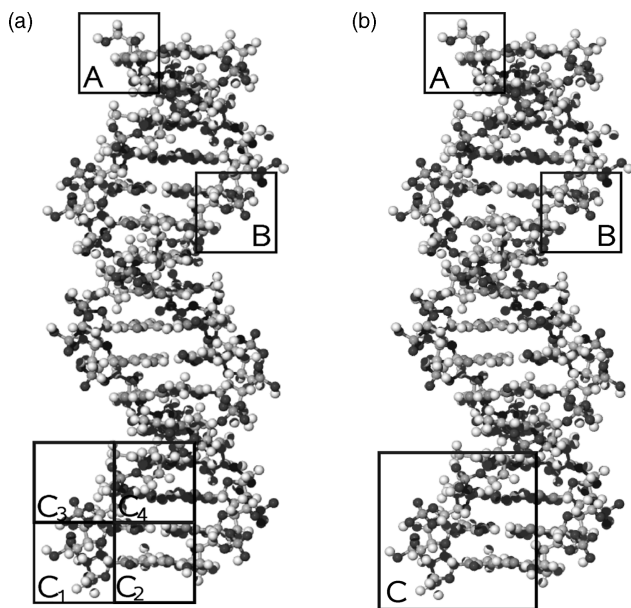
To achieve a specified numerical accuracy for the translation, it is sufficient to truncate the translation series roughly after  $\mathcal{O}(L^2)$  terms. Translating the expansion center of a charge distribution is therefore an  $\mathcal{O}(L^4)$  step. In the next section, we will see that the translation operator is necessary to obtain a true linear-scaling method.

## The Fast Multipole Method: Breaking the Quadratic Wall

We have seen in the introductory example that the multipole expansion may be used to replace particle interactions with their net effects (see Figure 7). But this alone does not reduce the scaling exponent: If there are  $d$  distributions with  $p$  particles on average, we will have to evaluate  $\mathcal{O}(d^2)$  interaction terms. As  $d = M/p$ , this so-called “Naive Multipole Method” still has a

complexity of  $\mathcal{O}(M^2)$  (but with a reduced scaling prefactor). For very large molecules, we would again run into problems due to quadratic scaling. We now outline the FMM of Greengard and coworkers (cf. Refs. 45–47) as a way to reduce the scaling exponent to linear without sacrificing accuracy. An efficient derivation and implementation of the FMM has been presented by White and Head-Gordon.<sup>48,49</sup> It is important to know that FMM was designed for point charges. Molecular integrals, however, involve (Gaussian) charge distributions, which are continuous in space. A generalization of FMM to continuous charge distributions was developed by White, Johnson, Gill, and Head-Gordon<sup>31</sup> and is explained in the next section.

First let us consider an example that captures one of the essentials of FMM. Imagine six boxes  $A$ ,  $B$  and  $C_1$  to  $C_4$ , which, for example, represent fragments of a DNA molecule [Figure 9(a)]. Say that  $A$  and  $B$  are rather close, whereas  $C_1$  to  $C_4$  are approximately four times as far from  $A$ . Throughout the following discussion, we will assume that the interacting particles are evenly distributed among the boxes (for a discussion of approaches to nonevenly distributed charges, see Ref. 46). The interaction energies  $U_{AC_1}, \dots, U_{AC_4}$  will then be approximately four times smaller than  $U_{AB}$ , because the interaction energy is proportional to  $1/R$ .



**Figure 9** In the naive multipole method, (a) the molecule is divided into boxes of equal sizes (for simplicity, only a few selected boxes are shown). For calculating interactions between remote boxes, e.g.,  $A$  and  $C_1$ – $C_4$ , one can use much larger boxes (b) and still achieve high numerical accuracy. This reduces the computational complexity considerably and is an important stepstone on the way to linear scaling.

The absolute error caused by truncating the multipole expansion is also proportional to  $1/R$ .<sup>13</sup> The error is therefore larger for close pairs of boxes of the same size and smaller for remote pairs of the same size. In our example, we could thus combine all four boxes  $C_1$  to  $C_4$  into a four times larger box  $C$  [Figure 9(b)], and the errors in the interaction energies of the close boxes ( $U_{AB}$ ) and of the remote boxes ( $U_{AC}$ ) would still be on the same order of magnitude. The computational cost, however, is reduced drastically by increasing the size of remote boxes.

This idea of resorting to fine grains when describing close interactions and coarse grains for remote interactions is one of the key concepts of the FMM.<sup>45</sup> In the framework of FMM, the graining is realized by a hierarchical boxing scheme, which we outline now for the sake of simplicity in one (rather than three) dimension.

The molecule is divided into a hierarchy of boxes. Each box is divided into two smaller boxes when going from one level to the next: At level 0 there is one box, at level 1 there are two, level 2 contains four boxes, and so on. This is illustrated for four hierarchy levels in Figure 10 (in general there will be more levels). The first box we will call parent (abbreviated as P), and the latter are the child boxes (abbreviated as C). In Figure 10 at the bottom, for example, the parent of box A is called P(A).

In FMM, one distinguishes between near-field (NF) and far-field (FF) interactions. All interactions that can be treated by multipole expansions belong to the far field, all others are near-field. In our illustration of Figure 10, all charges separated by more than two boxes ( $WS = 2$ ) are far-field. Here, WS is the so-called well-separatedness (WS) criterion.

The NF interactions are calculated by conventional methods, e.g., summing over all pairs of charges within the near field. For each level 3 box, there are two to four near-field boxes, depending on whether the box is located at the end or in the inner region of the molecule. The number of particles within the near-field of a particular box therefore remains constant. Altogether,



Figure 10 Illustration of the far-field calculation in FMM.

there are  $\mathcal{O}(M)$  lowest level boxes, and therefore, the total computational cost for evaluating the near-field interactions scales linearly.

The calculation of the far-field interactions in  $\mathcal{O}(M)$  work is the crucial step of the algorithm. We outline it now for the electrostatic field  $\mathbf{V}$  felt by the charges in box  $A$  (see Figure 10). Note that  $\mathbf{V}$  is the algebraized form of the multipole expanded electrostatic potential  $\phi^{(n)}(\mathbf{r})$  we used in the introductory example. For  $A$ 's parent  $P(A)$ , there is only one far-field box with which it interacts: the "parent far-field" PFF( $A$ ), as indicated by an arrow in the illustration. We denote this field as  $\mathbf{V}_{P(A)}^{\text{PFF}}$ . Box  $A$  feels the field of, in principle, five level 3 far-field boxes. Three of them are marked "FF" in the figure. Their field will be denoted as  $\mathbf{V}_A^{\text{FF}}$ . The interaction with the remaining two level 3 boxes can be described at a coarser grain, that is, at level 2. We call every far-field interaction that can be described at a higher boxing level parent far-field (PFF) and all remaining (same-level) interactions are denoted far-field (FF). The union of PFF and FF we call total far-field.

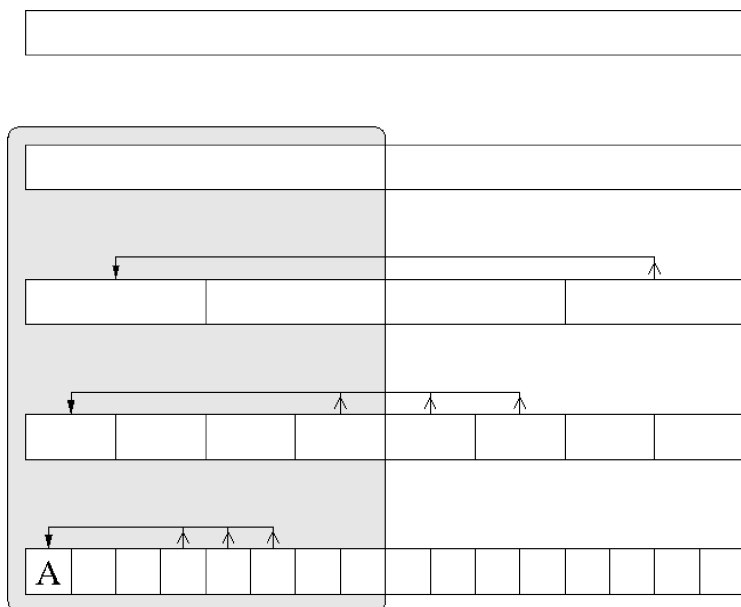
The interaction of box  $A$  with the remaining two far-field boxes is contained in the parent far-field  $\mathbf{V}_{P(A)}^{\text{PFF}}$ . Here the translation operator  $\mathbf{W}$  comes into play:  $\mathbf{V}_{P(A)}^{\text{PFF}}$  is the field felt at the center of  $P(A)$ ; to obtain the field experienced at the center of  $A$ , we must apply the translation operator. Altogether, the total far field of  $A$  is given as

$$\mathbf{V}_A^{\text{total FF}} = \mathbf{V}_{\text{FF}}^A + \mathbf{W}_{AP(A)} \cdot \mathbf{V}_{\text{PFF}}^{P(A)} \quad [72]$$

Note that the interaction of  $A$  with the three closest boxes is calculated at level 3, whereas the more remote interactions are calculated at level 2. In general, FF interactions are evaluated at the highest possible FMM level, i.e., using the largest possible boxes.

Consider now a molecule that is twice as large as the previous one, as illustrated in Figure 11. The area shaded in gray denotes a subunit of the size of our previous example (for comparison). Considering again the FF of a box  $A$ , then in proceeding from the gray subunit to the total system (doubling the size of the system) displayed in Figure 11, there are only three additional interactions with the new part of the system at levels 2 and 3. The number of these additional interactions is constant and does not scale with the total system size in choosing larger systems, because they are done at the highest possible level of boxes. Therefore, because we have a total number of  $\mathcal{O}(M)$  child boxes, the total effort scales with  $\mathcal{O}(M)$ .

Note that the inheritance of the field vectors from parent to child (and vice versa) is crucial for the overall linear scaling of the algorithm. If this could not be done, forming a  $\mathbf{V}$  vector would involve a summation over the higher level boxes, which would lead to an  $\mathcal{O}(M \log M)$  scaling. Finally we note that "inheritance" is only possible through the multipole translation operator, because the field experienced at a parent's center must be shifted to the child's center.



**Figure 11** Illustration of the far field calculation in FMM for a molecule that is twice as large as the example of Figure 10.

Now we are ready to give a detailed description of the algorithm. The FMM can be divided into four steps or, in FMM language, “passes.” In Pass 1, the multipole expansions of all boxes at the lowest level are calculated. In addition, multipole expansions at higher levels are formed by translating the expansions of the lowest box level to higher ones. These are then converted to the multipole-expanded far fields in Pass 2. Then, in Pass 3, the FF is calculated for all boxes by adding their own FF and their parent’s FF (the inheritance trick). Finally, in Pass 4, the interaction energies of all lowest level boxes with their NF and FF is evaluated to yield the total interaction energy. Altogether, FMM calculates the total interaction energy in  $\mathcal{O}(M)$  work.

**Pass 1:**

1. Calculate multipole expansions for all particles  $i$  and boxes  $A$  at the lowest level:

$$\mathbf{q}_A = \sum_{i \in A} \mathbf{q}_{iA}$$

2. Generate multipole expansions for all boxes at higher levels by translating the children’s expansions to the parent’s center:

$$\mathbf{q}_A = \sum_{B \in C(A)} \mathbf{W}_{AB} \cdot \mathbf{q}_B$$

**Pass 2:**

3. Calculate far-field vector for each box at every level:

$$\mathbf{V}_A^{\text{FF}} = \sum_{B \in \text{FF}(A)} \mathbf{T}_{AB} \mathbf{q}_B$$

**Pass 3:**

4. Generate FF vector for each box at every level by adding current boxes' FF and the parent boxes' FF:

$$\mathbf{V}_A^{\text{total FF}} = \mathbf{V}_A^{\text{FF}} + \mathbf{V}_A^{\text{PFF}} = \mathbf{V}_A^{\text{FF}} + \mathbf{W}_{AP(A)} \mathbf{V}_{P(A)}^{\text{FF}}$$

Special cases: Levels 0 and 1 have no FF. For Level = 2 there is no PFF:

$$\mathbf{V}_A^{\text{total FF}} = \mathbf{V}_A^{\text{FF}}$$

**Pass 4:**

5. Calculate interaction energy of each lowest level box with its total FF:

$$U_A^{\text{total FF}} = \sum_{i \in A} \mathbf{q}_{iA} \cdot \mathbf{V}_A^{\text{total FF}}$$

6. Calculate total interaction energy of each lowest level box (NF + total FF):

$$U_A = U_A^{\text{NF}} + U_A^{\text{total FF}} = \sum_{i,j \in \text{NF}(A)} \frac{q_i q_j}{R_{ij}} + U_A^{\text{total FF}}$$

7. The total Coulomb interaction energy is given by the sum over the interaction energies of all lowest-level boxes (factor of one half accounts for double counting of interactions):

$$U = \frac{1}{2} \sum_A U_A$$

## Fast Multipole Methods for Continuous Charge Distributions

So far the FMM considers only point charges. In quantum chemistry, however, we must deal with continuous charge distributions as, for example, in the form of Gaussian distributions. For these continuous distributions, one encounters two difficulties: How to define the spatial extent of a continuous charge distribution (they may extend over the whole space in general), and

how to treat different extents of charge distributions in an efficient way. We discuss here a prominent generalization of FMM to continuous charge distributions, the CFMM,<sup>31</sup> which addresses these issues.

When treating continuous charge distributions with the multipole expansion, we must ensure that the “bra” and “ket” distributions are nonoverlapping to guarantee convergence of the multipole series. Because Gaussian distributions extend over the whole space, they are never nonoverlapping in a strict sense. However, if the contributions of the overlapping regions to the two-integrals are numerically negligible, employing the multipole series causes no problems in practice.<sup>13</sup>

From the analytic expressions for the two-electron integral  $(\Omega_A|\Omega_B)$  over  $s$  Gaussians, we pointed out that the error caused by employing the multipole expansion for calculating the ERI is on the order of  $\epsilon$ , if the “bra-ket” distance  $R$  is chosen such that the following equation holds:

$$\begin{aligned}
 R &> R_A + R_B \\
 R_A &= \frac{\operatorname{erfc}^{-1}(\epsilon)}{\sqrt{\zeta_A}} \\
 R_B &= \frac{\operatorname{erfc}^{-1}(\epsilon)}{\sqrt{\zeta_B}}
 \end{aligned}
 \tag{73}$$

Extended criteria can be derived for higher angular momenta, but in practice, it is usually sufficient to use Eq. [73]. With it, we have a criterion at hand that ensures convergence of the multipole series to the exact value of the integral within numerical accuracy of  $\mathcal{O}(\epsilon)$  even for Gaussians.

We now outline the CFMM as first formulated by White et al.<sup>31</sup> For a discussion of performance issues and a detailed description of implementational considerations, the interested reader is referred to the original literature. A pedagogical introduction to (C)FMM can be found in the book by Helgaker, Jørgensen and Olsen.<sup>13</sup> It is important to notice that although the error estimate of Eq. [73] holds only rigorously for  $s$  functions, the maximum box–box interaction error is used in CFMM. The CFMM error estimate is therefore generally considered an upper bound to the true error.<sup>31</sup>

We must keep track of the spatial extents of Gaussian distributions in order to know when the multipole expansion is applicable and when it is not; in other words, we must know the size of the NF for each Gaussian. To that end, a “well-separatedness criterion”(WS) or, synonymously, an NF width parameter is introduced, which stores the number of boxes by which a pair of equal Gaussian distributions have to be separated to be treated as an FF pair:

$$\text{WS}_n = 2n, \quad n = 1, 2, \dots
 \tag{74}$$

With this nomenclature,  $WS_1$  means that the NF is two boxes wide, whereas  $WS_2$  stands for four boxes and so on (in one dimension).

All Gaussian distributions are sorted into boxes (like in FMM) and into branches of WS parameters according to their extents. That is, a Gaussian of extent  $R_A$  is sorted into the branch with

$$WS_n = \max\left(WS_1, 2 \left\lceil \frac{R_A}{L_{\text{box}}} \right\rceil\right) \quad [75]$$

where  $\lceil \cdot \rceil$  is the ceiling function (smallest integer greater than or equal to argument) and  $L_{\text{box}}$  denotes the size of a lowest level box. Accordingly, the tightest distributions, which have only very small extents, are assigned to the  $WS_1$  branch; less tight distributions are assigned to branches  $WS_n$  with larger  $n$ ; and the most diffuse distributions belong to the branch with largest well-separatedness criterion. WS is chosen to be as small as possible while containing the distribution completely; i.e., the far field is chosen as large as possible so as to benefit from the multipole expansion.

The WS criterion for two distributions of (in general) different extents is given by

$$WS_{mm} = \frac{WS_n + WS_m}{2} \quad [76]$$

This means the two distributions have to be separated by  $WS_{mm}$  boxes in order to be treated as FF interactions.

Apart from the additional assignment of charge distributions into WS branches, the CFMM steps are formally similar to those of FMM. The only important difference between CFMM and FMM is that the width of the NF is chosen according to the spatial extent of the charge distributions and the multiple moments are calculated by integration (instead of summation for point charges) with the former method; everything else stays essentially the same.

Finally, we comment on the computational complexity of CFMM. It is important to notice that the overall computational complexity of CFMM is  $\mathcal{O}(M)$  for calculating the Coulomb integral matrix, if the Gaussian charge distributions are not extremely diffuse. In the limit of exceedingly diffuse distributions, the NF would extend over the whole molecule, which ultimately results in a late onset of linear scaling (but with reduced prefactor in comparison with conventional methods). In practice, however, this problem is usually not observed for the basis sets commonly used in quantum chemistry for calculations on large molecules: Calculating the Coulomb matrix via CFMM is a linear-scaling step.

## Other Approaches

We concentrated here on the linear-scaling calculation of the Coulomb matrix in the frame of (C)FMM, which are used commonly in quantum chemical calculations of large molecules. It should be noted that other tree

codes for large molecules exist like, for example, Barnes–Hut (BH) tree methods<sup>50</sup> or the quantum chemical tree code (QCTC) of Challacombe, Schwegler, and Almlöf.<sup>51,52</sup> These differ in the structure of the tree used for organizing the boxes, or the kinds of expansions used for the integrals. BH methods traditionally use Cartesian multipole expansions,<sup>29,53</sup> whereas the QCTC employs the fast Gauss transform.<sup>54</sup> Several variations of these methods have been reported; see the references in Ref. 52 for example.

Another variation of FMM are the very fast multipole methods (vFMM).<sup>55</sup> In the original FMM formulation, the multipole series is truncated after a maximum angular momentum  $L$ , which is kept constant during the whole calculation. Depending on the shape of the charge distributions and the box size, it may not be necessary to carry out the summation up to  $L$  but, rather, to a smaller angular momentum  $L_{\text{eff}} < L$  to reach a defined level of accuracy for some boxes. This is essentially the idea behind vFMM, which truncates the multipole series at a certain angular momentum  $L_{\text{eff}}$  based on an empirical criterion. Strain, Scuseria, and Frisch developed a variation of CFMM, the Gaussian very fast multipole method (GvFMM),<sup>56</sup> which truncates the series after  $L_{\text{eff}} < L$  terms in the spirit of vFMM.<sup>55</sup>

Finally, we note that one of the most costly steps in calculating the Coulomb matrix using CFMM or GvFMM is the explicit evaluation of the near-field integrals. Although this step is scaling linearly with the size of the molecule, one can further decrease the prefactor by resorting to special methods that speed up the near-field integral calculation. Here we would like to mention just two recently developed methods: The use of auxiliary basis set expansions<sup>57–61</sup> in the multipole accelerated resolution of the identity (MARI-J) approach<sup>62</sup> and the Fourier transform Coulomb (FTC) method.<sup>63–66</sup>

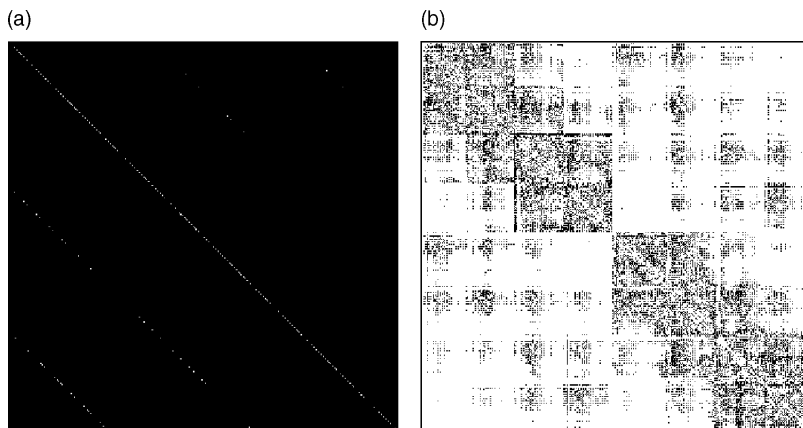
---

## EXCHANGE-TYPE CONTRACTIONS

Now that we have described how to reduce the scaling behavior for the construction of the Coulomb part in the Fock matrix (Eq. [9]), the remaining part within HF theory, which is as well required in hybrid DFT, is the exchange part. The exchange matrix is formed by contracting the two-electron integrals with the one-particle density matrix  $\mathbf{P}$ , where the density matrix elements couple the two sides of the integral:

$$K_{\mu\nu} = \sum_{\lambda\sigma} P_{\lambda\sigma} (\mu\lambda|\nu\sigma) \quad [77]$$

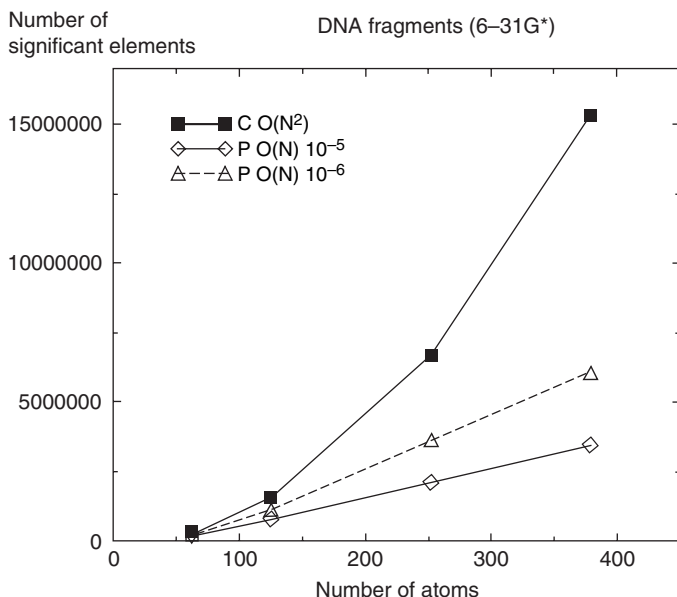
At first sight, it seems that the formation requires asymptotically  $\mathcal{O}(M^2)$  four-center two-electron integrals, so that the overall scaling of the computational effort for building the exchange part would be quadratic. However, the coupling of the two charge distributions of the two-electron integral by the



**Figure 12** Significant elements in (a) the canonical MO coefficient matrix (**C**) and (b) the one-particle density matrix (**P**) computed at the HF/6-31G\* level for four DNA base pairs (DNA<sub>4</sub>). Significant elements with respect to a threshold of  $10^{-7}$  are colored in black.

one-particle density matrix is of central importance for the scaling behavior as will become clear from the discussion below. Therefore, it is crucial to discuss first the behavior of the one-particle density matrix (Equation [8]).

The canonical MO coefficient matrix (**C**) and the one-particle density matrix (**P**) are depicted in Figure 12 as computed at the HF/6-31G\* level for a DNA fragment with four base pairs (DNA<sub>4</sub>). Here, negligible matrix elements below a threshold of  $10^{-7}$  are plotted in white, whereas significant elements are shown in black. The figure clearly illustrates that basically no nonsignificant elements occur in the canonical MO coefficient matrix, because the canonical MOs are typically delocalized over the entire system. This is different for the one-particle density matrix where a considerable number of negligible elements occurs. To reduce the computational scaling behavior by exploiting the localization of the one-particle density matrix, it is not sufficient to have many zero elements in the matrix, but it is necessary that the number of significant elements scales only linearly with system size. This favorable behavior of the density matrix is indeed observed, as shown in Figure 13 again for DNA fragments, in comparison with the  $\mathcal{O}(M^2)$  behavior of the canonical coefficient matrix. It is important to note that the scaling behavior of the number of significant elements in the one-particle density matrix is closely related to the highest occupied molecular orbital-lowest unoccupied molecular orbital (HOMO–LUMO) gap of molecular systems; see Refs. 67–70. Therefore, the asymptotic linear scaling behavior holds only for systems with a nonvanishing HOMO–LUMO gap, so that for a “truly metallic” system, for instance, a quadratic behavior would result. Nevertheless, for a multitude of important chemical and biochemical systems, the scaling of the one-particle density matrix is



**Figure 13** Scaling behavior of significant elements in the one-particle density matrix for DNA fragments computed at the HF/6-31G\* level as compared with the scaling behavior of the MO coefficient matrix. Two different thresholds of  $10^{-5}$  and  $10^{-6}$  are shown as compared with the  $M^2$  behavior of the coefficient matrix.

asymptotically linear. Therefore, the main goal in many linear-scaling theories is to exploit the scaling behavior of the density matrix and to avoid entirely the use of the nonlocal molecular orbital coefficient matrix.

For the linear-scaling formation of the exchange part of the Hamiltonian, the favorable scaling behavior of the one-particle density matrix  $\mathbf{P}$  needs to be exploited. If we have a linear scaling density  $\mathbf{P}$ , then to each index  $\mu$  of a matrix element  $P_{\mu\nu}$ , there can be only a constant number of elements with an index  $\nu$  that is significant with respect to a given threshold. This is nothing else than the definition of a linear-scaling matrix. However, this means that if we consider the formation of the exchange part in the Fock matrix (Eq. [77]), the asymptotically  $\mathcal{O}(M^2)$  number of four-center two-electron integrals are coupled over the one-particle density matrix elements, so that the overall number of required two-electron integrals is reduced to linear for a linear-scaling density matrix:

$$K_{\mu\nu} = \sum_{\lambda\sigma} P_{\lambda\sigma} \underbrace{\left( \underbrace{\mu\lambda}_M \mid \underbrace{\nu\sigma}_M \right)}_M \leftarrow P_{\lambda\sigma} \quad [78]$$

This result also becomes clear in looking at the pseudo-code for the formation of the exchange part:

```
Loop over  $\mu \rightarrow \mathcal{O}(M)$ 
  Loop over  $\lambda \rightarrow \mathcal{O}(1)$ : coupled to  $\mu$  via overlap
    Loop over  $\sigma \rightarrow \mathcal{O}(1)$ : coupled to  $\lambda$  via  $P_{\lambda\sigma}$ 
      Loop over  $\nu \rightarrow \mathcal{O}(1)$ : coupled to  $\sigma$  via overlap
         $K_{\mu\nu} += P_{\lambda\sigma}(\mu\lambda|\nu\sigma)$ 
      endloop
    endloop
  endloop
endloop
```

Here, the outside loop runs over all basis functions  $\mu$  [scaling as  $\mathcal{O}(M)$ ]. The second loop is over the second index of all significant charge distributions  $\Omega_{\mu\lambda}$  (see above), so that this second loop scales as  $\mathcal{O}(1)$  (asymptotically independent of molecular size). The third loop is coupled to the  $\mathcal{O}(1)\lambda$  loop over the linear-scaling one-particle density matrix, therefore scaling as well independently of system size [ $\mathcal{O}(1)$ ]. And finally the last loop is behaving as  $\mathcal{O}(1)$  due to the coupling in forming the charge distributions  $\Omega_{\nu\sigma}$ .

These two simple considerations illustrate that the scaling of the exchange part is linked closely to the scaling behavior of the one-particle density matrix. It is important to note, however, that the onset of the scaling behavior for the exchange formation can be significantly earlier than the one for the density matrix using the same thresholds due to the coupling over the two-electron integrals. We will discuss this in more detail and present timings in the context of the calculation of energy gradients.

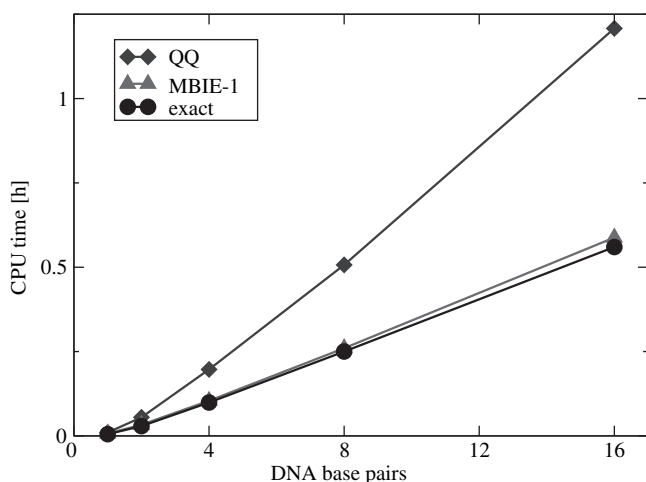
The key to a truly linear-scaling exchange is the efficient implementation of the screening for significant contributions to the exchange matrix in a non-quadratic manner. The time effort for a quadratic screening can be reduced significantly (see Ref. 71), but this depends strongly on the molecular system being studied and the type of exchange contraction (e.g., exchange-type contraction of perturbed densities). It is clear that asymptotically an  $\mathcal{O}(M^2)$  screening procedure would dominate the calculation.

The first attempts to reduce the scaling of the exchange formation required assumptions about the long-range density and exchange behavior.<sup>72,73</sup> Because of these assumptions they were not able to readily ensure a prescribed accuracy. This difficulty was overcome by Schwegler et al. in 1997<sup>74</sup> in their ONX (order N exchange) algorithm by employing the traditional density-weighted Schwarz integral estimates of direct SCF methods<sup>21</sup> within a novel loop structure. For nonmetallic systems, i.e., systems with a nonvanishing HOMO-LUMO gap, this achieves effective linear scaling by using preordered integral estimates, which allow the calculation to leave a loop early and to avoid unnecessary computational effort. However, unlike conventional direct

integral contraction schemes,<sup>19,21,22</sup> the original (compare as well Ref. 75) ONX does not exploit permutational symmetry of the two-electron integrals. It is clear that as long as the exchange formation is dominated by the integral computation, it is favorable to avoid sacrificing permutational symmetry that is, in an ideal case, a factor of four. The need to exploit permutational symmetry is of particular importance if the algorithm reverts to quadratic scaling for small molecules or small band gap systems.

To avoid these problems, the LinK method was introduced by Ochsenfeld et al. in 1998.<sup>71</sup> It reduces the computational scaling for the exchange part to linear for systems with a nonvanishing HOMO–LUMO gap, while preserving the highly optimized structure of conventional direct SCF methods with only negligible prescreening overhead and without imposing predefined decay properties. The LinK method leads to early advantages as compared with conventional methods for systems with larger band gaps. Due to negligible screening overhead, it is also competitive with conventional SCF schemes for both small molecules and for systems with small band gaps. For the formation of an exchange-type matrix in, e.g., coupled perturbed SCF theory, the LinK method achieves sub-linear scaling, or more precisely, independence of the computational effort from molecular size for local perturbations.<sup>71</sup> Because implementing the linear-scaling screening is tricky and does not provide much further insight for the current tutorial, we refer the reader to the original literature for details.<sup>71,76</sup>

We conclude this section by presenting some illustrative results for comparing Schwarz and MBIE screening using test calculations on DNA molecules



**Figure 14** Illustrative timings comparing Schwarz (QQ) and MBIE screening for calculating the exchange matrix for a series of DNA<sub>*n*</sub> molecules ( $n = 1, 2, 4, 8, 16$ ) with up to 1052 atoms (10674 basis functions). All calculations were performed within the LinK method and with a 6-31G\* basis at a threshold of  $10^{-7}$  on an Intel Xeon 3.6 GHz machine.

with up to 1052 atoms. In Figure 14, the calculation time for building one Hartree–Fock exchange matrix using Schwarz (QQ) and MBIE screening, respectively, is shown, while in both schemes, the LinK method is used. We observe that, in both schemes, the calculation time scales indeed linearly with the molecule size as pointed out in the foregoing discussion. A speed-up of the calculation by a factor of 2.1 is observed by employing MBIE as compared with the QQ screening, whereas the numerical error in the exchange energy is preserved and is on the order of 0.1 mHartree for both screening approaches using a threshold of  $10^{-7}$ . We have also compared these timings to “exact” screening, that is, the estimated calculation time that would result if the two-electron integrals were known exactly in the screening process. From the fact that the MBIE and “exact” graphs almost coincide, it is evident that MBIE screening is close to optimal for SCF.

---

## THE EXCHANGE-CORRELATION MATRIX OF KS-DFT

Although the Kohn–Sham-DFT method has been well established in solid-state physics for many years, it was introduced to the computational chemistry community by a reformulation within a finite Gaussian basis set.<sup>59,77–80</sup> Nowadays basically all popular ab initio packages provide a variety of exchange-correlation (XC) functionals that are widely used in computational chemistry and physics.

In this section, we will not present the different types of XC functionals (see Ref. 9 and references therein; Refs. 81 and 82 also treat the recently developed meta-GGA functionals) but discuss only briefly the  $\mathcal{O}(M)$  formation of the XC potential matrix  $\mathbf{V}_{xc}$  in the given basis.<sup>83,84</sup> It has to be mentioned that hybrid XC functionals<sup>85</sup> also contain a certain amount of exact exchange  $\mathbf{K}$ , which can be formed in  $\mathcal{O}(M)$  fashion within the LinK scheme<sup>71,76</sup> as described above.

The XC energy  $E_{xc}$  is in general a functional of the density  $\rho$ . Within the GGA,  $E_{xc}$  is also a functional of the density gradient  $\nabla\rho$ , and within the meta-GGA, it is a functional of  $\rho, \nabla\rho$ , and additionally the kinetic energy density  $\tau$ :

$$E_{xc} = \int f_{xc}[\rho_\alpha(\mathbf{r}), \rho_\beta(\mathbf{r}), \nabla\rho_\alpha(\mathbf{r}), \nabla\rho_\beta(\mathbf{r}), \tau_\alpha(\mathbf{r}), \tau_\beta(\mathbf{r})] d\mathbf{r} \quad [79]$$

The potential  $v_{xc}$  arising from exchange-correlation interactions between electrons is defined by the derivative of the XC energy functional  $E_{xc}$  with respect to the one-particle density  $\rho(\mathbf{r})$  as

$$v_{xc}(\mathbf{r}) = \frac{\partial f_{xc}(\mathbf{r})}{\partial \rho(\mathbf{r})} \quad [80]$$

and the discrete representation in the given basis results from integration over  $\mathbf{r}$  as

$$(\mathbf{V}_{xc})_{\mu\nu} = \langle \chi_\mu | \hat{v}_{xc} | \chi_\nu \rangle = \int \frac{\partial f_{xc}(\mathbf{r})}{\partial \rho(\mathbf{r})} \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r} \quad [81]$$

Because it is typically not possible to determine  $E_{xc}$  and  $\mathbf{V}_{xc}$  by analytic integration, a numerical quadrature has to be used. Therefore, Eq. [79] is rewritten to become

$$E_{xc} = \sum_A^{N_A} \sum_i^{N_{\text{grid}}^A} p_A w_i f_{xc}(\mathbf{r}_i) \quad [82]$$

where  $N_{\text{grid}}^A$  is the number of grid points  $\{\mathbf{r}_i\}$  and  $w_i$  is the weight to the given grid point of atom  $A$ .  $p_A$  is the nuclear partition function that enables a split of the molecular grid into single atomic integral contributions. In a first step, the atomic grids are constructed usually by a combination of radial and angular grid points.<sup>86</sup> After determining the partition factors  $p_A$  by, e.g., the popular method proposed by Becke,<sup>87</sup> the different atomic grids are merged to yield the molecular grid in  $\mathcal{O}(M)$  fashion.

For each atomic grid, the integral contribution is calculated with a scaling behavior independent of system size. After determining the constant number of basis functions  $\chi_\mu$  required for the actual subgrid, as well as the corresponding basis function pairs  $\chi_\mu \chi_\nu$ , the representation of the one-particle density within the partial grid is formed by

$$\rho(\mathbf{r}_i) = \sum_{\mu\nu} P_{\mu\nu} \chi_\mu(\mathbf{r}_i) \chi_\nu(\mathbf{r}_i) \quad [83]$$

with analogous equations for  $\nabla \rho(\mathbf{r}_i)$  and  $\tau(\mathbf{r}_i)$ . At this point it is important to note that the localization or delocalization of the electrons resulting in a sparse or dense discrete density matrix  $\mathbf{P}$  does not effect the scaling behavior of the algorithm; i.e., the strict  $\mathcal{O}(M)$  scaling holds even for metallic systems due to the overlap-type coupling of  $\chi_\mu$  and  $\chi_\nu$  (see also discussion for extremely diffuse basis functions in the context of CFMM).

The evaluation of the XC functional and its derivatives at each point of the sub-grid is followed by the summation of the energy functional values to yield the XC energy  $E_{xc}$ . To form the matrix representation of the corresponding XC potential  $\mathbf{V}_{xc}$  in the given basis, the different first-order derivatives have to be contracted with the corresponding basis function values as

$$\langle \chi_\mu | \hat{v}_{xc} | \chi_\nu \rangle = \sum_A^{N_A} \sum_i^{N_{\text{grid}}^A} p_A w_i \frac{\partial f_{xc}(\mathbf{r}_i)}{\partial \rho(\mathbf{r}_i)} \chi_\mu(\mathbf{r}_i) \chi_\nu(\mathbf{r}_i) \quad [84]$$

For determining higher order derivatives of the XC potential, which are needed for response properties, the implementation can be done in a similar fashion, so that an  $\mathcal{O}(M)$  scaling behavior is ensured as well.

## AVOIDING THE DIAGONALIZATION STEP—DENSITY MATRIX-BASED SCF

In the last sections, we have seen that the Fock matrix can be formed in a linear-scaling fashion. This however means that the second rate-determining step within the SCF approach becomes more important for large molecules due to its  $\mathcal{O}(M^3)$  scaling, although it shows a rather small prefactor: The solution of the generalized eigenvalue problem is typically done by a diagonalization of the Fock matrix. We now discuss general approaches of how to avoid the diagonalization step entirely and to reduce the cubic scaling to linear.

The necessity for diagonalization alternatives is illustrated by the following example: Both the Fock matrix diagonalization and the Fock matrix construction using LinK/CFMM require for a  $DNA_8$  molecule (eight stacked DNA base pairs) with 5290 basis functions approximately 22 minutes on a 3.6-GHz Xeon processor (HF/6-31G\*, threshold  $10^{-7}$ , MBIE screening).<sup>23</sup> This changes for a  $DNA_{16}$  system (with 1052 atoms and 10674 basis functions), where the diagonalization is already more costly than calculating the two-electron integral matrices—141 versus 51 minutes. Therefore it is clearly necessary to circumvent the diagonalization for large molecules.

By diagonalizing the Fock matrix, the canonical MO coefficient matrix (**C**) is obtained (see Eq. [7]). However, we have seen in a previous section that almost all elements in the coefficient matrix are significant, which contrasts with the favorable behavior of the one-particle density matrix (**P**). The density matrix is conventionally constructed from the coefficient matrix by a matrix product (Eq. [8]). Although the Roothaan–Hall equations are useful for small- to medium-sized molecules, it makes no sense to solve first for a nonlocal quantity (**C**) and generate from this the local quantity (**P**) in order to compute the Fock matrix or the energy of a molecule. Therefore, the goal is to solve directly for the one-particle density matrix as a local quantity and avoid entirely the use of the molecular orbital coefficient matrix.

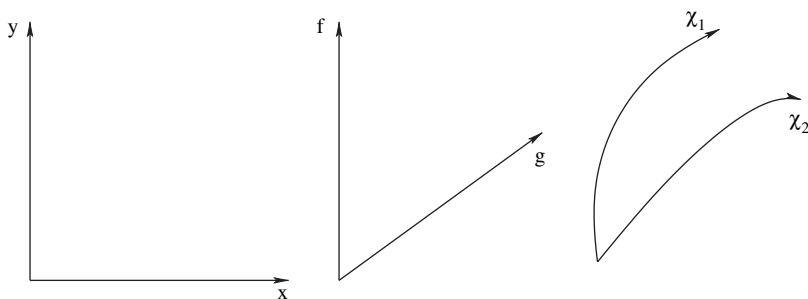
### General Remarks

In the following sections, we will provide an overview of density matrix-based SCF theory that allows one to exploit the naturally local behavior of the one-particle density matrix for molecular systems with a nonvanishing HOMO–LUMO gap. Besides the density matrix-based theories sketched below,<sup>68,88–94</sup> a range of other methods exists, including divide-and-conquer methods,<sup>95–98</sup> Fermi operator expansions (FOE),<sup>99,100</sup> Fermi operator projection (FOP),<sup>101</sup> orbital minimization (OM),<sup>102–105</sup> and optimal basis density-matrix minimization (OBDMM).<sup>106,107</sup> Although different in detail, many share as a common feature the idea of (imposed or natural) localization regions in order to achieve an overall  $\mathcal{O}(M)$  complexity. This notion implies that the density matrix (or the molecule) may be divided into smaller

submatrices (submolecules), of which only a linear-scaling number of fragments may interact with each other. For an overview of these methods, the reader is referred to reviews by Goedecker,<sup>108,109</sup> Scuseria,<sup>84,110</sup> and Bowler et al.<sup>111,112</sup> In the field of *ab initio* quantum chemistry, it seems that density matrix-based schemes are (so far) favored, whereas other diagonalization alternatives are mainly applied to large tight-binding or semi-empirical calculations. We will therefore focus on density matrix-based approaches, which not only allow one to avoid the diagonalization step, but also provide a way for the efficient calculation of molecular response properties such as NMR chemical shifts for large systems.<sup>113</sup> In the next section, we begin by describing some basics of tensor formalisms that are useful (but not necessary) for understanding methods employing nonorthogonal basis functions. That section is followed by a brief outline of selected properties of the density matrix. With this we then turn to the formulation of diagonalization alternatives based on solving directly for the one-particle density matrix.

## Tensor Formalism

To account correctly for the metric of the space spanned by the basis functions (overlap matrix  $S_{\mu\nu}$ ; see also Figure 15), it is convenient to handle operations in a nonorthogonal basis (like the AO basis) using tensor notation. As we will be concerned with AO-formulations of quantum chemical methods in the following, a basic understanding of this topic is useful for comprehending the succeeding sections, although we can only give a very brief introduction here. For a more thorough and yet pedagogical introduction to tensor theory in quantum chemistry along with a detailed list of references to the literature, the interested reader is referred to the review of Head-Gordon et al.<sup>114</sup>



**Figure 15** Basis vectors and functions may be (a) orthogonal, (b) nonorthogonal, or even (c) curvilinear like (a)  $(x, y)$ , (b)  $(f, g)$ , and (c)  $(\chi_1, \chi_2)$  in this illustration. The metric  $g_{\mu\nu} \equiv S_{\mu\nu}$  (which can be identified with the overlap matrix in quantum chemistry) describes uniquely the kind of coordinate system (a–c) spanned by the basis and provides a measure for distances, volumes, etc.

A general introduction to tensor analysis and its relation to Dirac's notation may be found in the book of Schouten.<sup>115</sup>

Every vector  $\mathbf{x}$  in an  $n$ -dimensional linear vector space may be expressed as a linear combination of basis vectors  $\mathbf{e}_i$ :

$$\mathbf{x} = \sum_{i=1}^n x^i \mathbf{e}_i \equiv x^i \mathbf{e}_i \quad [85]$$

where  $x^i$  are the components of  $\mathbf{x}$  in the  $\mathbf{e}_i$  representation. On the right-hand side, we used Einstein's sum convention, which we employ for the sake of brevity whenever applicable. Vectors with lower indices will be called "covariant," e.g., the  $\mathbf{e}_i$  are covariant basis vectors.

The basis vectors are nonorthogonal in general. That is, the scalar product of every pair gives a number  $S$ :

$$\mathbf{e}_i \cdot \mathbf{e}_j = S_{ij}, \quad 0 \leq |S_{ij}| \leq 1 \quad [86]$$

where we assume that the basis vectors are normalized such that  $S_{ii} = 1$ . One could wish, for reasons that will become clear later, to find a second set of basis vectors  $\mathbf{e}^j$ , such that for every  $\mathbf{e}_i$ , there is an  $\mathbf{e}^j$  with

$$\mathbf{e}_i \cdot \mathbf{e}^j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad [87]$$

Equation [87] is similar to the case of normalized orthogonal basis sets (where  $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$ ), with the difference being that one vector comes from the first basis set and the second from the other basis set. For that reason, Eq. [87] is referred to as the "biorthogonality" or "biorthonormality" condition. Basis vectors meeting the biorthogonality requirement with respect to a covariant basis  $\mathbf{e}_i$  will be denoted with an upper index and are called "contravariant." The contravariant basis vectors will also be nonorthogonal in general; i.e.,  $\mathbf{e}^i \cdot \mathbf{e}^j = S^{ij}$ .

Instead of expanding  $\mathbf{x}$  in terms of covariant basis vectors, one may use equally well the contravariant basis as

$$\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}^i \equiv x_i \mathbf{e}^i \quad [88]$$

Here the  $x_i$  are the components of  $\mathbf{x}$  in contravariant representation.

So far we have restricted ourselves to vectors so as to simplify the discussion. Now we turn to tensors. A tensor  $\mathbf{T}^{(k)}$  of rank  $k$  may be seen as an entity whose components are described by

$$\mathbf{T}^{(k)} = \sum_{i_1, i_2, \dots, i_k} T_{i_1, i_2, \dots, i_k} \mathbf{e}_{i_1} \mathbf{e}_{i_2} \dots \mathbf{e}_{i_k} \quad [89]$$

with  $k$  indices  $i_1, i_2, \dots, i_k$ . Note that here the indices are placed below  $T$  and the  $e$ 's to denote that each may either be co- or contravariant depending on the chosen representation.<sup>115</sup> For example, one could choose a set of covariant basis vectors ( $T^{i_1, i_2, \dots, i_k} e_{i_1} e_{i_2} \dots e_{i_k}$ ), a contravariant set ( $T_{i_1, i_2, \dots, i_k} e^{i_1} e^{i_2} \dots e^{i_k}$ ), or even a mixed representation. To be called a tensor, such an entity must obey certain rules concerning coordinate transformations, which we will, however, not discuss here and assume to be fulfilled in the following.

Consider the following examples for illustration: A vector  $\mathbf{a}$  in  $n$ -dimensional space is described completely by its  $n$  components  $a_i$ . It may therefore be seen as a one-index quantity or a tensor of rank one. A matrix  $\mathbf{A}$  has  $n^2$  components  $A_{ij}$  (two indices) and is a rank two tensor. A tensor of rank three has  $n^3$  components, and its components have three indices,  $T_{ijk}$ , and so on. As a special case, scalars have only  $n^0 = 1$  component and are tensors of rank zero.

The following important rules of tensor analysis should be mentioned: Addition and subtraction is only defined for tensors of the same rank and of the same transformation properties (co-/contravariance). For example, adding a matrix and a vector is not valid. Multiplication (also called tensor contraction) is only defined for pairs of indices, where one index is co- and the other is contravariant. As another example,  $x^i x_i$  is a valid tensor contraction, but  $x^i x^i$  is not.

Tensor notation may be applied to quantum chemical entities such as basis functions and matrix elements. For example,  $|\chi_\mu\rangle$  is a covariant tensor of rank one. Like before, superscripts, e.g.,  $|\chi^\mu\rangle$ , denote contravariant tensors. Co- and contravariant basis functions are defined to be *biorthogonal*; that is, they obey the conditions of

$$\langle \chi^\mu | \chi_\nu \rangle = \delta_\nu^\mu \text{ and } \langle \chi_\mu | \chi^\nu \rangle = \delta_\mu^\nu \quad [90]$$

where the first index refers to the “bra” side and the second to the “ket.” For the sake of simplicity, we do not pay attention to the order of indices in the following; e.g., we use  $\delta_\nu^\mu$  instead of  $\delta_\nu^\mu$  or  $\delta_\nu^\mu$ . For an in-depth discussion of this point, see Ref. 116.

Co- and contravariant basis functions are nonorthogonal in general; i.e., the following equation holds:

$$\langle \chi_\mu | \chi_\nu \rangle = S_{\mu\nu} \text{ and } \langle \chi^\mu | \chi^\nu \rangle = S^{\mu\nu} \quad [91]$$

It can be shown that co- and contravariant tensors may be converted into each other by applying the contravariant and covariant metric tensors  $g^{\mu\nu} = (S^{-1})_{\mu\nu} \equiv S^{\mu\nu}$  and  $g_{\mu\nu} = S_{\mu\nu}$  as

$$|\chi^\mu\rangle = g^{\mu\nu} |\chi_\nu\rangle \text{ and } |\chi_\mu\rangle = g_{\mu\nu} |\chi^\nu\rangle \quad [92]$$

where  $S_{\mu\nu}$  is the well-known overlap matrix and  $S^{\mu\nu} = (S^{-1})_{\mu\nu}$  is its inverse.

For tensors  $T^{(2)}$  of rank two, we have the following choices as far as co- and contravariance of the component indices are concerned: (1)  $T_{\mu\nu}$ , (2)  $T^{\mu\nu}$ , (3)  $T_{\mu}^{\nu}$ , and (4)  $T_{\nu}^{\mu}$ . Alternative (1) is said to be “fully covariant,” (2) is “fully contravariant,” and the other two are “mixed” representations. In principle, one is free to formulate physical laws and quantum chemical equations in any of these alternative representations, because the results are independent of the choice of representation. Furthermore, by applying the metric tensors, one may convert between all of these alternatives. It turns out, however, that it is convenient to use representations (3) or (4), which are sometimes called the “natural representation.” In this notation, every “ket” is considered to be a covariant tensor, and every “bra” is contravariant, which is advantageous as a result of the condition of biorthogonality; in the natural representation, one obtains equations that are formally identical to those in an orthogonal basis, and operator equations may be translated directly into tensor equations in this natural representation. On the contrary, in fully co- or contravariant equations, one has to take the metric into account in many places, leading to formally more difficult equations.

Let us look at how to translate the idempotency requirement of the density operator (which will be discussed in the next section more extensively) into a tensor equation as, for example,

$$\hat{\rho}^2 = \hat{\rho} \quad [93]$$

Introducing the matrix elements of the density operator in the natural representation,  $P_{\nu}^{\mu} = \langle \chi^{\mu} | \hat{\rho} | \chi_{\nu} \rangle$ , one may easily cast this operator equation into tensor form:

$$P_{\lambda}^{\mu} P_{\nu}^{\lambda} = P_{\nu}^{\mu} \quad [94]$$

This natural tensor equation is formally similar to the operator equation. If we wish to cast this equation into another (nonorthogonal) representation, we can do so by applying the metric tensor as described above. Let us, for example, rewrite Eq. [94] using the fully contravariant form of the density matrix:

$$\underbrace{P^{\mu\alpha} g_{\alpha\lambda} P^{\lambda\beta} g_{\beta\nu}}_{P_{\lambda}^{\mu} P_{\nu}^{\lambda}} = \underbrace{P^{\mu\beta} g_{\beta\nu}}_{P_{\nu}^{\mu}} \quad [95]$$

Because  $g_{\beta\nu}$  occurs on both sides of the equation, we can remove it, and inserting  $g_{\alpha\lambda} = S_{\alpha\lambda}$ , we arrive at

$$P^{\mu\alpha} S_{\alpha\lambda} P^{\lambda\nu} = P^{\mu\nu}, \text{ or in matrix notation : } \mathbf{PSP} = \mathbf{P} \quad [96]$$

This is the same result that is used in AO-based density matrix-based formulations of quantum chemistry discussed in later sections.

Atomic basis functions in quantum chemistry transform like covariant tensors. Matrices of molecular integrals are therefore fully covariant tensors; e.g., the matrix elements of the Fock matrix are  $F_{\mu\nu} = \langle \chi_\mu | \mathbf{F} | \chi_\nu \rangle$ . In contrast, the density matrix is a fully contravariant tensor,  $P^{\mu\nu} = \langle \chi^\mu | \hat{\rho} | \chi^\nu \rangle$ . This representation is called the “covariant integral representation.”<sup>114,116</sup> The derivation of working equations in AO-based quantum chemistry can therefore be divided into two steps: (1) formulation of the basic equations in natural tensor representation, and (2) conversion to covariant integral representation by applying the metric tensors. The first step yields equations that are similar to the underlying operator or orthonormal-basis equations and are therefore simple to derive. The second step automatically yields tensorially correct equations for nonorthogonal basis functions, whose derivation may become unwieldy without tensor notation because of the frequent occurrence of the overlap matrix and its inverse.

In the following we will tacitly assume some basic knowledge of tensor analysis, especially as far as co- and contravariance is concerned. We will, however, in general not use upper and lower indices to discriminate co- and contravariance, because this is traditionally omitted in quantum chemistry and would greatly complicate the notation. The rare occasions where this tensor notation is needed will be pointed out explicitly.

### Properties of the One-Particle Density Matrix

A system with  $N$  electrons is fully described by the corresponding wave function  $\Psi$  and, following the interpretation of Born,<sup>117</sup>

$$|\Psi|^2 d\mathbf{r}_1 d\mathbf{r}_2 \dots d\mathbf{r}_N = \Psi^*(\mathbf{r}_1 \mathbf{r}_2 \dots \mathbf{r}_N) \Psi(\mathbf{r}_1 \mathbf{r}_2 \dots \mathbf{r}_N) d\mathbf{r}_1 d\mathbf{r}_2 \dots d\mathbf{r}_N \quad [97]$$

represents the probability for finding electron 1 in  $d\mathbf{r}_1$ , electron 2 in  $d\mathbf{r}_2$ , and so on. The probability for an arbitrary electron to be found at  $\mathbf{r}_1$  is obtained by integrating over the positions of the remaining electrons and accounting for the indistinguishability of fermions:

$$\rho(\mathbf{r}_1) = N \int \Psi(\mathbf{r}_1 \mathbf{r}_2 \dots \mathbf{r}_N) \Psi^*(\mathbf{r}_1 \mathbf{r}_2 \dots \mathbf{r}_N) d\mathbf{r}_2 \dots d\mathbf{r}_N \quad [98]$$

which defines the so-called one-particle density function (see Ref. 118). It is important to note that these functions are quadratic in the wavefunction and invariant to unitary transformations of the wave function.

If we consider the Hartree–Fock approach of building a Slater determinant from one-particle functions  $\varphi_i$  (compare our section on the basics of SCF theory), we can similarly define the one-particle density as

$$\rho(\mathbf{r}) = \sum_{i \in \text{occ}} \varphi_i(\mathbf{r}) \varphi_i^*(\mathbf{r}) \quad [99]$$

and the corresponding density operator  $\hat{\rho}$  as

$$\hat{\rho} = \sum_{i \in \text{occ}} |\varphi_i\rangle\langle\varphi_i| \quad [100]$$

The density operator is a projector onto the occupied space, which becomes more clear if one considers, for example, an arbitrary function  $f$  that is expanded in the basis  $\{\varphi_r\}$  as

$$f = \sum_r a_r |\varphi_r\rangle \quad [101]$$

By projection with the density operator, the orthonormality causes all components other than those corresponding to the occupied space to disappear:

$$\hat{\rho}f = \sum_{i \in \text{occ}} \sum_r a_r |\varphi_i\rangle\langle\varphi_i|\varphi_r\rangle = \sum_{i \in \text{occ}} \sum_r a_r |\varphi_i\rangle\delta_{ir} = \sum_{i \in \text{occ}} a_i |\varphi_i\rangle \quad [102]$$

Projecting a second time gives the same result, leading to the idempotency property of projection operators ( $\hat{\rho}^2 = \hat{\rho}$ ) in

$$\hat{\rho}^2 = \sum_{\substack{i \in \text{occ} \\ j \in \text{occ}}} |\varphi_i\rangle\langle\varphi_i|\varphi_j\rangle\langle\varphi_j| = \sum_{i \in \text{occ}} |\varphi_i\rangle\langle\varphi_i| = \hat{\rho} \quad [103]$$

If the one-particle functions  $\varphi_i$  are expanded in basis functions  $\chi_\mu$  as

$$\varphi_i = \sum_\mu C_{\mu i} \chi_\mu \quad [104]$$

the density operator can be written as

$$\hat{\rho} = \sum_{i \in \text{occ}} |\varphi_i\rangle\langle\varphi_i| = \sum_{\mu\nu} \sum_{i \in \text{occ}} C_{\mu i} C_{\nu i}^* |\chi_\mu\rangle\langle\chi_\nu| = \sum_{\mu\nu} P_{\mu\nu} |\chi_\mu\rangle\langle\chi_\nu| \quad [105]$$

with the one-particle density matrix  $\mathbf{P}$  introduced in Eq. [8]. If we consider again the idempotency property of the density operator,

$$\hat{\rho}^2 = \sum_{\mu\nu} \sum_{\lambda\sigma} P_{\mu\nu} |\chi_\mu\rangle\langle\chi_\nu|\chi_\lambda\rangle\langle\chi_\sigma| P_{\lambda\sigma} = \sum_{\mu\nu} P_{\mu\nu} |\chi_\mu\rangle\langle\chi_\nu| = \hat{\rho} \quad [106]$$

then it becomes immediately clear that the following holds:

$$\mathbf{PSP} = \mathbf{P} \quad [107]$$

Note that we have already derived this equation by the help of tensor notation in the previous section. The overlap matrix  $\mathbf{S}$  appears in a nonorthogonal basis and is important for correct contraction with co- and contravariant basis sets. Therefore, either  $\mathbf{PS}$  or  $\mathbf{SP}$  is a projector onto the occupied space depending on the tensor properties of the quantity to which it is applied. The same holds for the complementary projector onto the virtual space  $(\mathbf{1} - \mathbf{PS})$  or  $(\mathbf{1} - \mathbf{SP})$ .

An important technique that we will exploit in the next sections is the decomposition of matrices into occupied–occupied (oo), occupied–virtual (ov/vo), and virtual–virtual (vv) blocks.<sup>118</sup> This is done by projecting these matrices onto the occupied and/or virtual space using the projectors  $\mathbf{PS}$  or  $\mathbf{SP}$  (occupied) and  $(\mathbf{1} - \mathbf{PS})$  or  $(\mathbf{1} - \mathbf{SP})$  (virtual) as it is shown in the following equation for a covariant matrix  $\mathbf{A}$ :

$$\begin{aligned} \mathbf{A} &= \mathbf{1A1} = \mathbf{SPAS} + \mathbf{SPA}(\mathbf{1} - \mathbf{PS}) + (\mathbf{1} - \mathbf{SP})\mathbf{AS} + (\mathbf{1} - \mathbf{SP})\mathbf{A}(\mathbf{1} - \mathbf{PS}) \\ &= \mathbf{A}_{oo} + \mathbf{A}_{ov} + \mathbf{A}_{vo} + \mathbf{A}_{vv} \end{aligned} \quad [108]$$

### Density Matrix-Based Energy Functional

As discussed, for a formulation of SCF theories suitable for large molecules, it is necessary to avoid the nonlocal MO coefficient matrix, which is conventionally obtained by diagonalizing the Fock matrix. Instead we employ the one-particle density matrix throughout. For achieving such a reformulation of SCF theory in a density matrix-based way, we can start by looking at SCF theory from a slightly different viewpoint. To solve the SCF problem, we need to minimize the energy functional of

$$E = \text{tr} \left( \mathbf{P} \mathbf{h} + \frac{1}{2} \mathbf{P} \mathbf{G}(\mathbf{P}) \right) \quad [109]$$

where  $\mathbf{G}(\mathbf{P})$  denotes the two-electron integral matrices contracted with the density matrix. We minimize the energy with respect to changes in the one-particle density matrix,

$$\frac{dE}{d\mathbf{P}} \stackrel{!}{=} 0 \quad [110]$$

enforcing two constraints: First, the idempotency condition of the following equation needs to be accounted for:

$$\mathbf{PSP} = \mathbf{P} \quad [111]$$

and, second, the number of electrons  $N$  must be correct:

$$\text{tr}(\mathbf{PS}) = N \quad [112]$$

These conditions are automatically fulfilled upon diagonalization of the Fock or the Kohn–Sham matrices and the formation of the density matrix (Eq. [8]).

The question now becomes: How do we impose these properties without diagonalization? Li, Nunes, and Vanderbilt (abbreviated as LNV)<sup>88</sup> first realized in the context of tight-binding (TB) calculations (see also the related work of Daw<sup>119</sup>) that insertion of a purification transformation introduced by McWeeny in 1959<sup>120,121</sup> allows one to incorporate the idempotency constraint directly into the energy functional (Eq. [109]). In addition, the constraint of having the correct number of electrons was imposed by fixing the chemical potential  $\mu_{\text{chem}}$  as in:

$$E_{\text{LNV}} = \text{tr} \left( \tilde{\mathbf{P}}(\mathbf{H}^{TB} - \mu_{\text{chem}}\mathbf{1}) \right) \quad [113]$$

where  $\mathbf{1}$  is the unit matrix and  $\tilde{\mathbf{P}}$  denotes the purified density matrix:

$$\tilde{\mathbf{P}} = 3\mathbf{P}\mathbf{S}\mathbf{P} - 2\mathbf{P}\mathbf{S}\mathbf{P}\mathbf{S}\mathbf{P} \quad [114]$$

This purification transformation of McWeeny<sup>120,121</sup> allows one to create, out of a nearly idempotent density matrix  $\mathbf{P}$ ; a more idempotent matrix  $\tilde{\mathbf{P}}$ . The method converges quadratically toward the fully idempotent matrix.<sup>120</sup> The function  $\tilde{x}(x) = 3x^2 - 2x^3$  (for  $\mathbf{S} = \mathbf{1}$ ) is shown in Figure 16 and possesses two stationary points at  $f(0) = 0$  and  $f(1) = 1$ . The purification transformation (Eq. [114]) converges quadratically to an idempotent density matrix whose eigenvalues are either 0 or 1, which correspond to virtual or occupied states, respectively. The necessary convergence condition is that the starting eigenvalues of  $\mathbf{P}$  are in the range  $(-0.5, 1.5)$ .

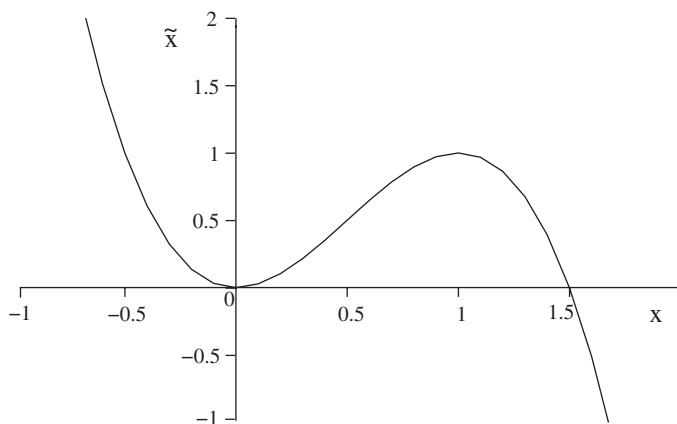


Figure 16 “Purification” function  $\tilde{x} = 3x^2 - 2x^3$ .

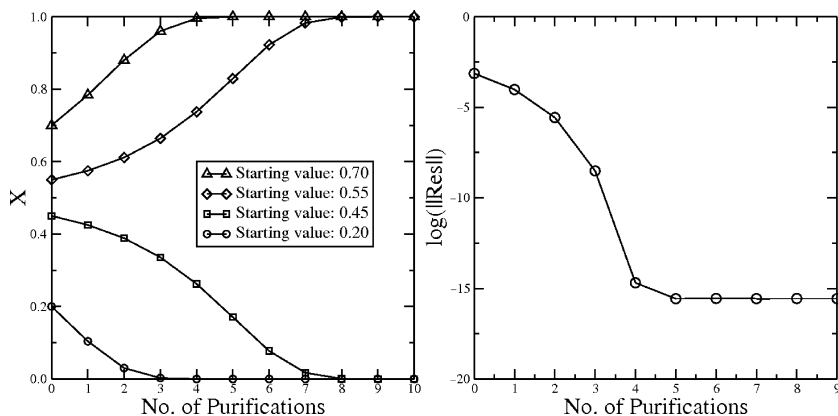
Let us illustrate the purification transformation by some numerical examples. Suppose  $x$  is close to zero, say,  $x = 0.1$ . Purification will then bring it closer to zero:  $\tilde{x} = 3(0.1)^2 - 2(0.1)^3 = 0.028$ . Suppose, on the other hand, that  $x = 0.9$ , that is, close to one. This time the purity transformation brings it closer to 1:  $\tilde{x} = 3(0.9)^2 - 2(0.9)^3 = 0.972$ . We have also illustrated the convergence of the purification transformation for several starting values of  $x$  and the density matrix in Figure 17.

In passing by, we note that the original LNV approach was designed for orthogonal basis functions. Nunes and Vanderbilt later presented a generalization to nonorthogonal problems<sup>89</sup> (see as well later work by White et al. in Ref. 122). A modified LNV scheme for SCF theories was introduced by Ochsenfeld and Head-Gordon.<sup>68</sup> Similarly Millam and Scuseria<sup>90</sup> presented as well an extension of the LNV algorithm to the HF method.

In the derivation of density matrix-based SCF theory below, we do not employ the chemical potential introduced by LNV,<sup>88</sup> but instead we follow the derivation of Ochsenfeld and Head-Gordon, because McWeeny's purification automatically preserves the electron number.<sup>68</sup> Therefore, to avoid the diagonalization within the SCF procedure, we minimize the energy functional

$$\tilde{\mathbf{E}} = \text{tr} \left( \tilde{\mathbf{P}}\mathbf{h} + \frac{1}{2}\tilde{\mathbf{P}}\mathbf{G}(\tilde{\mathbf{P}}) \right) \quad [115]$$

with respect to density-matrix changes, where  $\tilde{\mathbf{P}}$  is the inserted purified density. The simplest approach is therefore to optimize the density matrix (e.g., starting



**Figure 17** Convergence of purification transformation for different starting values (left). Purification of the density matrix after a typical geometry optimization step within D-QCSCF (see the following section for a definition) calculation (right), the logarithmic value of the norm of the residual ( $\log\|\mathbf{P}^{(i)} - \mathbf{P}^{(i-1)}\|$ ) is plotted.

with a trial density matrix  $\mathbf{P}^{(i)}$  by searching for an energy minimum along the direction of the negative energy gradient:

$$\mathbf{P}^{(i+1)} = \mathbf{P}^{(i)} - s \cdot \frac{\partial E[\mathbf{P}^{(i)}]}{\partial \mathbf{P}^{(i)}} \quad [116]$$

where  $s$  is the step length. The gradient is built by forming the derivative of the energy functional (Eq. [115]):

$$\frac{\partial \tilde{E}}{\partial \mathbf{P}} = \frac{\partial \tilde{E}}{\partial \tilde{\mathbf{P}}} \frac{\partial \tilde{\mathbf{P}}}{\partial \mathbf{P}} = 3\text{FPS} + 3\text{SPF} - 2\text{SPFPS} - 2\text{FPSPS} - 2\text{SPSPF} \quad [117]$$

At convergence this energy gradient expression reduces to the usual criterion of  $\text{FPS} - \text{SPF} = 0$ . It is important to note that the covariant energy gradient (Eq. [117]) cannot be added directly to the contravariant one-particle density matrix, so that a transformation with the metric is required.

Therefore, let us look briefly at the tensor properties of the energy gradient. Rewriting the energy gradient (Eq. [117]) in tensor notation,

$$\begin{aligned} \left( \frac{\partial \tilde{E}}{\partial \mathbf{P}} \right)_{\mu\nu} &= (\nabla E)_{\mu\nu} = 3F_{\mu\lambda} P^{\lambda\sigma} S_{\sigma\nu} + 3S_{\mu\lambda} P^{\lambda\sigma} F_{\sigma\nu} \\ &\quad - 2S_{\mu\lambda} P^{\lambda\sigma} F_{\sigma\alpha} P^{\alpha\beta} S_{\beta\nu} \\ &\quad - 2F_{\mu\lambda} P^{\lambda\sigma} S_{\sigma\alpha} P^{\alpha\beta} S_{\beta\nu} \\ &\quad - 2S_{\mu\lambda} P^{\lambda\sigma} S_{\sigma\alpha} P^{\alpha\beta} F_{\beta\nu} \end{aligned} \quad [118]$$

we note immediately that this gradient is a fully *covariant* tensor. Because the density matrix is fully *contravariant* in “covariant integral representation,” it is tensorially inconsistent to generate a new density matrix by adding the fully covariant gradient to the fully contravariant density matrix  $\mathbf{P}^{(i)}$ . Converting the covariant to contravariant indices by applying the inverse metric, we find the tensorially consistent formulation of the energy gradient<sup>116</sup> as follows:

$$\begin{aligned} (\nabla E)^{\mu\nu} &= g^{\mu\lambda} (\nabla E)_{\lambda\sigma} g^{\sigma\nu} \\ &= (S^{-1})^{\mu\lambda} (\nabla E)_{\lambda\sigma} (S^{-1})^{\sigma\nu} \end{aligned} \quad [119]$$

With this the fully contravariant energy gradient results:

$$(\nabla E) = \frac{\partial E}{\partial \mathbf{P}} = 3\mathbf{S}^{-1}\mathbf{FP} + 3\mathbf{PFS}^{-1} - 2\mathbf{PFP} - 2\mathbf{S}^{-1}\mathbf{FPSP} - 2\mathbf{PSPFS}^{-1} \quad [120]$$

Because all matrices for its formation can be built in a linear-scaling fashion and because they are sparse for systems with a nonvanishing

HOMO–LUMO gap, the energy gradient with respect to the density matrix can be built with linear-scaling effort. Due to symmetries, only a few sparse matrix multiplications are required for the computation of the gradient. In this way, it is possible to avoid the diagonalization in the SCF procedure, thereby reducing the computational scaling asymptotically to linear<sup>68,122</sup> for large molecules.

### “Curvy Steps” in Energy Minimization

In the simplest density matrix-based method, optimization steps are taken in the direction of the negative gradient  $-\nabla E$ . The best one can therefore do is to find the minimum energy along a straight line defined by the gradient direction in each step. One can show, however, that it is possible to find the minimum along a curved path at essentially no additional cost,<sup>91,92</sup> which potentially leads to more efficient minimization steps. With this approach, the idempotency condition is fulfilled through higher orders than in the density matrix-based scheme described above, where just the lowest-order purification transformation of McWeeny<sup>121</sup> enters the formulation.

It is useful to describe the generation of a new density matrix from the previous matrix by unitary transformation:

$$\tilde{\mathbf{P}} = \mathbf{U}^\dagger \mathbf{P} \mathbf{U} \quad [121]$$

Every unitary matrix  $\mathbf{U}$  can be represented by an exponential function of an anti-Hermitian matrix  $\mathbf{\Lambda}$ <sup>13</sup>

$$\mathbf{U} = e^{\mathbf{\Lambda}} = \mathbf{1} + \mathbf{\Lambda} + \frac{1}{2!} \mathbf{\Lambda}^2 + \dots \quad [122]$$

or in the tensor notation introduced in the previous section as

$$U_{\nu}^{\mu} = (e^{\mathbf{\Lambda}})_{\nu}^{\mu} \equiv e^{\Delta_{\nu}^{\mu}} \quad [123]$$

In this notation, the exponential parametrization of the new density matrix becomes

$$\tilde{\mathbf{P}}_{\nu}^{\mu} = (\mathbf{U}^{\dagger})_{\lambda}^{\mu} \mathbf{P}_{\sigma}^{\lambda} \mathbf{U}_{\nu}^{\sigma} = e^{-\Delta_{\lambda}^{\mu}} \mathbf{P}_{\sigma}^{\lambda} e^{\Delta_{\nu}^{\sigma}} \quad [124]$$

The density matrix (and thus the Hartree–Fock energy) can now be seen as functions of the parameter  $\mathbf{\Lambda}$  and the requirement for an energy minimum becomes

$$(\nabla E)_{\mu}^{\nu} = \frac{\partial E}{\partial \Delta_{\nu}^{\mu}} = \frac{\partial E}{\partial \tilde{\mathbf{P}}_{\sigma}^{\lambda}} \frac{\partial \tilde{\mathbf{P}}_{\sigma}^{\lambda}}{\partial \Delta_{\nu}^{\mu}} = F_{\lambda}^{\sigma} \frac{\partial \tilde{\mathbf{P}}_{\sigma}^{\lambda}}{\partial \Delta_{\nu}^{\mu}} = 0 \quad [125]$$

If one is not yet at a minimum, one can for example use the method of steepest descent to arrive at the optimum density matrix. Inserting the explicit dependence of  $\tilde{\mathbf{P}}_v^\mu$  on  $\Delta_v^\mu$ , e.g., in the form of the Taylor expansion of Eq. [124] around  $\Delta_v^\mu = 0$ , one obtains for the direction of steepest descent:

$$\Delta_v^\mu = - \left( \frac{\partial E}{\partial \Delta_v^\mu} \right) \Bigg|_{\Delta=0} = - (F_\lambda^\mu P_v^\lambda - P_\lambda^\mu F_v^\lambda) \quad [126]$$

Until now this ansatz is similar to the LNV approach (reformulated in natural tensor notation), starting from an almost idempotent density matrix. (Inserting the purity-transformed density matrix and going to covariant integral representation, the previous equation yields exactly the same result as in the LNV approach.) One could search along this direction in a steepest descent manner to reach the energy minimum. It is instructive to notice that these searches along a straight line may be interpreted as truncating the Taylor series of the exponential transformation after the linear term:

$$\tilde{\mathbf{P}} = \left( \mathbf{1} - \Delta + \frac{1}{2!} \Delta^2 + \dots \right) \mathbf{P} \left( \mathbf{1} + \Delta + \frac{1}{2!} \Delta^2 + \dots \right) \approx \mathbf{P} - \Delta \mathbf{P} + \mathbf{P} \Delta \quad [127]$$

Now we introduce “curvy steps.” An intuitive interpretation of what is done here is to expand the Taylor series of the exponential transformation to higher orders, such that the step directions are no longer straight lines, but instead they are curved. Invoking the Baker–(Campbell)–Hausdorff lemma (see, e.g., Ref. 123), the unitary transformation of the density matrix can be written as

$$\tilde{\mathbf{P}} = \mathbf{P} + [\mathbf{P}, \Delta] + \frac{1}{2!} [[\mathbf{P}, \Delta], \Delta] + \frac{1}{3!} [[[ \mathbf{P}, \Delta ], \Delta ], \Delta] + \dots \quad [128]$$

or

$$\tilde{\mathbf{P}}_v^\mu = \sum_{j=0} \frac{1}{j!} \left( P^{[j]} \right)_v^\mu$$

where the  $P^{[j]}$  are short-hand notations for nested commutators, which can be calculated by recursion using

$$\left( P^{[j+1]} \right)_v^\mu = [P^{[j]}, \Delta]_v^\mu \quad [129]$$

In a similar way, the Hartree–Fock trial energy, as a function of the transformed density matrix  $\tilde{\mathbf{P}}$ , can be written as a series in the step length  $s$ , as

$$\tilde{E} = \sum_{j=0} \frac{s^j}{j!} \tilde{E}^{[j]}, \quad \text{with } \tilde{E}^{[j]} = \left( P^{[j]} \right)_v^\mu F_\mu^v \quad [130]$$

This equation describes the dependence of the trial energy on the step length along a *curved* step direction given by the  $\tilde{E}^{[j]}$ 's and  $P^{[j]}$ 's.

In the “curvy steps” approach, higher terms of the Taylor expansion may be retained by including higher order commutators in the sum of Eq. [130] (letting  $j$  run to high orders), which corresponds to making steps along curved (polynomial) directions. If the series of Eq. [130] was truncated after  $j = 1$ , one would obtain essentially the same step directions as in the LNV approach (starting from an idempotent density matrix).

A step along a curved direction is superior to one along a straight line and should lead to faster convergence as far as the number of iterations is concerned, because higher order terms of the Taylor expansion are kept in the transformation of the density matrix. If all intermediate matrices are stored in memory, searching along curved directions is not more expensive than for straight-line steps; therefore, Head-Gordon and coworkers.<sup>91</sup> find the curvy steps method to be faster than the LNV approach.

### Density Matrix-Based Quadratically Convergent SCF (D-QCSCF)

We have shown, in principle, how to circumvent the diagonalization and introduced two alternatives for choosing the density updates—the methods of steepest descent and “curvy steps.” Now we derive another density update, on which the density matrix-based quadratically convergent SCF method (D-QCSCF) of Ochsenfeld and Head-Gordon<sup>68</sup> is based. This will also be our starting point in deriving linear-scaling methods for energy derivatives needed to determine response properties like vibrational frequencies or NMR chemical shifts, which are described in the next two sections.

To minimize the energy functional (Eq. [115]) with respect to density changes

$$\frac{d\tilde{E}}{d\mathbf{P}} \stackrel{!}{=} 0 \tag{131}$$

we can use, for example, a Newton–Raphson scheme.<sup>124</sup> The Taylor expansion of the energy functional around  $\mathbf{P}$  in changes of the density matrix ( $\mathbf{P}_\Delta$ ) is given as

$$\tilde{E}(\mathbf{P} + \mathbf{P}_\Delta) = \tilde{E}(\mathbf{P}) + \left. \frac{d\tilde{E}}{d\mathbf{P}} \right|_{\mathbf{P}_\Delta=0} (\mathbf{P}_\Delta) + \frac{1}{2} \left. \frac{d^2\tilde{E}}{d\mathbf{P}^2} \right|_{\mathbf{P}_\Delta=0} (\mathbf{P}_\Delta)^2 + \dots \tag{132}$$

For small changes  $\mathbf{P}_\Delta$  in the density  $\mathbf{P}$ , terms higher than linear in the expansion can be discarded. We want to minimize the energy gradient of Eq. [132] as

$$\frac{d\tilde{E}(\mathbf{P} + \mathbf{P}_\Delta)}{d\mathbf{P}} \stackrel{!}{=} 0 \quad [133]$$

Neglecting all terms higher than linear and differentiating Eq. [132], we immediately arrive at the Newton–Raphson equation, which has to be solved iteratively to obtain the density update  $\mathbf{P}_\Delta$ :

$$\frac{d^2\tilde{E}}{d\mathbf{P}^2}(\mathbf{P}_\Delta) = -\frac{d\tilde{E}}{d\mathbf{P}} \quad [134]$$

The term on the right-hand side of Eq. [134] is already known from the simple energy gradient (Eq.[117]) and the left-hand side can be calculated as

$$\begin{aligned} \frac{\partial}{\partial \mathbf{P}} \text{tr} \left( \frac{\partial \tilde{E}(\mathbf{P})}{\partial \mathbf{P}} \mathbf{P}_\Delta \right) &= 3\mathbf{F}\mathbf{P}_\Delta\mathbf{S} + 3\mathbf{S}\mathbf{P}_\Delta\mathbf{F} - 2\mathbf{F}\mathbf{P}_\Delta\mathbf{S}\mathbf{P}\mathbf{S} - 2\mathbf{F}\mathbf{P}\mathbf{S}\mathbf{P}_\Delta\mathbf{S} \\ &\quad - 2\mathbf{S}\mathbf{P}_\Delta\mathbf{F}\mathbf{P}\mathbf{S} - 2\mathbf{S}\mathbf{P}\mathbf{F}\mathbf{P}_\Delta\mathbf{S} - 2\mathbf{S}\mathbf{P}_\Delta\mathbf{S}\mathbf{P}\mathbf{F} - 2\mathbf{S}\mathbf{P}\mathbf{S}\mathbf{P}_\Delta\mathbf{F} \\ &\quad + 3\mathbf{G}(\mathbf{X})\mathbf{P}\mathbf{S} + 3\mathbf{S}\mathbf{P}\mathbf{G}(\mathbf{X}) - 2\mathbf{G}(\mathbf{X})\mathbf{P}\mathbf{S}\mathbf{P}\mathbf{S} \\ &\quad - 2\mathbf{S}\mathbf{P}\mathbf{S}\mathbf{P}\mathbf{G}(\mathbf{X}) - 2\mathbf{S}\mathbf{P}\mathbf{G}(\mathbf{X})\mathbf{P}\mathbf{S} \end{aligned} \quad [135]$$

with

$$\mathbf{X} = 3\mathbf{P}_\Delta\mathbf{S}\mathbf{P} + 3\mathbf{P}\mathbf{S}\mathbf{P}_\Delta - 2\mathbf{P}_\Delta\mathbf{S}\mathbf{P}\mathbf{S}\mathbf{P} - 2\mathbf{P}\mathbf{S}\mathbf{P}_\Delta\mathbf{S}\mathbf{P} - 2\mathbf{P}\mathbf{S}\mathbf{P}\mathbf{S}\mathbf{P}_\Delta \quad [136]$$

After the density update has been determined using the Newton–Raphson equations shown above, the density matrix may be updated as

$$\mathbf{P}^{(i+1)} = \mathbf{P}^{(i)} + s\mathbf{P}_\Delta^{(i)} \quad [137]$$

The procedure of determining  $\mathbf{P}_\Delta$  and updating the density is iterated until self-consistency is obtained. For molecules with a nonvanishing HOMO–LUMO gap, all matrices involved are sparse, such that solving the SCF eigenvalue problem is altogether an asymptotically linear-scaling step.

## Implications for Linear-Scaling Calculation of SCF Energies

This concludes our derivation of tools for the linear-scaling calculation of SCF energies. We have outlined methods that enable the linear-scaling execution of the two expensive steps of SCF calculations: first, efficient integral screening and linear-scaling formation of Fock-type matrices, and second, methods for circumventing the diagonalization step, used conventionally for

solving the SCF pseudo-eigenvalue problem. With these methods, it is now possible to calculate HF and DFT energies with an effort scaling asymptotically linear, so that molecular systems with 1000 and more atoms can be handled with today's computers.

---

## SCF ENERGY GRADIENTS

Up to this stage of our review, we were mainly focusing on the computation of the energy of a molecule. However, to obtain suitable instruments for studying molecular systems and to be able to establish useful connections to experimental investigations, we also need to compute other molecular properties than just the energy. The first step toward this is the calculation of energy gradients, e.g., with respect to nuclear coordinates, which allow us to locate stationary points on a potential energy surface. In addition, energy gradients are crucial for performing direct Born–Oppenheimer molecular dynamics.

The energy gradients with respect to a nuclear coordinate, as an example, can be obtained by differentiating the SCF energy expression of Eq. [109]:

$$\frac{\partial E}{\partial \mathbf{x}} = \text{tr}(\mathbf{P}\mathbf{h}^{\mathbf{x}}) + \frac{1}{2} \text{tr}(\mathbf{P}\mathbf{G}^{\mathbf{x}}(\mathbf{P})) + \text{tr}(\mathbf{P}^{\mathbf{x}}\mathbf{F}) \quad [138]$$

where  $\mathbf{G}^{\mathbf{x}}(\mathbf{P})$  denotes the contraction of derivative two-electron integrals with the one-particle density matrix and  $\mathbf{h}^{\mathbf{x}}$  the derivative of the core-Hamiltonian matrix. Note that the computation of these integral derivatives can be done in a linear-scaling fashion by slight modifications of the previously introduced  $\mathcal{O}(M)$  algorithms like CFMM and LinK, for example. Although the derivative density matrix  $\mathbf{P}^{\mathbf{x}}$  occurs, Pulay pointed out<sup>125–127</sup> that it can be avoided by exploiting the solved Roothaan–Hall equations and the derivative of the orthonormality relation. In this way, the perturbed density matrix is replaced by an expression containing the overlap integral derivative  $\mathbf{S}^{\mathbf{x}}$  as

$$\frac{\partial E}{\partial \mathbf{x}} = \text{tr}(\mathbf{P}\mathbf{h}^{\mathbf{x}}) + \frac{1}{2} \text{tr}(\mathbf{P}\mathbf{G}^{\mathbf{x}}(\mathbf{P})) - \text{tr}(\mathbf{W}\mathbf{S}^{\mathbf{x}}) \quad [139]$$

where  $\mathbf{W}$  is the “energy-weighted density matrix” expressed as

$$W_{\mu\nu} = \sum_i \varepsilon_i C_{\mu i}^* C_{\nu i} \quad [140]$$

This formulation requires one to compute the energy-weighted density matrix using the molecular orbital coefficient matrix  $\mathbf{C}$ , which must be avoided to

achieve an overall linear-scaling behavior. Therefore, we need to derive an alternative expression for substituting  $\text{tr}(\mathbf{P}^x \mathbf{F})$  (see also Ref. 128).

To obtain equations that are independent of  $\mathbf{P}^x$ , it is necessary to consider the different contributions to the derivative density. As for any matrix representation of operators, it is possible to split the contributions into different subspace projections (compare Eq. [108]):

$$\mathbf{P}^x = \mathbf{P}_{\text{oo}}^x + \mathbf{P}_{\text{ov}}^x + \mathbf{P}_{\text{vo}}^x + \mathbf{P}_{\text{vv}}^x \quad [141]$$

These projections will be analyzed below. At SCF convergence, the following equations are valid:

$$\mathbf{FPS} = \mathbf{SPF} \quad [142]$$

$$\mathbf{P} = \mathbf{PSP} \quad [143]$$

In addition, after introducing the perturbation “x,” the derivative of the idempotency relation (Eq. [143]) has to be fulfilled:

$$\mathbf{P}^x = \mathbf{P}^x \mathbf{SP} + \mathbf{P} \mathbf{S}^x \mathbf{P} + \mathbf{PS} \mathbf{P}^x \quad [144]$$

Projecting Eq. [144] onto the occupied space and employing the idempotency relation (Eq. [143]) allows us to identify  $\mathbf{P}_{\text{oo}}^x$ :

$$\mathbf{P}_{\text{oo}}^x = \mathbf{PSP}^x \mathbf{SP} = \mathbf{PSP}^x \mathbf{SP} + \mathbf{PS}^x \mathbf{P} + \mathbf{PSP}^x \mathbf{SP} = -\mathbf{PS}^x \mathbf{P} \quad [145]$$

This shows that the occupied–occupied part of  $\mathbf{P}^x$  is directly linked to the derivative of the overlap integrals. In addition, the virtual–virtual block of  $\mathbf{P}^x$  vanishes:

$$\mathbf{P}_{\text{vv}}^x = (\mathbf{1} - \mathbf{PS})\mathbf{P}^x(\mathbf{1} - \mathbf{SP}) = \underbrace{\mathbf{P}^x - \mathbf{PSP}^x - \mathbf{P}^x \mathbf{SP}}_{\mathbf{PS}^x \mathbf{P}} + \underbrace{\mathbf{PSP}^x \mathbf{SP}}_{-\mathbf{PS}^x \mathbf{P}} = 0 \quad [146]$$

With these properties of  $\mathbf{P}^x$  at hand, the remaining part of  $\text{tr}(\mathbf{P}^x \mathbf{F})$  is the occ/virt and the virt/occ blocks:

$$\text{tr}(\mathbf{P}_{\text{ov}}^x \mathbf{F}) = \text{tr}(\mathbf{PSP}^x(\mathbf{1} - \mathbf{SP})\mathbf{F}) = \text{tr}(\mathbf{P}^x(\mathbf{1} - \mathbf{SP})\mathbf{FPS}) = \text{tr}(\mathbf{P}^x \mathbf{F}_{\text{vo}}) \quad [147]$$

$$\text{tr}(\mathbf{P}_{\text{vo}}^x \mathbf{F}) = \text{tr}(\mathbf{P}^x \mathbf{F}_{\text{ov}}) \quad [148]$$

In Eqs. [147] and [148], the cyclic permutation possible within a trace has been exploited, which shows that the projection of  $\mathbf{P}^x$  can be transferred to the Fock matrix. At this stage it is important to think again on what is done

in solving the SCF equations at the HF or DFT level: The Fock matrix is diagonalized, so that blocks coupling virtual and occupied parts vanish:

$$\mathbf{F}_{ov} = \mathbf{SPF}(\mathbf{1} - \mathbf{PS}) = \mathbf{SPF} - \mathbf{SP} \underbrace{\mathbf{FPS}}_{\mathbf{SPF}} = \mathbf{SPF} - \mathbf{S} \underbrace{\mathbf{PSP}}_{\mathbf{P}} \mathbf{F} = \mathbf{0} \quad [149]$$

$$\mathbf{F}_{vo} = \mathbf{0} \quad [150]$$

Therefore, the term involved in the energy gradient simplifies to

$$\text{tr}(\mathbf{P}^x \mathbf{F}) = \text{tr}(\mathbf{P}_{oo}^x \mathbf{F}) = \text{tr}(-\mathbf{PS}^x \mathbf{PF}) = \text{tr}(-\mathbf{PFPS}^x) \quad [151]$$

In this way the final density matrix-based energy gradient expression results, which avoids the use of the conventional energy-weighted density matrix (which is conventionally constructed via the delocalized, canonical MO coefficients):

$$\frac{\partial E}{\partial x} = \text{tr}(\mathbf{P} \mathbf{h}^x) + \frac{1}{2} \text{tr}(\mathbf{P} \mathbf{G}^x(\mathbf{P})) - \text{tr}(\mathbf{PFPS}^x) \quad [152]$$

Here, only quantities that are sparse for systems with a nonvanishing HOMO–LUMO gap enter the formulation, whereas the derivative quantities  $\mathbf{h}^x$ ,  $\mathbf{S}^x$ , and the contraction of the one-particle density matrix with the derivative integrals can be computed in a linear-scaling fashion. For the integral contractions slightly modified CFMM- and LinK-type schemes can be used to reduce the scaling.<sup>76,129</sup> It is important to note that, for example, in the formation of the exchange energy gradient  $\mathbf{EK}^x$ , the derivative two-electron integrals are contracted with *two* one-particle density matrices:

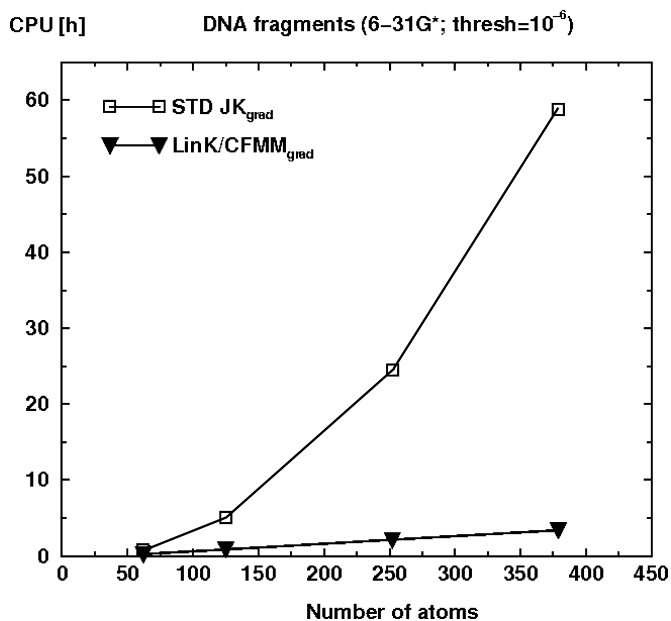
$$\mathbf{EK}^x = \sum_{\mu\nu} \sum_{\lambda\sigma} \mathbf{P}_{\mu\nu} (\mu\lambda|\nu\sigma)^x \mathbf{P}_{\lambda\sigma} \quad [153]$$

Therefore, the coupling in reducing the scaling behavior is even stronger than for the calculation of the Fock matrix. An example for the scaling behavior of the energy-gradient calculation as compared with the conventional quadratic scaling behavior is displayed in Figure 18 for DNA fragments. It is important to note that the numerical accuracy is the same in both cases.

---

## MOLECULAR RESPONSE PROPERTIES AT THE SCF LEVEL

With the energy gradients at hand, one can now locate stationary points on a potential hypersurface or study the dynamics of a system. However, a huge class of important molecular properties is more complicated to calculate,



**Figure 18** Comparison of timings for standard  $\mathcal{O}(M^2)$  (STD JK<sub>grad</sub>) and linear-scaling energy gradients for DNA fragments (HF/6-31G\*) using the LinK<sup>76</sup> and CFMM<sup>129</sup> methods.

because they are linked to the response of the molecular system with respect to a perturbation. Examples include vibrational frequencies, NMR chemical shifts, and polarizabilities. An excellent review is given in Ref. 130, so that we focus in the following on the key issues that need to be resolved to reduce the scaling behavior for the computation of response properties to being linear at the SCF level.

In the following, we first describe briefly how properties such as vibrational frequencies and NMR chemical shifts are computed. Then we focus on how to calculate the common difficult part, namely the response of the one-particle density matrix with respect to a perturbation and how to reduce the computational scaling to linear.

## Vibrational Frequencies

The second derivatives with respect to nuclear displacements are crucial for characterizing stationary points on a potential hypersurface. They provide as well the normal modes of the system and can be linked within the harmonic approximation to the vibrational frequencies of the system, which can be measured experimentally by IR or Raman spectroscopy. By taking the derivative of the SCF energy gradient expression (Eq. [152]) with respect to another

nuclear coordinate  $y$ , we obtain the following expression for the second derivatives:

$$\begin{aligned} \frac{\partial^2 E}{\partial x \partial y} = & \operatorname{tr} \left( \mathbf{P} \frac{\partial^2 \mathbf{h}}{\partial x \partial y} \right) + \frac{1}{2} \operatorname{tr} \left( \mathbf{P} \frac{\partial^2 \mathbf{\Pi}}{\partial x \partial y} \mathbf{P} \right) - \operatorname{tr} \left( \mathbf{P} \mathbf{F} \mathbf{P} \frac{\partial^2 \mathbf{S}}{\partial x \partial y} \right) + \operatorname{tr} \left( \mathbf{P} \frac{\partial \mathbf{F}}{\partial y} \mathbf{P} \frac{\partial \mathbf{S}}{\partial x} \right) \\ & + \operatorname{tr} \left( \frac{\partial \mathbf{P}}{\partial y} \frac{\partial \mathbf{h}}{\partial x} \right) + \operatorname{tr} \left( \frac{\partial \mathbf{P}}{\partial y} \frac{\partial \mathbf{\Pi}}{\partial x} \mathbf{P} \right) - \operatorname{tr} \left( \frac{\partial \mathbf{P}}{\partial y} \mathbf{F} \mathbf{P} \frac{\partial \mathbf{S}}{\partial x} + \mathbf{P} \mathbf{F} \frac{\partial \mathbf{P}}{\partial y} \frac{\partial \mathbf{S}}{\partial x} \right) \end{aligned} \quad [154]$$

where  $\mathbf{\Pi}$  abbreviates the antisymmetrized two-electron integrals. In contrast to the simple energy gradient expression, the computation of the perturbed one-particle density matrix cannot be avoided anymore for the second derivatives. To obtain this response of the one-particle density matrix with respect to the perturbation  $y$ , the coupled-perturbed Hartree–Fock or coupled-perturbed Kohn–Sham equations need to be solved.<sup>131–136</sup> The standard path for a solution in the MO basis scales as  $\mathcal{O}(M^5)$ ,<sup>135,136</sup> whereas an AO formulation introduced later<sup>128,136</sup> reduces the computational effort to  $\mathcal{O}(M^4)$ . The scaling behavior of the latter scheme is due to a partial solution of the coupled-perturbed self-consistent field (CPSCF) equations in the MO basis. To reduce this scaling behavior, Ochsenfeld and Head-Gordon<sup>68</sup> reformulated the CPSCF theory in a density matrix-based scheme (D-CPSCF), so that asymptotically a linear-scaling behavior becomes possible. Closely related to this density matrix-based approach, Ochsenfeld, Kussmann, and Koziol recently introduced an even more efficient density-based approach for the solution of the CPSCF equations in the context of the calculation of NMR chemical shieldings.<sup>113</sup> Therefore, we focus in the following on the calculation of NMR shieldings and formulate the corresponding D-CPSCF theory within this context.

## NMR Chemical Shieldings

The routine calculation of NMR chemical shifts<sup>137–141</sup> using quantum chemical methods has become possible since the introduction of local gauge-origin methods,<sup>142–148</sup> which provide a solution to the gauge-origin problem within approximated schemes. In our formulation, we use gauge-including atomic orbitals (GIAO):<sup>144–146</sup>

$$\chi_{\mu}(\mathbf{B}) = \chi_{\mu}(0) \exp \left( -\frac{i}{2c} \mathbf{B} \times (\mathbf{R}_{\mu} - \mathbf{R}_0) \cdot \mathbf{r} \right) \quad [155]$$

where  $\mathbf{B}$  is the magnetic field vector and  $\chi_{\mu}(0)$  denotes the standard field-independent basis functions. The location of the basis functions and the gauge origin are described by  $\mathbf{R}_{\mu}$  and  $\mathbf{R}_0$ , respectively. The use of the GIAO functions permits us to avoid the gauge-origin problem and has proven to be particularly successful.<sup>137</sup> In the following, we constrain ourselves to the HF method (GIAO-HF),<sup>145,146,148</sup> which provides useful results for many molecular systems.

For example, in many systems, we found the GIAO-HF method to yield  $^1\text{H}$ -NMR chemical shifts with an accuracy of typically 0.2–0.5 ppm.<sup>149–152</sup> For other nuclei the inclusion of correlation effects can become more important.<sup>137,139, 153,154</sup> The computation at the GIAO-DFT level is closely related.

NMR chemical shifts are calculated as second derivatives of the energy with respect to the external magnetic field  $\mathbf{B}$  and the nuclear magnetic spin  $m_{N_i}$  of a nucleus  $N$ :

$$\sigma_{ij}^N = \frac{\partial^2 E}{\partial B_i \partial m_{N_j}} \quad [156]$$

where  $i, j$  are  $x, y, z$  coordinates. This leads to:

$$\sigma_{ij}^N = \sum_{\mu\nu} P_{\mu\nu} \frac{\partial^2 h_{\mu\nu}}{\partial B_i \partial m_{N_j}} + \sum_{\mu\nu} \frac{\partial P_{\mu\nu}}{\partial B_i} \frac{\partial h_{\mu\nu}}{\partial m_{N_j}} \quad [157]$$

The equation shows that, similar to the calculation of vibrational frequencies, the response of the one-particle density matrix to a perturbation is necessary, which is in the case of NMR shieldings the magnetic field  $B_i$ . Therefore, the CPSCF equations need to be solved for the perturbed one-particle density matrices  $\frac{\partial P_{\mu\nu}}{\partial B_i}$  (short:  $\mathbf{P}^{B_i}$ ). In the context of NMR shieldings, the computational effort of conventional schemes<sup>146,148</sup> scales cubically with molecular size.

To reduce the scaling behavior for the calculation of response properties, we focus in the following on a reformulation of the CPSCF equations in a density matrix-based scheme, so that the scaling of the computational effort can be reduced to linear for systems with a nonvanishing HOMO–LUMO gap.

### Density Matrix-Based Coupled Perturbed SCF (D-CPSCF)

The solution of the CPSCF equations is necessary for obtaining the response of the one-particle density matrix with respect to a perturbation. As mentioned, in conventional formulations of CPSCF theory, AO–MO transformations are involved, so that again the delocalized, canonical MO coefficients are required. In this way, it is not possible to reduce the computational effort to linear. Therefore, the key feature of linear-scaling CPSCF theory is to avoid these MO transformations and, instead, to solve directly for the perturbed one-particle density matrices. The quadratically convergent density matrix-based SCF method described above in the context of avoiding the diagonalization within the SCF cycle can be used as the basis for the reformulation of the response equations.<sup>68</sup> Related alternative approaches have been later proposed in Refs. 155 and 156, but we follow in this review our derivation presented in Refs. 68 and 113, which we have found to be useful in obtaining an efficient density matrix-based CPSCF scheme for large molecules.

In the following, we focus on the determination of the density matrix as perturbed with respect to the magnetic field ( $\mathbf{P}^{B_i}$ ), whereas an extension to other perturbations is straightforward. Within a linear response formalism (only terms linear in the external perturbation are considered), we can solve for the perturbed density matrix  $\mathbf{P}^{B_i}$  directly

$$\left[ \frac{\partial^2 \tilde{E}[\mathbf{P}]}{\partial \mathbf{P}^2} \right] \mathbf{P}^{B_i} = - \frac{\partial}{\partial B_i} \left[ \frac{\partial \tilde{E}[\mathbf{P}]}{\partial \mathbf{P}} \right] \quad [158]$$

where  $\tilde{E}$  is the functional described in Eq. [115]. Inserting Eq. [135] and the derivative of Eq. [117] with respect to the perturbation  $B_i$  into Eq. [158], we obtain<sup>68</sup>

$$\begin{aligned} & 3\mathbf{F}\mathbf{P}^{B_i}\mathbf{S} + 3\mathbf{S}\mathbf{P}^{B_i}\mathbf{F} - 2\mathbf{F}\mathbf{P}^{B_i}\mathbf{S}\mathbf{P} - 4\mathbf{F}\mathbf{P}\mathbf{S}\mathbf{P}^{B_i}\mathbf{S} - 4\mathbf{S}\mathbf{P}^{B_i}\mathbf{S}\mathbf{P}\mathbf{F} - 2\mathbf{S}\mathbf{P}\mathbf{S}\mathbf{P}^{B_i}\mathbf{F} \\ & \quad + \mathbf{G}(\mathbf{X})\mathbf{P}\mathbf{S} + \mathbf{S}\mathbf{P}\mathbf{G}(\mathbf{X}) - 2\mathbf{S}\mathbf{P}\mathbf{G}(\mathbf{X})\mathbf{P}\mathbf{S} \\ & = \mathbf{F}\mathbf{P}\mathbf{S}^{B_i} + \mathbf{S}^{B_i}\mathbf{P}\mathbf{F} + 2\mathbf{F}\mathbf{P}\mathbf{S}^{B_i}\mathbf{P}\mathbf{S} + 2\mathbf{S}\mathbf{P}\mathbf{S}^{B_i}\mathbf{P}\mathbf{F} \\ & \quad - \mathbf{F}^{(B_i)}\mathbf{P}\mathbf{S} - \mathbf{S}\mathbf{P}\mathbf{F}^{(B_i)} + 2\mathbf{S}\mathbf{P}\mathbf{F}^{(B_i)}\mathbf{P}\mathbf{S} \end{aligned} \quad [159]$$

with

$$\mathbf{F}^{(B_i)} = \mathbf{h}^{B_i} + \mathbf{G}^{B_i}(\mathbf{P}) \quad [160]$$

$$\mathbf{X} = \mathbf{P}^{B_i}\mathbf{S}\mathbf{P} + \mathbf{P}\mathbf{S}\mathbf{P}^{B_i} - 2\mathbf{P}\mathbf{S}\mathbf{P}^{B_i}\mathbf{S}\mathbf{P} - \mathbf{P}\mathbf{S}^{B_i}\mathbf{P} = \mathbf{P}^{B_i} \quad [161]$$

At this stage, it is worthwhile to consider some properties of the derivative density matrix, which can—as described in the section on density matrix-based energy gradients—be split into a sum of subspace projections (Eq. [108]):

$$\mathbf{P}^{B_i} = \mathbf{P}_{oo}^{B_i} + \mathbf{P}_{ov}^{B_i} + \mathbf{P}_{vo}^{B_i} + \mathbf{P}_{vv}^{B_i} \quad [162]$$

The comparison with the derivative of the idempotency relation

$$\mathbf{P}^{B_i} = \mathbf{P}^{B_i}\mathbf{S}\mathbf{P} + \mathbf{P}\mathbf{S}^{B_i}\mathbf{P} + \mathbf{P}\mathbf{S}\mathbf{P}^{B_i} \quad [163]$$

clarifies the different contributions of  $\mathbf{P}^{B_i}$ :

$$\mathbf{P}_{oo}^{B_i} = -\mathbf{S}_{oo}^{B_i} \quad [164]$$

$$\mathbf{P}_{ov}^{B_i} = -\mathbf{P}_{vo}^{B_i} \quad [165]$$

$$\mathbf{P}_{vv}^{B_i} = \mathbf{0} \quad [166]$$

where the sign on the right of Eq. [164] originates from the fact that the first-order matrices with respect to the magnetic perturbation are skew-symmetric. As we can directly calculate  $S^{B_i}$ , we only have to determine the occupied–virtual part  $P_{ov}^{B_i}$ . To solve only for  $P_{ov}^{B_i}$  and  $P_{vo}^{B_i}$ , the equation system of Eq. [159] can be projected by  $SP$  from the left and  $PS$  from the right, respectively, and the two resulting equations are added. In this way, we obtain the following density matrix-based CPSCF equations,<sup>113</sup> which provide superior convergence properties, in particular if sparse algebra is employed:

$$\begin{aligned} &FP^{B_i}SPS + SPSP^{B_i}F - FPSP^{B_i}S - SP^{B_i}SPF + G(P^{B_i})PS + SPG(P^{B_i}) - 2SPG(P^{B_i})PS \\ &= FPS^{B_i} + S^{B_i}PF - FPS^{B_i}PS - SPS^{B_i}PF - F^{(B_i)}PS - SPF^{(B_i)} + 2SPF^{(B_i)}PS \end{aligned} \quad [167]$$

with

$$F_{\mu\nu}^{(B_i)} = \frac{\partial h_{\mu\nu}}{\partial B_i} + \sum_{\lambda\sigma} P_{\lambda\sigma} \frac{\partial[(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\lambda|\nu\sigma)]}{\partial B_i} \quad [168]$$

$$G_{\mu\nu}(P_{\lambda\sigma}^{B_i}) = -\frac{1}{2} \sum_{\lambda\sigma} \frac{\partial P_{\lambda\sigma}}{\partial B_i} (\mu\lambda|\nu\sigma) \quad [169]$$

The convergence properties of the density matrix-based equations, i.e., the number of iterations to converge  $P^{B_i}$ , are similar to the ones encountered for a solution in the MO space, so that the advantage of using sparse multiplications within the density-based approach allows us to reduce the scaling property of the computational effort in an efficient manner. In this way, NMR chemical shift calculations with linear-scaling effort become possible and systems with 1000 and more atoms can be treated at the HF or DFT level on today’s computers.<sup>113</sup> Extensions to other molecular properties can be formulated in a similar fashion.

---

## OUTLOOK ON ELECTRON CORRELATION METHODS FOR LARGE SYSTEMS

Although the main focus of the current review is to provide insights into reducing the scaling behavior of HF and DFT methods, it seems appropriate to provide a brief outlook on the behavior of post-HF methods. The importance of these methods cannot be overemphasized, because it is the systematic hierarchy of approaches to the exact solution of the electronic Schrödinger equation that allows for systematic and reliable studies of molecular systems. A concise overview of the huge amount of interesting and successful work done in the field of reducing the scaling behavior of post-HF methods is beyond the scope of this chapter; therefore,

we just provide some insights into why a reduction of the scaling behavior with rigorous error bounds should also be possible here. For a firsthand account of the impressive progress made in the field, the reader is referred to the work of Pulay and Saebø,<sup>157–160</sup> Werner and coworkers,<sup>161–163</sup> Head-Gordon and coworkers,<sup>164,165</sup> Almlöf and Häser,<sup>166–168</sup> Ayala and Scuseria,<sup>169,170</sup> Friesner and coworkers,<sup>171</sup> Carter and coworkers,<sup>172</sup> Schütz and Werner,<sup>173</sup> and Schütz.<sup>174</sup> Recent reviews may be found in Refs. 175 and 114.

To explain some principles, we focus here on the most simple of these approaches, the Møller–Plesset perturbation theory to second order (MP2). In the conventional, canonical MO-based formulation, the closed-shell MP2 correlation energy is given by

$$E_{\text{MP2}} = - \sum_{ijab} \frac{(ia|jb)[2(ia|jb) - (ib|ja)]}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j} \quad [170]$$

with the MO integrals

$$(ia|jb) = \int \varphi_i^*(\mathbf{r}_1)\varphi_a(\mathbf{r}_1) \frac{1}{r_{12}} \varphi_j^*(\mathbf{r}_2)\varphi_b(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad [171]$$

Indices  $i, j$  denote occupied orbitals, whereas  $a, b$  are virtuals. The difficulty is that the integrals computed in the AO basis need to be transformed into the MO basis:

$$(ia|jb) = \sum_{\mu\nu\lambda\sigma} C_{\mu i} C_{\nu a} C_{\lambda j} C_{\sigma b} \quad (\mu\nu|\lambda\sigma) \quad [172]$$

If the transformations are done in a successive instead of a simultaneous way, the computational effort reduces from formally  $\mathcal{O}(M^8)$  to  $\mathcal{O}(M^5)$ . However, due to the nonlocality of the canonical MOs (discussed above in the context of SCF methods), this  $\mathcal{O}(M^5)$  effort holds in the asymptotic limit, so that no reduction can be expected. The only factor is that the four transformations scale differently depending on whether occupied or virtual indices are transformed.

To avoid the canonical, delocalized orbitals, Almlöf suggested in 1991<sup>166</sup> using a Laplace transform for eliminating the disturbing denominator  $x_q \equiv \varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j$ :

$$\frac{1}{x_q} = \int_0^\infty \exp(-x_q t) dt \approx \sum_{\alpha=1}^{\tau} \omega^{(\alpha)} \exp(-x_q t^{(\alpha)}) \quad [173]$$

where the integral can be replaced by a summation over a few grid points. In typical applications, it has been found by Häser and Almlöf<sup>167</sup> that  $\tau = 5 - 8$

provides  $\mu$ Hartree accuracy. This approach was employed by Häser to formulate an AO-MP2 formalism,<sup>168</sup> which we briefly revise in the following: With the definition of two pseudo-density matrices,

$$\underline{\mathbf{P}}_{\mu\nu}^{(\alpha)} = |\omega^{(\alpha)}|^{1/4} \sum_i^{\text{occ}} C_{\mu i} \exp((\varepsilon_i - \varepsilon_F)t^{(\alpha)}) C_{\nu i} \quad [174]$$

and

$$\bar{\mathbf{P}}_{\mu\nu}^{(\alpha)} = |\omega^{(\alpha)}|^{1/4} \sum_a^{\text{virt}} C_{\mu a} \exp((\varepsilon_F - \varepsilon_a)t^{(\alpha)}) C_{\nu a} \quad [175]$$

where  $\varepsilon_F$  is  $(\varepsilon_{\text{HOMO}} + \varepsilon_{\text{LUMO}})/2$ ,<sup>166–168</sup> the MP2 energy expression becomes

$$E_{\text{MP2}} = - \sum_{\alpha=1}^{\tau} e_{JK}^{(\alpha)} \quad [176]$$

with

$$e_{JK}^{(\alpha)} = \sum_{\mu\nu\lambda\sigma} \sum_{\mu'\nu'\lambda'\sigma'} \underline{\mathbf{P}}_{\mu\mu'}^{(\alpha)} \bar{\mathbf{P}}_{\nu\nu'}^{(\alpha)} (\mu'\nu'|\lambda'\sigma') \underline{\mathbf{P}}_{\lambda\lambda'}^{(\alpha)} \bar{\mathbf{P}}_{\sigma\sigma'}^{(\alpha)} [2(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu)] \quad [177]$$

For each integration point ( $\alpha = 1 \dots \tau$ ), four, formally  $\mathcal{O}(M^5)$  scaling transformations are necessary to yield the transformed two-electron integrals

$$(\underline{\mu}\bar{\nu}|\underline{\lambda}\bar{\sigma})^{(\alpha)} = \sum_{\mu'\nu'\lambda'\sigma'} \underline{\mathbf{P}}_{\mu\mu'}^{(\alpha)} \bar{\mathbf{P}}_{\nu\nu'}^{(\alpha)} (\mu'\nu'|\lambda'\sigma') \underline{\mathbf{P}}_{\lambda\lambda'}^{(\alpha)} \bar{\mathbf{P}}_{\sigma\sigma'}^{(\alpha)} \quad [178]$$

which are contracted in a final, formally  $\mathcal{O}(M^4)$  scaling step in a Coulomb- ( $e_J^{(\alpha)}$ ) and an exchange-type ( $e_K^{(\alpha)}$ ) fashion:

$$e_{JK}^{(\alpha)} = 2e_J^{(\alpha)} - e_K^{(\alpha)} = \sum_{\mu\nu\lambda\sigma} (\underline{\mu}\bar{\nu}|\underline{\lambda}\bar{\sigma})^{(\alpha)} [2(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu)] \quad [179]$$

Here  $\mu$  and  $\bar{\mu}$  denote the same index, where the bar of  $\mu$  (or  $\bar{\mu}$ ) only indicates that the index has been transformed with  $\underline{\mathbf{P}}$  (or  $\bar{\mathbf{P}}$ , respectively).

In contrast to the conventional MO-based formulation, the AO-based Laplace formalism allows one to reduce the conventional  $\mathcal{O}(N^5)$  scaling of the computational cost for large molecules. However, for small molecules, the overhead consists of the need to compute  $\tau = 5\text{--}8$  exponentials and the larger prefactor for the transformations scaling formally as  $N^5$  compared with the  $n_{\text{occ}} \cdot N^4$ ,  $\dots$ , and  $n_{\text{occ}}^2 \cdot n_{\text{virt}}^2 \cdot N$  scaling for the different MO-based

transformations ( $n_{\text{occ}}$  and  $n_{\text{virt}}$  denote the number of occupied and virtual orbitals, respectively). Despite this overhead for small molecules, the central drawback in MO-based transformations caused by the delocalized nature of canonical MOs is avoided and the scaling can be reduced for large molecules.

The AO–MP2 method introduced in 1993 by Häser<sup>168</sup> applies screening criteria to the intermediate four-index quantities in order to reduce the computational scaling for larger molecules. Here, the Schwarz inequality introduced earlier in this review<sup>21,168,177</sup>

$$|(\mu\nu|\lambda\sigma)| \leq |(\mu\nu|\mu\nu)|^{\frac{1}{2}} |(\lambda\sigma|\lambda\sigma)|^{\frac{1}{2}} = Q_{\mu\nu} Q_{\lambda\sigma} \quad [180]$$

is used, which we denote as QQ-screening. Häser<sup>168</sup> adapted the Schwarz screening for estimating the transformed quantities occurring in AO–MP2 theory, which we abbreviate in the following as QQZZ or Pseudo-Schwarz screening, where  $Z$  is defined as an upper bound approximation to the transformed Schwarz criterium (see Ref. 168):

$$|(\underline{\mu}\bar{\nu}|\underline{\mu}\bar{\nu})|^{\frac{1}{2}} \leq Z_{\mu\nu} \quad [181]$$

As pointed out by Häser,<sup>168</sup> this screening protocol yields asymptotically a quadratically scaling MP2 method for systems with a significant HOMO–LUMO gap. This quadratic scaling of AO–MP2 was further reduced to become linear by Ayala and Scuseria by “*introducing interaction domains and neglecting selective domain-domain interactions*”.<sup>169</sup>

In this tutorial, we use the Laplace approach to explain some aspects of the long-range behavior of electron-correlation methods, without commenting on which one of the many approaches for reducing the computational effort will become the standard replacement of conventional correlation formulations. We follow here our discussion presented in a recent publication,<sup>24</sup> which permits for the first time to determine rigorously which of the transformed integral products contribute to the MP2 energy.

## Long-Range Behavior of Correlation Effects

The formation of the correlation energy in AO–MP2 consists of the transformation (Eq. [178]) and the contraction step (Eq. [179]). We start our discussion by considering the distance dependence of correlation contributions.

### *Transformed Integrals*

Some discussion in this section is similar to considerations of Ayala and Scuseria.<sup>169</sup> However, we present here a different argument for deriving rigorous and tight upper bounds for estimating transformed integral products following the work in Ref. 24.

For nonoverlapping charge distributions  $\Omega_A = \Omega_{\mu\nu} = \chi_\mu\chi_\nu$  and  $\Omega_B = \Omega_{\lambda\sigma}$  centered at **A** and **B**, respectively, the two-electron integral  $(\mu\nu|\lambda\sigma)$  is bound from above (see Refs. 23 and 31) by

$$\left\langle \frac{1}{r_{12}} \right\rangle \leq \frac{1}{R} \left| \sum_{n=0}^{\infty} \frac{\langle (\mathbf{r}_{1A} - \mathbf{r}_{2B})^n \rangle}{R^n} \right| \quad [182]$$

with  $R = |\mathbf{B} - \mathbf{A}|$  and the position of the electrons  $\mathbf{r}_1 = \mathbf{r}_{1A} + \mathbf{A}$  and  $\mathbf{r}_2 = \mathbf{r}_{2B} + \mathbf{B}$ , whereas  $\langle \rangle$  abbreviates the two-electron integral. This expansion in multipoles such as overlap  $M^{(0)} = S$ , dipole  $M^{(1)}$ , and higher order terms  $M^{(2)}$ ,  $M^{(3)}$ , ... leads to

$$\begin{aligned} |(\mu\nu|\lambda\sigma)| &\leq R^{-1} \left| M_{\mu\nu}^{(0)} M_{\lambda\sigma}^{(0)} \right| \\ &\quad + R^{-2} \left| M_{\mu\nu}^{(1)} M_{\lambda\sigma}^{(0)} - M_{\mu\nu}^{(0)} M_{\lambda\sigma}^{(1)} \right| \\ &\quad + R^{-3} \left| M_{\mu\nu}^{(2)} M_{\lambda\sigma}^{(0)} - 2M_{\mu\nu}^{(1)} M_{\lambda\sigma}^{(1)} + M_{\mu\nu}^{(0)} M_{\lambda\sigma}^{(2)} \right| \\ &\quad + R^{-4} \left| M_{\mu\nu}^{(3)} M_{\lambda\sigma}^{(0)} - 3M_{\mu\nu}^{(2)} M_{\lambda\sigma}^{(1)} + 3M_{\mu\nu}^{(1)} M_{\lambda\sigma}^{(2)} - M_{\mu\nu}^{(0)} M_{\lambda\sigma}^{(3)} \right| \\ &\quad + \mathcal{O}(R^{-5}) \end{aligned} \quad [183]$$

Due to the orthogonality properties of  $\underline{\mathbf{P}}$  and  $\bar{\mathbf{P}}$  (similar to the standard one-particle density matrix  $\mathbf{P}$  of SCF theory and its complement  $(1 - \mathbf{P})$  in an orthogonal basis), the transformation of the overlap leads to

$$\sum_{\mu'} \sum_{\nu'} \underline{P}_{\mu\mu'} S_{\mu'\nu'} \bar{P}_{\nu\nu'} = S_{\underline{\mu}\bar{\nu}} = M_{\underline{\mu}\bar{\nu}}^{(0)} = 0 \quad [184]$$

so that all terms involving the overlap ( $M^{(0)}$ ) are zero. Therefore, the expansion for the transformed integrals becomes

$$\begin{aligned} |(\underline{\mu}\bar{\nu}|\underline{\lambda}\bar{\sigma})| &\leq R^{-3} \left| -2M_{\underline{\mu}\bar{\nu}}^{(1)} M_{\underline{\lambda}\bar{\sigma}}^{(1)} \right| \\ &\quad + R^{-4} \left| -3M_{\underline{\mu}\bar{\nu}}^{(2)} M_{\underline{\lambda}\bar{\sigma}}^{(1)} + 3M_{\underline{\mu}\bar{\nu}}^{(1)} M_{\underline{\lambda}\bar{\sigma}}^{(2)} \right| \\ &\quad + \mathcal{O}(R^{-5}) \end{aligned} \quad [185]$$

and an  $\mathcal{O}(\frac{1}{R^3})$  dependence for the transformed integrals results. Together with the  $\mathcal{O}(\frac{1}{R})$  behavior of the untransformed integrals, this leads to an overall  $\mathcal{O}(\frac{1}{R^4})$  decay in the contraction step (Eq. [179]). It is important to note that this distance dependence results only from the orthogonality properties of the pseudo-density matrices, where the only requirement is the validity of

the multipole expansion (Eq. [182]) for the untransformed integrals. No locality of the pseudo-density matrices has been exploited at this stage, which leads to an even stronger decay as discussed below.

### Coulomb-Type Contraction

If the charge distributions  $\Omega_{\mu\nu}$ ,  $\Omega_{\lambda\sigma}$ , and  $\Omega_{\mu'\nu'}$ ,  $\Omega_{\lambda'\sigma'}$ , respectively, are non-overlapping in the sense that the multipole expansion (Eq. [183]) is applicable to the untransformed two-electron integrals  $(\mu\nu|\lambda\sigma)$  and  $(\mu'\nu'|\lambda'\sigma')$ , then the corresponding Coulomb terms can be written as

$$\begin{aligned}
 |e_j^{(\alpha)}| &\leq \sum_{\mu\nu\lambda\sigma} \sum_{\mu'\nu'\lambda'\sigma'} |(\mu\nu|\lambda\sigma)| \cdot \left| \underline{\mathbf{P}}_{\mu\mu'}^{(\alpha)} \bar{\mathbf{P}}_{\nu\nu'}^{(\alpha)}(\mu'\nu'|\lambda'\sigma') \underline{\mathbf{P}}_{\lambda\lambda'}^{(\alpha)} \bar{\mathbf{P}}_{\sigma\sigma'}^{(\alpha)} \right| \\
 &\leq \sum_{\mu\nu\lambda\sigma} R^{-1} \left| M_{\mu\nu}^{(0)} M_{\lambda\sigma}^{(0)} \right| \cdot R^{-3} \left| -2M_{\underline{\mu\nu}}^{(1)} M_{\underline{\lambda\sigma}}^{(1)} \right|^{(\alpha)} \\
 &\quad + R^{-1} \left| M_{\mu\nu}^{(0)} M_{\lambda\sigma}^{(0)} \right| \cdot R^{-4} \left| 3M_{\underline{\mu\nu}}^{(1)} M_{\underline{\lambda\sigma}}^{(2)} - 3M_{\underline{\mu\nu}}^{(2)} M_{\underline{\lambda\sigma}}^{(1)} \right|^{(\alpha)} \\
 &\quad + R^{-2} \left| M_{\mu\nu}^{(1)} M_{\lambda\sigma}^{(0)} - M_{\mu\nu}^{(0)} M_{\lambda\sigma}^{(1)} \right| \cdot R^{-3} \left| -2M_{\underline{\mu\nu}}^{(1)} M_{\underline{\lambda\sigma}}^{(1)} \right|^{(\alpha)} \\
 &\quad + R^{-2} \left| M_{\mu\nu}^{(1)} M_{\lambda\sigma}^{(0)} - M_{\mu\nu}^{(0)} M_{\lambda\sigma}^{(1)} \right| \cdot R^{-4} \left| 3M_{\underline{\mu\nu}}^{(1)} M_{\underline{\lambda\sigma}}^{(2)} - 3M_{\underline{\mu\nu}}^{(2)} M_{\underline{\lambda\sigma}}^{(1)} \right|^{(\alpha)} \\
 &\quad + \dots
 \end{aligned} \tag{186}$$

For the sake of notational simplicity, we have not made a distinction between distances of centers of untransformed or transformed charge distributions, because it is clear from the context. Considering in more detail the summation over the  $\mu\nu$ -part (the  $\lambda\sigma$  terms are omitted) of the first term of order  $\frac{1}{R^4}$ .

$$\sum_{\mu\nu} \sum_{\mu'\nu'} \frac{M_{\mu\nu}^{(0)}}{R} \cdot \left[ \underline{\mathbf{P}}_{\mu\mu'}^{(\alpha)} \frac{M_{\mu'\nu'}^{(1)}}{R^3} \bar{\mathbf{P}}_{\nu\nu'}^{(\alpha)} \right] = \sum_{\mu'\nu'} \sum_{\mu\nu} \left[ \underline{\mathbf{P}}_{\mu\mu'}^{(\alpha)} \frac{M_{\mu\nu}^{(0)}}{R} \bar{\mathbf{P}}_{\nu\nu'}^{(\alpha)} \right] \cdot \frac{M_{\mu'\nu'}^{(1)}}{R^3} \tag{187}$$

makes clear that we can either perform first the  $\mu', \nu'$  summation or the summation over  $\mu, \nu$ . In the second representation, the  $M^{(0)}$  term is multiplied by  $\underline{\mathbf{P}}$  (and  $\bar{\mathbf{P}}$ ), which would show that the  $\frac{1}{R^4}$  term becomes zero. However, this is only true if  $\underline{\mathbf{P}}$  and  $\bar{\mathbf{P}}$  are still fully orthogonal in the restricted space of indices where the multipole expansion is valid. Otherwise, missing indices would lead to nonvanishing overlap contributions after the projection with  $\underline{\mathbf{P}}$  and  $\bar{\mathbf{P}}$ . This leads to the following requirements:

- $\Omega_{\mu\nu}$  and  $\Omega_{\lambda\sigma}$  are nonoverlapping: valid multipole expansion for  $(\mu\nu|\lambda\sigma)$ .
- $\Omega_{\mu'\nu'}$  and  $\Omega_{\lambda'\sigma'}$  are nonoverlapping: valid multipole expansion for  $(\mu'\nu'|\lambda'\sigma')$ .

- For each  $\mu$  and  $\nu$  of  $\Omega_{\mu\nu}$ , the elements  $\mu'$  and  $\nu'$  coupled via the significant elements of  $\underline{P}_{\mu\mu'}$  and  $\underline{P}_{\nu\nu'}$  have to be contained in the sum  $\sum_{\mu'\nu'}$ . In other words: For each  $\mu$  and  $\nu$  of  $\Omega_{\mu\nu}$ , the shell pairs coupled via  $\underline{P}_{\mu\mu'}$  and  $\underline{P}_{\nu\nu'}$  (the significant elements) have to be nonoverlapping with  $\Omega_{\lambda'\sigma'}$ , so that the multipole expansion can be applied.

If these criteria are fulfilled in the restricted space of indices defined by the multipole expansion within a threshold, then  $(\underline{\mathbf{P}}\mathbf{M}^{(0)}\underline{\mathbf{P}})_{\mu'\nu'}$  is zero and with that the  $\frac{1}{R^4}$  term in Eqs. [186] and [187], and two of the three  $\frac{1}{R^3}$  terms in Eq. [186] disappear. If the analogous argumentation holds at the same time for  $\Omega_{\lambda\sigma}$ , the third  $\frac{1}{R^3}$  contribution is zero as well, so that an overall  $\frac{1}{R^6}$  dependence of the required transformed integrals for the Coulomb-type contraction results.

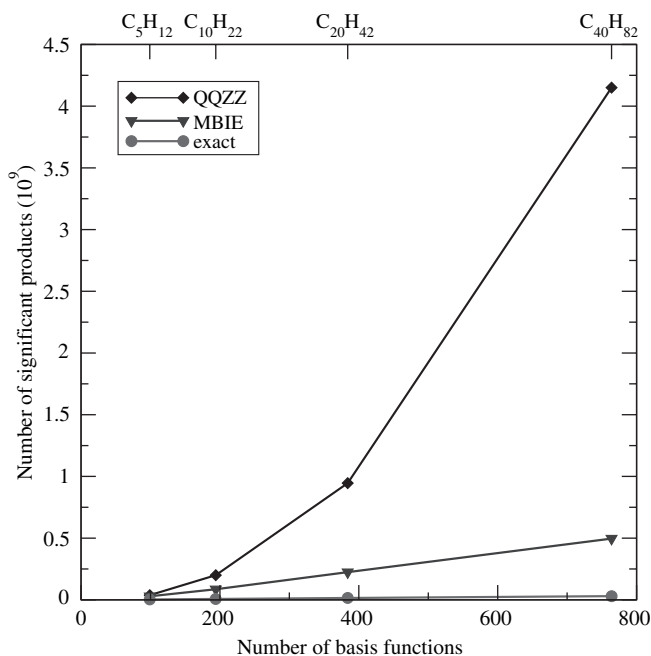
Therefore, for well-separated charge distributions in the above sense, the  $\frac{1}{R^4}$  behavior of transformed integrals turns into a  $\frac{1}{R^6}$  distance dependence. Such a behavior is well known for van der Waals/dispersion-type interactions. In contrast to the  $\frac{1}{R^4}$  dependence, however, the  $\frac{1}{R^6}$  behavior is linked closely to the exponential decay of the pseudo-density matrices  $\underline{\mathbf{P}}$  and  $\underline{\mathbf{P}}$  for systems with nonvanishing HOMO–LUMO gaps. Our experience from SCF theories shows that the one-particle density matrix is fairly long-ranged. Although we get for, e.g., DNA fragments a relatively early onset of a linear-scaling behavior for the computation of the Hartree–Fock exchange, it has to be stressed that this feature is strongly enhanced by the integral contractions (see Refs. 71 and 76) and not solely due to the locality of the one-particle density by itself. In this context, the true locality of the one-particle density matrix is needed for the  $\frac{1}{R^6}$  decay, so that this behavior is expected to start only at significantly larger distances as compared with the  $\frac{1}{R^4}$  decay. Nevertheless, it is clear that the  $\frac{1}{R^6}$  decay can be exploited in an analogous fashion by imposing the criteria listed above.

The implications of this decay behavior for the Coulomb-type products are illustrated in Figure 19 for the example of linear alkanes. For an alkane with four to five carbon atoms, the exact number of required transformed products (MP2, 6-31G\*, providing an accuracy of 0.1 mHartree for the first Laplace coefficient) scales already as low as  $N^{1.48}$  approaching the asymptotic linear scaling.<sup>24</sup> However, the pseudo-Schwarz screening drastically overestimates the number of required products and no linear-scaling can be achieved using this criterion, because the distance dependence of the transformed products is not accounted for.

### *Exchange-Type Contraction*

The exchange-type part of the AO–MP2 energy is computed as

$$e_K^{(\alpha)} = \sum_{\mu\nu\lambda\sigma} (\underline{\mu}\bar{\nu}|\underline{\lambda}\bar{\sigma})^{(\alpha)} (\mu\sigma|\lambda\nu) \quad [188]$$



**Figure 19** Comparison of the number of significant Coulomb-type integral products ( $C_nH_{2n+2}/6-31G^*$  basis; in units of  $10^9$ ) as estimated over shells by Schwarz-type screening (QQZZ;  $10^{-5}$ ) and MBIE ( $10^{-5}$ ) with the exact number of products selected via basis functions. For the latter, a threshold of  $10^{-8}$  has been selected to provide comparable accuracy in the absolute energies of 0.1 mH (only data for the first Laplace coefficient in computing the MP2 energy is listed).

As discussed, the transformed two-electron integrals decay as  $\frac{1}{R^3}$ , whereas the untransformed ones decay as  $\frac{1}{R}$ , resulting in a total distance dependence of  $\frac{1}{R^4}$ . In addition, in the exchange contraction step, the exponentially decaying *charge densities* of the untransformed integral  $\Omega_{\mu\sigma}$  and  $\Omega_{\lambda\nu}$  couple the two sides of the transformed integral. Therefore, as long as the transformed charge distributions  $\Omega_{\underline{\mu}\nu}$  and  $\Omega_{\underline{\lambda}\sigma}$  decay exponentially, an overall exponential decay for the exchange-type contraction results. The exponential coupling is similar to the one encountered for the formation of exchange-type contributions in SCF theories using our LinK method for computing energies,<sup>71</sup> energy gradients,<sup>76</sup> or NMR chemical shifts,<sup>113</sup> where the coupling of the two sides of the two-electron integrals is mediated over the one-particle densities or its derivatives. Therefore, in this context, the exchange-type contribution (Eq. [188]) to the correlation energy decays not only as  $\frac{1}{R^3}$ , but exponentially for systems with a nonvanishing HOMO-LUMO gap.<sup>24</sup>

## Rigorous Selection of Transformed Products via Multipole-Based Integral Estimates (MBIE)

The discussion above suggests that for the exploitation of the strong long-range decay behavior of at least  $\frac{1}{R^4}$  for electron correlation effects, it is crucial to introduce distance dependence into the integral estimates for transformed and untransformed two-electron integrals. Here, the MBIE scheme introduced by Lambrecht and Ochsenfeld<sup>23</sup> discussed in the introductory parts of the current review allows one to rigorously preselect which of the transformed products actually contribute to the correlation energy.<sup>24</sup>

To preselect the transformed integral products required for computing the MP2 energy, one can modify the MBIE integral bounds<sup>23,24</sup> so that an upper bound to the transformed integrals is obtained. In addition to this screening for the number of contributing products, one needs to select the significant untransformed integrals required for integral transformations. We will not discuss the derivation of MBIE bounds for AO-MP2 further in the current context, because the details would not provide more insight. For a detailed derivation the reader is referred to Ref. 24.

The performance of our MBIE method in its current stage for preselecting the significant number of contributing transformed products is illustrated in Figure 19. Although MBIE in its current stage still overestimates the number of products, it is always a true upper bound, so that noncontributing products can be safely discarded and only a linear-scaling number of products can be preselected. The MBIE estimate in Figure 19 has been optimized with respect to the transformation as compared with the one described in Ref. 24. We expect further improvements in the future in order to approach the true number of required products, so as to reduce the computational effort.

The MBIE screening formulas are crucial for the correct estimation of the long-range behavior of correlations effects:

- First, MBIE describes the exponential coupling of the “bra” and “ket” indices as does the QQZZ screening.
- Second, and most importantly, MBIE correctly describes the  $\frac{1}{R^4}$  dependence of the transformed products, so that for larger separations between “bra” and “ket” centers, the integral products vanish.
- Third, the MBIE estimates are rigorous upper bounds.

In addition, it is possible to exploit the  $\frac{1}{R^6}$  behavior as described above; however, the onset is expected to occur for a significantly larger “bra-ket” separation, so we focused here on the exploitation of the  $\frac{1}{R^4}$  decay.

## Implications

The considerations presented in this last section of the chapter illustrate that dynamic correlation is a local effect and that its description should,

therefore, scale linearly with the size of the molecule. This is not only true for the simple MP2 theory (where the “correlation interaction” between electrons decays as  $1/R^4$  and faster) on which we based our argumentation, but also for more sophisticated approaches like, for example, coupled-cluster theory. Although a tremendous amount of work has been done by many research groups in this field, much remains to be done. The path to such improvements is, in principle, set, and based on the example of the foregoing analysis, we can be optimistic that the scientific community will eventually reach this exciting goal of performing highly accurate *ab initio* calculations for very large molecular systems.

---

## CONCLUSIONS

Much work has been done by many scientists over the last decades to bring quantum chemistry to the impressive stage it is today. Thinking back to just a bit more than 15 years ago, computing a non-symmetric molecule with, say, 10–20 atoms at the Hartree–Fock level was painful. Today molecules with more than 1000 atoms can be tackled at the HF or DFT level on one-processor computers, and widespread applicability to a multitude of chemical and biochemical problems has been achieved. Although advances in quantum chemistry certainly go hand in hand with the fast-evolving increase of computer speed, it is clear that the introduction of linear-scaling methods over the last ten years, or so, has made important contributions to this success.

In this tutorial we have described some of the basic ideas for reducing the computational scaling of quantum chemical methods, without going into the details of the many different approaches followed by the numerous research groups involved in this field. We have presented linear-scaling methods for the calculation of SCF energies, energy gradients, and response properties, which open new possibilities for studying molecular systems with 1000 and more atoms on today’s computers. In addition, the given outlook on linear-scaling electron correlation methods indicates that much more can be expected and that more and more highly accurate approaches in the *ab initio* hierarchy will be available as well for large molecules.

Despite the success of linear-scaling methods, a multitude of challenges and open questions remain in the linear-scaling community. Some of the more important challenges include the following issues:

- Many molecular properties remain for which so far no linear-scaling methods have been devised and implemented.
- Reducing the prefactors stays an important issue and becomes even more important for linear-scaling methods (because any gain in the prefactor directly translates into the treatable molecule size).

- The results for some molecular properties or electron correlation energies depend strongly on the size of the basis set; post-HF methods, in particular, require large basis sets. Even if a method scales linearly with molecular size, the computational cost may increase dramatically with the basis set size. Therefore, much more work needs to be devoted for tackling this basis set problem.
- The more “metallic” a system is (small HOMO–LUMO gap), the less local is the one-particle density matrix. Therefore, the question about how to deal with strong delocalization in an efficient manner remains an important challenge.
- Because matrix multiplications are central to many aspects of linear-scaling schemes, any further speed-up in sparse matrix multiplications will be of importance, in particular if the systems are more “metallic”.
- Although a multitude of open questions still exists even for HF and DFT linear-scaling schemes, the rigorous and efficient reduction of scaling in post-HF methods to account for the missing electron correlation effects remains one of the central challenges for the success of quantum chemistry.
- Many large molecular systems are flexible, and dynamic effects are necessary for a realistic description. Therefore, molecular dynamics simulations are needed that require the computation of a huge number of points on a hypersurface, resulting in an extremely high computational cost for reliable methods.

From this small list of challenges, it becomes clear that there is still a great need for developing and improving linear-scaling methods. Nevertheless, the foregoing discussion has shown that much has been achieved for the approximate solution of the Schrödinger equation even for large molecules. For the future, the ultimate goal of solving the molecular Schrödinger equation to highest accuracy and efficiency appears to be reachable. Accomplishing this goal will allow us to rationalize and understand, to predict, and ultimately, to control the chemical and biochemical processes of very large molecular systems.

---

## REFERENCES

1. E. Schrödinger, *Ann. Phys.*, **79**, 361 (1926). Quantisierung als Eigenwertproblem (erste Mitteilung).
2. D. R. Hartree, *Proc. Cambridge Philos. Soc.*, **24**, 89 (1928). The Wave Mechanics of an Atom with a Non-Coulomb Central Field. I. Theory and Methods.
3. V. Fock, *Z. Phys.*, **61**, 126 (1930). Näherungsmethode zur Lösung des Quantenmechanischen Mehrkörperproblems.; *Z. Phys.*, **62**, 795 (1930). “Self-Consistent Field” mit Austausch für Natrium.

4. A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry - Introduction to Advanced Electronic Structure Theory*, Dover Publications, Inc., Mineola, New York, 1989.
5. C. Møller and M. S. Plesset, *Phys. Rev.*, **46**, 618 (1934). Note on an Approximation Treatment for Many-Electron Systems.
6. R. J. Bartlett and J. F. Stanton, in *Reviews in Computational Chemistry*, Vol. 5, K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, pp. 65–169, 1990. Application of Post-Hartree-Fock Methods: A Tutorial.
7. P. Hohenberg and W. Kohn, *Phys. Rev. B*, **136**, 864 (1964). Inhomogeneous Electron Gas.
8. W. Kohn and L. J. Sham, *Phys. Rev.*, **140**, A1133 (1965). Self-Consistent Equations Including Exchange and Correlation Effects.
9. R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, International Series of Monographs on Chemistry 16, Oxford Science Publications, Oxford, United Kingdom, 1989.
10. G. E. Moore, *Electronics Magazine*, 19 April, 1965. Cramming More Components onto Integrated Circuits.
11. W. Kutzelnigg, *Einführung in die Theoretische Chemie*, VCH Weinheim, Weinheim, Germany, 2001.
12. I. N. Levine, *Quantum Chemistry*, Fifth ed., Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 2000.
13. T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic Structure Theory*, Wiley, Chichester, United Kingdom, 2000.
14. M. Born and R. A. Oppenheimer, *Ann. Phys.*, **84**, 457 (1927) Zur Quantentheorie der Molekeln.
15. B. T. Sutcliffe, in *Computational Techniques in Quantum Chemistry*, G. H. F. Diercksen, B. T. Sutcliffe, and A. Veillard, Eds., Reidel, Boston, Massachusetts, 1975, pp. 1–105. Fundamentals of Computational Quantum Chemistry.
16. J. C. Slater, *Quantum Theory of Matter*, 2nd ed., McGraw-Hill, New York, 1968.
17. C. C. J. Roothaan, *Rev. Mod. Phys.*, **23**, 69 (1951). New Developments in Molecular Orbital Theory.
18. G. G. Hall, *Proc. Roy. Soc.*, **A205**, 541 (1951). The Molecular-Orbital Theory of Chemical Valency. VIII. A Method of Calculating Ionization Potentials.
19. J. Almlöf, K. Faegri, and K. Korsell, *J. Comput. Chem.*, **3**, 385 (1982). Principles for a Direct SCF Approach to LCAO-MO Ab-initio Calculations.
20. V. Dyzmoms, *Theoret. Chim. Acta*, **28**, 307 (1973). No  $N^4$ -dependence in the Calculation of Large Molecules.
21. M. Häser and R. Ahlrichs, *J. Comput. Chem.*, **10**, 104 (1989). Improvements on the Direct SCF Method.
22. D. Cremer and J. Gauss, *J. Comput. Chem.*, **7**, 274 (1986). An Unconventional SCF Method for Calculations on Large Molecules.
23. D. S. Lambrecht and C. Ochsenfeld, *J. Chem. Phys.*, **123**, 184101 (2005). Multipole-Based Integral Estimates for the Rigorous Description of Distance Dependence in Two-Electron Integrals.
24. D. S. Lambrecht, B. Doser, and C. Ochsenfeld, *J. Chem. Phys.*, **123**, 184102 (2005). Rigorous Integral Screening for Electron Correlation Methods.
25. J. E. Almlöf, USIP Report 72-09 (1972), republished in *Theor. Chem. Acc. memorial issue: P. R. Taylor* *Theor. Chem. Acc.*, **97**, 10 (1997). Methods for the Rapid Evaluation of Electron Repulsion Integrals in Large-scale LCGO Calculations.
26. J. E. Almlöf, in *Modern Electronic Structure Theory*, D. Yarkony, C.-Y. Ng, Eds., World Scientific Singapore, 1994, pp. 121–151. Direct Methods in Electronic Structure Theory.
27. H. Eyring, J. Walter, and G. E. Kimball, *Quantum Chemistry*, Wiley, New York, 1947.

28. J. O. Hirschfelder, C. F. Curtiss, and R. B. Byron, *Molecular Theory of Gases and Liquids*, Wiley, New York, 1954.
29. A. D. Buckingham, in *Intermolecular Interactions: From Diatomics to Biopolymers*, B. Pullman, Ed., Wiley, New York, 1987, pp. 1–67. Basic Theory of Intermolecular Forces: Applications to Small Molecules.
30. P. M. W. Gill, B. G. Johnson, and J. A. Pople, *Chem. Phys. Lett.*, **217**, 65 (1994). A Simple yet Powerful Upper Bound for Coulomb Integrals.
31. C. A. White, B. G. Johnson, P. M. W. Gill, and M. Head-Gordon, *Chem. Phys. Lett.*, **230**, 8 (1994). The Continuous Fast Multipole Method.
32. P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, Volume I, McGraw-Hill Education, Tokyo, Japan, 1953.
33. P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, Volume II, McGraw-Hill Education, Tokyo, Japan, 1953.
34. D. E. Williams, in *Reviews in Computational Chemistry*, Vol. 2, K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1991, pp. 219–271. Net Atomic Charge and Multipole Models for the ab Initio Molecular Electric Potential.
35. G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists*, Academic Press, London, United Kingdom, 2001.
36. S. Obara and A. Saika, *J. Chem. Phys.*, **84**, 3963 (1985). Efficient Recursive Computation of Molecular Integrals over Cartesian Gaussian Functions.
37. M. Head-Gordon and J. A. Pople, *J. Chem. Phys.*, **89**, 5777 (1988). A Method for Two-electron Gaussian Integral and Integral Derivative Evaluation Using Recurrence Relations.
38. P. M. W. Gill and J. A. Pople, *J. Quantum Chem.*, **40**, 753 (1991). The Prism Algorithm for Two-Electron Integrals.
39. L. E. McMurchie and E. R. Davidson, *J. Comput. Phys.*, **26**, 218 (1978). One- and Two-Electron Integrals over Cartesian Gaussian Functions.
40. P. M. W. Gill, B. G. Johnson, and J. A. Pople, *Int. J. Quantum Chem.*, **40**, 745 (1991). Two-Electron Repulsion Integrals Over Gaussian s Functions.
41. P. M. W. Gill, M. Head-Gordon, and J. A. Pople, *Int. J. Quantum Chem.*, **23**, 269 (1989). An Efficient Algorithm for the Generation of Two-Electron Repulsion Integrals over Gaussian Basis Functions.
42. A. V. Scherbinin, V. I. Pypyshev, and N. F. Stepanov *Int. J. Quantum Chem.*, **60**, 843 (1996). On the Use of Multipole Expansion of the Coulomb Potential in Quantum Chemistry.
43. V. R. Saunders, in *Methods in Computational Molecular Physics*, G. H. F. Dierksen and S. Wilson, Eds., NATO ASI Series, Series C: Mathematical and Physical Sciences, Vol. **113**, D. Reidel Publishing Company, Dordrecht, The Netherlands, 1983, pp. 1–36. Molecular Integrals for Gaussian Type Functions.
44. T. Helgaker and P. R. Taylor, *Modern Electronic Structure Theory*, Vol. 2, D. Yarkony, Ed., World Scientific, Singapore, 1995, pp. 725–856. Gaussian Basis Sets and Molecular Integrals.
45. L. Greengard and V. Rokhlin, *J. Comput. Phys.*, **60**, 187 (1990). Rapid Solution of Integral Equations of Classical Potential Theory.
46. R. Beatson and L. Greengard, Available: [www.math.nyu.edu/faculty/greengar/shortcourse\\_fmm.pdf](http://www.math.nyu.edu/faculty/greengar/shortcourse_fmm.pdf). A Short Course on Fast Multipole Methods.
47. L. Greengard, *Science*, **265**, 909 (1994). Fast Algorithms for Classical Physics.
48. C. A. White and M. Head-Gordon, *J. Chem. Phys.*, **101**, 6593 (1994). Derivation and Efficient Implementation of the Fast Multipole Method.
49. C. A. White and M. Head-Gordon, *J. Chem. Phys.*, **105**, 5061 (1996). Rotating Around the Angular Momentum Barrier in Fast Multipole Method Calculations.
50. J. Barnes and P. Hut, *Nature* (London), **324**, 446 (1986). A Hierarchical  $\mathcal{O}(N \log N)$  Force-Calculation Algorithm.

51. M. Challacombe, E. Schwegler, and J. Almlöf, in *Computational Chemistry: Review of Current Trends*, Vol. 53, J. Leszczynski, Ed., World Scientific, Singapore, 1996, pp. 4685–4695. Modern Developments in Hartree-Fock Theory: Fast Methods for Computing the Coulomb Matrix.
52. M. Challacombe, E. Schwegler, and J. Almlöf, *J. Chem. Phys.*, **104**, 4685 (1995). Fast Assembly of the Coulomb Matrix: A Quantum Chemical Tree Code.
53. J. Cipriani and B. Silvi, *Mol. Phys.*, **45**, 259 (1982). Cartesian Expression of Electric Multipole Moments.
54. L. Greengard and J. Strain, *J. Sci. Stat. Comp.*, **12**, 79 (1991). The Fast Gauss Transform.
55. H. G. Petersen, D. Soelvason, J. W. Perram, and E. R. Smith, *J. Chem. Phys.*, **101**, 8870 (1994). The very fast multipole method.
56. M. C. Strain, G. E. Scuseria, and M. J. Frisch, *Science*, **271**, 51 (1996). Achieving Linear Scaling for the Electronic Quantum Coulomb Problem.
57. O. Vahtras, J. Almlöf, and M. W. Feyereisen, *Chem. Phys. Lett.*, **213**, 514 (1993). Integral Approximations for LCAO-SCF Calculations.
58. R. A. Kendall and H. A. Früchtel, *Theor. Chem. Acc.*, **97**, 158 (1997). The Impact of the Resolution of the Identity Approximate Integral Method on Modern Ab Initio Algorithm Development.
59. B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin, *J. Chem. Phys.*, **71**, 3396 (1979). On Some Approximations in Applications of  $X\alpha$  Theory.
60. K. Eichkorn, O. Treutler, H. Oehm, M. Häser, and R. Ahlrichs, *Chem. Phys. Lett.*, **240**, 283 (1995). Auxiliary Basis Sets to Approximate Coulomb Potentials.
61. F. Weigend, *Phys. Chem. Chem. Phys.*, **4**, 4285 (2002). A Fully Direct RI-HF Algorithm: Implementation, Optimised Auxiliary Basis Sets, Demonstration of Accuracy and Efficiency.
62. M. Sierka, A. Hogekamp, and R. Ahlrichs, *J. Chem. Phys.*, **118**, 9136 (2003). Fast Evaluation of The Coulomb Potential for Electron Densities Using Multipole Accelerated Resolution of Identity Approximation.
63. L. Füsti-Molnar and P. Pulay, *J. Chem. Phys.*, **117**, 7827 (2002). The Fourier Transform Coulomb Method: Efficient and Accurate Calculation of the Coulomb Operator in a Gaussian Basis.
64. L. Füsti-Molnar and P. Pulay, *J. Mol. Struct. (THEOCHEM)*, 666–667, **25** (2003). Gaussian-based First-principles Calculations on Large Systems Using the Fourier Transform Coulomb Method.
65. L. Füsti-Molnar and P. Pulay, *J. Chem. Phys.*, **119**, 11080 (2003). New Developments in the Fourier Transform Coulomb Method: Efficient and Accurate Localization of the Filtered Core Functions and Implementation of the Coulomb Energy Forces.
66. L. Füsti-Molnar and J. Kong, *J. Chem. Phys.*, **122**, 074108 (2005). Fast and Accurate Coulomb Calculation with Gaussian Functions.
67. R. Ahlrichs, M. Hoffmann-Ostenhof, T. Hoffmann-Ostenhof, and J. D. Morgan III, *Phys. Rev. A*, **23**, 2106 (1981). Bounds on Decay of Electron Densities with Screening.
68. C. Ochsenfeld and M. Head-Gordon, *Chem. Phys. Lett.*, **270**, 399 (1997). A Reformulation of the Coupled Perturbed Self-consistent Field Equations Entirely Within a Local Atomic Orbital Density Matrix-based Scheme.
69. P. E. Maslen, C. Ochsenfeld, C. A. White, M. S. Lee, and M. Head-Gordon, *J. Phys. Chem.*, **102**, 2215 (1998). Locality and Sparsity of ab initio One-Particle Density Matrices and Localized Orbitals.
70. W. Kohn, *Phys. Rev. Lett.*, **76**, 3168 (1996). Density Functional and Density Matrix Method Scaling Linearly with the Number of Atoms.
71. C. Ochsenfeld, C. A. White, and M. Head-Gordon, *J. Chem. Phys.*, **109**, 1663 (1998). Linear and Sublinear Scaling Formation of Hartree-Fock-type Exchange Matrices.

72. E. Schwegler and M. Challacombe, *J. Chem. Phys.*, **105**, 2726 (1996). Linear Scaling Computation of the Hartree-Fock Exchange Matrix.
73. J. C. Burant, G. E. Scuseria, and M. J. Frisch, *J. Chem. Phys.*, **105**, 8969 (1996). A Linear Scaling Method for Hartree-Fock Exchange Calculations of Large Molecules.
74. E. Schwegler, M. Challacombe, and M. Head-Gordon, *J. Chem. Phys.*, **106**, 9708 (1997). Linear Scaling Computation of the Fock Matrix. II. Rigorous Bounds on Exchange Integrals and Incremental Fock Build.
75. E. Schwegler and M. Challacombe, *Theoret. Chim. Acta*, **104**, 344 (2000). Linear Scaling Computation of the Hartree-Fock Exchange Matrix. III. Formation of the Exchange Matrix with Permutational Symmetry.
76. C. Ochsenfeld, *Chem. Phys. Lett.*, **327**, 216 (2000). Linear Scaling Exchange Gradients for Hartree-Fock and Hybrid Density Functional Theory.
77. H. Sambe and R. H. Felton, *J. Chem. Phys.*, **62**, 1122 (1975). A New Computational Approach to Slater's SCF- $X\alpha$  Equation.
78. C. Satoko, *Chem. Phys. Lett.*, **82**, 111 (1981). Direct Force Calculation in the  $X\alpha$  Method and its Application to Chemisorption of an Oxygen Atom on the Al(111) Surface.
79. R. Fournier, J. Andzelm, and D. R. Salahub, *J. Chem. Phys.*, **90**, 6371 (1989). Analytical Gradient of the Linear Combination of Gaussian-type Orbitals — Local Spin Density Energy.
80. J. A. Pople, P. M. W. Gill, and B. G. Johnson, *Chem. Phys. Lett.*, **199**, 557 (1992). Kohn-Sham Density-Functional Theory within a Finite Basis Set.
81. J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, *Phys. Rev. Lett.*, **91**, 146401 (2003). Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids.
82. J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria, and G. I. Csonka, *J. Chem. Phys.*, **123**, 062201 (2005). Prescription for the Design and Selection of Density Functional Approximations: More Constraint Satisfaction with Fewer Fits.
83. B. G. Johnson, C. A. White, Q. Zhang, B. Chen, R. L. Graham, P. M. W. Gill, and M. Head-Gordon, in *Recent Developments in Density Functional Theory*, J. M. Seminario, Ed., Vol. 4, Elsevier, Amsterdam, The Netherlands, 1996, pp. 441–463. Advances in Methodologies for Linear-Scaling Density Functional Calculations.
84. G. E. Scuseria, *J. Phys. Chem. A*, **103**, 4782 (1999). Linear Scaling Density Functional Calculations with Gaussian Orbitals.
85. A. D. Becke, *J. Chem. Phys.*, **98**, 5648 (1992). Density-Functional Thermochemistry. III. The Role of Exact Exchange.
86. P. M. W. Gill, B. G. Johnson, and J. A. Pople, *Chem. Phys. Lett.*, **209**, 506 (1993). A Standard Grid for Density Functional Calculations.
87. A. D. Becke, *J. Chem. Phys.*, **88**, 2547 (1988). A Multicenter Numerical Integration Scheme for Polyatomic Molecules.
88. X.-P. Li, R. W. Nunes, and D. Vanderbilt, *Phys. Rev. B*, **47**, 10891 (1993). Density-matrix Electronic-structure Method with Linear System-size Scaling.
89. R. W. Nunes and D. Vanderbilt, *Phys. Rev. B*, **50**, 17611 (1994). Generalization of the Density-matrix Method to a Nonorthogonal Basis.
90. J. M. Milliam and G. E. Scuseria, *J. Chem. Phys.*, **106**, 5569 (1997). Linear Scaling Conjugate Gradient Density Matrix Search as an Alternative to Diagonalization for First Principles Electronic Structure Calculations.
91. M. Head-Gordon, Y. Shao, C. Saravanan, and C. A. White, *Mol. Phys.*, **101**, 37 (2003). Curvy Steps for Density Matrix Based Energy Minimization: Tensor Formulation and Toy Applications.
92. T. Helgaker, H. Larsen, J. Olsen, and P. Jørgensen, *Chem. Phys. Lett.*, **327**, 397 (2000). Direct Optimization of the AO Density Matrix in Hartree-Fock and Kohn-Sham Theories.

93. H. Larsen, J. Olsen, P. Jørgenson, and T. Helgaker, *J. Chem. Phys.*, **115**, 9685 (2001). Direct Optimization of the Atomic-orbital Density Matrix Using the Conjugate-gradient Method with a Multilevel Preconditioner.
94. M. Challacombe, *J. Chem. Phys.*, **110**, 2332 (1999). A Simplified Density Matrix Minimization for Linear Scaling Self-Consistent Field Theory.
95. W. Yang, *Phys. Rev. Lett.*, **66**, 1438 (1991). Direct Calculation of Electron Density in Density-Functional Theory.
96. W. Yang, *J. Chem. Phys.*, **94**, 1208 (1991). A Local Projection Method for the Linear Combination of Atomic Orbital Implementation of Density-Functional Theory.
97. Q. Zhao and W. Yang, *J. Chem. Phys.*, **102**, 9598 (1995). Analytical Energy Gradients and Geometry Optimization in the Divide-and-conquer Method for Large Molecules.
98. W. Yang and T.-S. Lee, *J. Chem. Phys.*, **103**, 5674 (1995). A Density-matrix Divide-and-conquer Approach for Electronic Structure Calculations of Large Molecules.
99. S. Goedecker and L. Colombo, *Phys. Rev. Lett.*, **73**, 122 (1994). Efficient Linear Scaling Algorithm for Tight-binding Molecular Dynamics.
100. S. Goedecker and M. Teter, *Phys. Rev. B*, **51**, 9455 (1995). Tight-binding Electronic-structure Calculations and Tight-binding Molecular Dynamics with Localized Orbitals.
101. S. Goedecker, *J. Comput. Phys.*, **118**, 261 (1995). Low Complexity Algorithms for Electronic Structure Calculations.
102. J. Kim, F. Mauri, and G. Galli, *Phys. Rev. B*, **52**, 1640 (1995). Total-energy Global Optimizations using Nonorthogonal Localized Orbitals.
103. F. Mauri and G. Galli, *Phys. Rev. B*, **50**, 4316 (1994). Electronic-structure Calculations and Molecular-dynamics Simulations with Linear System-size Scaling.
104. F. Mauri, G. Galli, and R. Car, *Phys. Rev. B*, **47**, 9973 (1993). Orbital Formulation for Electronic-structure Calculations with Linear System-size Scaling.
105. P. Ordejon, *Comput. Mater. Sci.*, **12**, 157 (1998). Order-N Tight-binding Methods for Electronic-structure and Molecular Dynamics.
106. E. Hernandez and M. Gillan, *Phys. Rev. B*, **51**, 10157 (1995). Self-consistent First-principles Technique with Linear Scaling.
107. W. Hiese and E. Stechel, *Phys. Rev. B*, **50**, 17811 (1994). Order-N Methods in Self-consistent Density-functional Calculations.
108. S. Goedecker, *Rev. Mod. Phys.*, **71**, 1085 (1999). Linear Scaling Electronic Structure Methods.
109. S. Goedecker and G. E. Scuseria, *Commun. Science & Engineering*, **5**, 14 (2003). Linear Scaling Electronic Structure Methods in Chemistry and Physics.
110. A. D. Daniels and G. E. Scuseria, *J. Chem. Phys.*, **110**, 1321 (1999). What Is the Best Alternative to Diagonalization of the Hamiltonian in Large Scale Semiempirical Calculations?
111. D. R. Bowler, T. Miyazaki, and M. J. Gillan, *J. Phys.: Condens. Matter*, **14**, 2781 (2002). Recent Progress in Linear Scaling Ab Initio Electronic Structure Techniques.
112. D. R. Bowler, I. J. Bush, and M. J. Gillan, *Int. J. Quantum Chem.*, **77**, 831 (2000). Practical Methods for Ab Initio Calculations on Thousands of Atoms.
113. C. Ochsenfeld, J. Kussmann, and F. Koziol, *Angew. Chem., Int. Ed.*, **43**, 4485 (2004); *Angew. Chem. Int. Ed.*, **43**, 4485 (2004). Ab Initio NMR Spectra for Molecular Systems with a Thousand and More Atoms: A Linear-Scaling Method.
114. M. Head-Gordon, M. S. Lee, P. E. Maslen, T. van Voorhis, and S. Gwaltney, *Modern Methods and Algorithms of Quantum Chemistry*, Proceedings, Second ed., J. Grotendorst, Ed., John von Neumann Institute for Computing, Jülich, Germany, NIC Series, Vol. 3, 2000, pp. 593–638. Tensors in Electronic Structure Theory: Basic Concepts and Applications to Electron Correlation Models.

115. J. A. Schouten, *Tensor Analysis for Physicists*, 2nd ed., Dover Publications, Mineola, New York, 1988.
116. M. Head-Gordon, P. E. Maslen, and C. A. White, *J. Chem. Phys.*, **108**, 616 (1998). A Tensor Formulation of Many-electron Theory in a Nonorthogonal Single-particle Basis.
117. A. Messiah, *Quantum Mechanics*, Dover Publications, Mineola, New York, 1999.
118. R. McWeeny, *Methods of Molecular Quantum Mechanics* (Theoretical Chemistry), 2nd ed., Academic Press Limited, London, United Kingdom, 1989.
119. M. S. Daw, *Phys. Rev. B*, **47**, 10895 (1993). Model for Energetics of Solids based on the Density Matrix.
120. R. McWeeny, *Phys. Rev.*, **114**, 1528 (1959). Hartree-Fock Theory with Nonorthogonal Basis Functions.
121. R. McWeeny, *Rev. Mod. Phys.*, **32**, 335 (1960). Some Recent Advances in Density Matrix Theory.
122. C. A. White, P. E. Maslen, M. S. Lee, and M. Head-Gordon, *Chem. Phys. Lett.*, **276**, 133 (1997). The Tensor Properties of Energy Gradients Within a Non-orthogonal Basis.
123. J. J. Sakurai, *Modern Quantum Mechanics*, Addison Wesley, Reading, Massachusetts, 1993.
124. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran*, 2nd ed., Cambridge University Press, Cambridge, United Kingdom, 1996.
125. P. Pulay, *Mol. Phys.* **17**, 197 (1969). Ab initio Calculation of Force Constants and Equilibrium Geometries. I. Theory.
126. P. Pulay, in *Modern Electronic Structure Theory*, D. Yarkony, Ed., World Scientific, Singapore, 1995, pp. 1191–1240. Analytical Derivative Techniques and the Calculation of Vibrational Properties, in *Modern Electronic Structure Theory*.
127. P. Pulay, in *Ab Initio Methods in Quantum Chemistry*, K. P. Lawley, Ed., Wiley, New York, 1987, pp. 241–286. Analytic Derivative Methods in Quantum Chemistry.
128. M. Frisch, M. Head-Gordon, and J. A. Pople, *Chem. Phys.*, **141**, 189 (1990). Direct Analytic SCF Second Derivatives and Electric Field Properties.
129. Y. Shao, C. A. White, and M. Head-Gordon, *J. Chem. Phys.*, **114**, 6572 (2001). Efficient Evaluation of the Coulomb Force in Density Functional Theory Calculations.
130. J. Gauss, in *Modern Methods and Algorithms of Quantum Chemistry*, Proceedings, 2nd ed., J. Grotendorst, Ed., John von Neumann Institute for Computing, Jülich, Germany, NIC Series Vol. 3, 2000, pp. 541–592. Molecular Properties.
131. J. Gerratt and I. M. Mills, *J. Chem. Phys.*, **49**, 1968 (1979). Force Constants and Dipole-Moment Derivatives of Molecules from Perturbed Hartree-Fock Calculations.
132. C. E. Dykstra and P. G. Jasien, *Chem. Phys. Lett.*, **109**, 388 (1984). Derivative Hartree-Fock Theory to all Orders.
133. N. C. Handy, D. J. Tozer, G. J. Laming, C. W. Murray, and R. D. Amos, *Isr. J. Chem.* **33**, 331 (1993). Analytic Second Derivatives of the Potential Energy Surface.
134. B. G. Johnson and M. J. Fisch, *J. Chem. Phys.*, **100**, 7429 (1994). An Implementation of Analytic Second Derivatives of the Gradient-corrected Density Functional Energy.
135. J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley, *Int. J. Quantum Chem. Symp.*, **S13**, 225 (1979). Derivative Studies in Hartree-Fock and Møller-Plesset Theories.
136. Y. Osamura, Y. Yamaguchi, P. Saxe, D. J. Fox, M. A. Vincent, and H. F. Schafer III., *J. Mol. Struct.: THEOCHEM*, **103**, 183 (1983). Analytic Second Derivative Techniques for Self-Consistent-Field Wave Functions. A new Approach to the Solution of the coupled Perturbed Hartree-Fock Equations.
137. J. Gauss, *Ber. Bunsenges. Phys. Chem.*, **99**, 1001 (1995). Accurate Calculation of NMR Chemical Shifts.
138. T. Helgaker, M. Jaszunski, and K. Ruud, *Chem. Rev.*, **99**, 293 (1999). Ab Initio Methods for the Calculation of NMR Shielding and Indirect Spin-Spin Coupling Constants.

139. U. Fleischer, W. Kutzelnigg, and C. van Wüllen, in *Encyclopedia of Computational Chemistry*, P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, and P. R. Schreiner, Eds., Wiley, Chichester, United Kingdom, 1998, pp. 1827. Ab initio NMR Chemical Shift Computation.
140. T. Helgaker, P. J. Wilson, R. D. Amos, and N. C. Handy, *J. Chem. Phys.*, **113**, 2983 (2000). Nuclear Shielding Constants by Density Functional Theory with Gauge Including Atomic Orbitals.
141. G. Schreckenbach and T. Ziegler, *J. Phys. Chem.*, **99**, 606 (1995). Calculation of NMR Shielding Tensors Using Gauge-Including Atomic Orbitals and Modern Density Functional Theory.
142. W. Kutzelnigg, *Isr. J. Chem.*, **19**, 193 (1980). Theory of Magnetic Susceptibilities and NMR Chemical Shifts in Terms of Localized Quantities.
143. A. E. Hansen and T. D. Bouman, *J. Chem. Phys.*, **82**, 5035 (1985). Localized Orbital/Local Origin Method for Calculation and Analysis of NMR Shieldings. Applications to  $^{13}\text{C}$  Shielding Tensors.
144. F. London, *J. Phys. Radium*, **8**, 397 (1937). Quantum Theory of Interatomic Currents in Aromatic Compounds.
145. R. Ditchfield, *Molecular Physics*, **27**, 789 (1974). Self-consistent Perturbation Theory of Diamagnetism. I. A Gauge-invariant LCAO Method for N.M.R. Chemical Shifts.
146. K. Wolinski, J. F. Hinton, and P. Pulay, *J. Am. Chem. Soc.*, **112**, 8251 (1990). Efficient Implementation of the Gauge-Independent Atomic Orbital Method for NMR Chemical Shift Calculations.
147. J. R. Cheeseman, G. W. Trucks, T. A. Keith, and M. J. Frisch, *J. Chem. Phys.*, **104**, 5497 (1996). A Comparison of Models for Calculating Nuclear Magnetic Resonance Shielding Tensors.
148. M. Häser, R. Ahlrichs, H. P. Baron, P. Weis, and H. Horn, *Theoret. Chim. Acta*, **83**, 455 (1992). Direct Computation of Second-order SCF Properties of Large Molecules on Workstation Computers with an Application to Large Carbon Clusters.
149. C. Ochsenfeld, *Phys. Chem. Chem. Phys.*, **2**, 2153 (2000). An Ab Initio Study of the Relation between NMR Chemical Shifts and Solid-State Structures: Hexabenzocoronene Derivatives.
150. C. Ochsenfeld, S. P. Brown, I. Schnell, J. Gauss, and H. W. Spiess, *J. Am. Chem. Soc.*, **123**, 2597 (2001). Structure Assignment in the Solid State by the Coupling of Quantum Chemical Calculations with NMR Experiments: A Columnar Hexabenzocoronene Derivative.
151. S. P. Brown, T. Schaller, U. P. Seelbach, F. Koziol, C. Ochsenfeld, F.-G. Klärner, and H. W. Spiess, *Angew. Chem. Int. Ed.*, **40**, 717 (2001). Structure and Dynamics of the Host-Guest Complex of a Molecular Tweezer: Coupling Synthesis, Solid-State NMR, and Quantum-Chemical Calculations.
152. C. Ochsenfeld, F. Koziol, S. P. Brown, T. Schaller, U. P. Seelbach, and F.-G. Klärner, *Solid State Nucl. Magn. Reson.*, **22**, 128 (2002). A Study of a Molecular Tweezer Host-Guest System by a Combination of Quantum-Chemical Calculations and Solid-State NMR Experiments.
153. J. Gauss, and J. F. Stanton, *Adv. Chem. Phys.* **123**, 355 (2002). Electron-Correlated Approaches for the Calculation of NMR Chemical Shifts.
154. M. Kaupp, M. Bühl, and V. G. Malkin (Eds.), *Calculation of NMR and EPR Parameters*, Wiley-VCH Weinheim, 2004.
155. H. Larsen, T. Helgaker, J. Olsen, and P. Jørgensen, *J. Chem. Phys.*, **115**, 10344 (2001). Geometrical Derivatives and Magnetic Properties in Atomic-orbital Density-based Hartree-Fock Theory.
156. V. Weber and M. Challacombe, *J. Chem. Phys.*, **123**, 044106 (2005). Higher-Order Response in  $\mathcal{O}(N)$  by Perturbed Projection.
157. P. Pulay, *Chem. Phys. Lett.*, **100**, 151 (1983). Localizability of Dynamic Electron Correlation.

158. S. Saebø and P. Pulay, *Chem. Phys. Lett.*, **113**, 13 (1985). Local Configuration Interaction: An Efficient Approach for Larger Molecules.
159. P. Pulay and S. Saebø, *Theoret. Chim. Acta*, **69**, 357 (1985). Orbital-invariant Formulation and Second-order Gradient Evaluation in Møller-Plesset Perturbation Theory.
160. S. Saebø and P. Pulay, *J. Chem. Phys.*, **86**, 914 (1987). Fourth-order Møller-Plesset Perturbation Theory in the Local Correlation Treatment. I. Method.
161. C. Hampel and H.-J. Werner, *J. Chem. Phys.*, **104**, 6286 (1996). Local Treatment of Electron Correlation in Coupled Cluster Theory.
162. M. Schütz, G. Hetzer, and H.-J. Werner, *J. Chem. Phys.*, **111**, 5691 (1999). Low-order Scaling Local Electron Correlation Methods. I. Linear Scaling Local MP2.
163. G. Hetzer, M. Schütz, H. Stoll, and H.-J. Werner, *J. Chem. Phys.*, **113**, 9443 (2000). Low-Order Scaling Local Correlation Methods II: Splitting the Coulomb Operator in Linear Scaling Local Second-Order Møller-Plesset Perturbation Theory.
164. P. E. Maslen and M. Head-Gordon, *Chem. Phys. Lett.*, **283**, 102 (1998). Non-iterative Local Second Order Møller-Plesset Theory.
165. M. S. Lee, P. E. Maslen, and M. Head-Gordon, *J. Chem. Phys.*, **112**, 3592 (2000). Closely Approximating Second-order Møller-Plesset Perturbation Theory with a Local Triatomics in Molecules Model.
166. J. Almlöf, *Chem. Phys. Lett.*, **181**, 319 (1991). Elimination of Energy Denominators in Møller-Plesset Perturbation Theory by a Laplace Transform Approach.
167. M. Häser and J. Almlöf, *J. Chem. Phys.*, **96**, 489 (1992). Laplace Transform Techniques in Møller-Plesset Perturbation Theory.
168. M. Häser, *Theoret. Chim. Acta*, **87**, 147 (1993). Møller-Plesset (MP2) Perturbation Theory for Large Molecules.
169. P. Y. Ayala and G. E. Scuseria, *J. Chem. Phys.*, **110**, 3660 (1999). Linear Scaling Second-order Møller-Plesset Theory in the Atomic Orbital Basis for Large Molecular Systems.
170. G. E. Scuseria and P. Y. Ayala, *J. Chem. Phys.*, **111**, 8330 (1999). Linear Scaling Coupled Cluster and Perturbation Theories in the Atomic Orbital Basis.
171. R. Friesner, R. B. Murphy, M. D. Beachy, M. N. Ringnalda, W. T. Pollard, B. D. Dunietz, and Y. Cao, *J. Phys. Chem.*, **103**, 1913 (1999). Correlated ab Initio Electronic Structure Calculations for Large Molecules.
172. D. Walter, A. B. Szilva, K. Niedfeldt, and E. A. Carter, *J. Chem. Phys.*, **117**, 1982 (2002). Local Weak-pairs Pseudospectral Multireference Configuration Interaction.
173. M. Schütz and H.-J. Werner, *J. Chem. Phys.*, **114**, 661 (2001). Low-order Scaling Local Electron Correlation Methods. IV. Linear Scaling Local Coupled-Cluster (LCCSD).
174. M. Schütz, *J. Chem. Phys.*, **116**, 8772 (2002). Low-order Scaling Local Electron Correlation Methods. V. Connected Triples beyond (T): Linear Scaling Local CCSDT-1b.
175. G. E. Scuseria and P. Y. Ayala, *J. Chem. Phys.*, **111**, 8330 (1999). Linear Scaling Coupled Cluster and Perturbation Theories in the Atomic Orbital Basis.
176. P. Knowles, M. Schütz, and H.-J. Werner, in *Modern Methods and Algorithms of Quantum Chemistry*, Proceedings, Second ed., J. Grotendorst, Ed., John von Neumann Institute for Computing, Jülich, Germany NIC Series, Vol. 3, 2000, pp. 97–197. Ab Initio Methods for Electron Correlation in Molecules.
177. J. L. Whitten, *J. Chem. Phys.*, **58**, 4496 (1973). Coulombic Potential Energy Integrals and Approximations.